WHEN BETTER DRAFTS COST MORE VERIFICATION: SAFETY BOUNDARIES FOR MULTI-TOKEN DRAFTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Speculative-decoding draft heads are trained to match the target model at each position. Yet a better local draft can make a whole response more expensive to verify: accepting more tokens moves where verification restarts, possibly to a costlier continuation. We characterize this effect for a fixed target path with a fixed proposal cap, full-prefix commitment and unit cost per verifier call. First, same-rank readouts can raise some target-match probabilities, lower none, yet raise verifier calls in proportion to the cap. Second, every such improvement is safe when the expected accepted length is at most one at every restart, the largest uniform threshold, or when prefix-survival probabilities are consistent across restarts; a sharp bound prices departures from consistency. Third, an exact backward recursion decides safety outside both regions, certifying 370 of 10,752 fitted probability fields against every coordinatewise increase in match probabilities, where the threshold alone certifies two. Finally, in a preregistered ten-seed comparison on two frozen backbones with matched anchors and evaluation budgets, whole-response training wins none of eight comparisons against a local objective covering every anchor, and the local objective reduces sampled verification calls in four. Draft-head changes should therefore be judged by whole-response verification work; our conditions and exact test show when a local improvement is guaranteed safe.

1 INTRODUCTION

Speculative decoding lets a drafter propose several tokens that the target model verifies in parallel (Leviathan et al., 2023; Chen et al., 2023). Draft heads are trained for local target agreement, through cross entropy or acceptance-aware objectives (Zhang et al., 2025; Samarin et al., 2026; Wu et al., 2026; Yang & Li, 2026; Zou et al., 2026). These objectives reward agreement at each anchor, but they do not price where the next verification starts. This matters: a draft head can improve its token predictions yet make a complete response more expensive. A longer accepted prefix changes where verification resumes, so it can skip an inexpensive continuation and land at a costlier one (Figure 1).

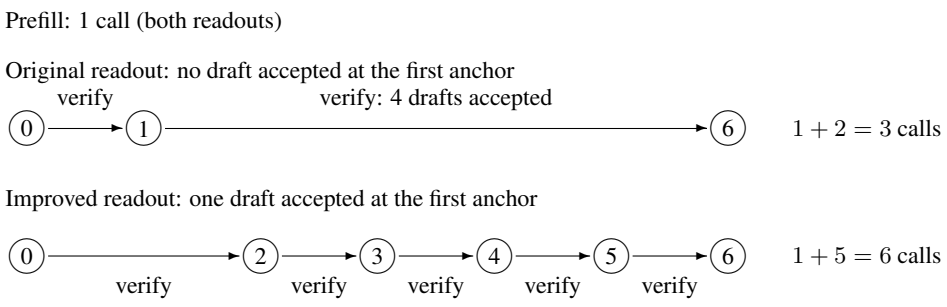


Figure 1: Improving one draft raises verification calls from three to six. With four proposals and seven output tokens, both same-rank argmax readouts pay one prefill call plus one call per arrow; circles mark positions, and 6 is terminal. Improving only the first draft at anchor 0 skips anchor 1, where all four drafts would match.

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.

We study this effect in its cleanest form. We fix independent proposal sampling, a proposal cap, maximal-prefix commitment and unit cost per verifier call, which isolates the anchor effect from hardware-dependent costs (Liu et al., 2026c). A probability field records match probabilities across restarts and draft depths. SpecDec++ models adaptive whole-response cost (Huang et al., 2025); we characterize when a fixed verification rule is safe against better drafts. Our contributions are:

- We construct same-rank readouts that raise some target-match probabilities and lower none, yet increase verifier calls by a factor proportional to the proposal cap (Theorem 1).
- We give two safety structures. Expected accepted length at most one at every anchor is the largest uniform safety threshold (Theorem 2). Prefix-survival probabilities that are consistent across restart anchors are also safe and permit high acceptance; the greatest nested field below the original field gives a sharp bound on the cost of inconsistency (Theorem 3). Both structures ensure that moving to a later destination cannot increase remaining work, so they protect against every improvement in the distribution of accepted lengths.
- We give an exact backward recursion that decides safety for the smaller class of increases in individual conditional match probabilities and returns a harmful change whenever one exists (Equation (6)). With strictly interior probabilities and a fixed cap, the two safety judgments can differ only through cost differences at early destinations. On 10,752 fitted fields, the recursion certifies 370 and the uniform threshold alone certifies two.
- In a preregistered ten-seed comparison with the same head class, anchors and evaluation budget, whole-response training wins none of eight comparisons against a local objective covering every anchor, and the local objective reduces sampled verification calls in four (Section 4).

The practical message is to judge a draft-head change by its effect on whole-response verification work, and to use the safety conditions and the exact test to know when a local improvement is guaranteed to be safe.

2 Problem: Heads, Work and the Local Objective

Local prediction quality and response work are different functionals of the same frozen-representation head.

A fixed head class. Let $h \in \mathbb{R}^d$ be a frozen target-model representation and $W \in \mathbb{R}^{d\times|\mathcal{V}|}$ its frozen unembedding. For proposal depth j, the trainable logits are

$$z_j(h) = \big[h + (h\Omega)\Gamma_j\big]W + b_j, \qquad \Omega \in \mathbb{R}^{d\times r}, \quad \Gamma_j \in \mathbb{R}^{r\times d}. \tag{1}$$

The principal local and whole-response training arms fit the same full-vocabulary factors and biases. They use four active depths and rank $r = 108$. The backbone is not updated. The probability-field theorems apply to any drafter whose decoder and probability field satisfy their stated conditions; our experiments use these linear readouts.

The complete-response objective. Fix a target-greedy path of length n, ending at its first end-of-sequence token or the response-length cap. Prefill emits its first token. There are $M = n - 1$ subsequent anchors, including the final anchor with no active proposal. At anchor a, set $H_a = \min(K, M-a-1)$ and let $p_{a,j}$ be the probability that an independent categorical sample from draft head j matches the corresponding token on this fixed target-greedy path. Thus $p_{a,j}$ is that token's softmax mass. Define

$$A_{a,k} = \prod_{j=1}^{k} p_{a,j}, \qquad S_p(a) = \sum_{k=1}^{H_a} A_{a,k}.$$

Thus $S_p(a)$ is the expected accepted prefix length. Accepting ℓ drafts moves the verifier to $a+\ell+1$. With $A_{a,0} = 1$, its transition probabilities are $A_{a,\ell} - A_{a,\ell+1}$ for $\ell < H_a$, and A_{a,H_a} for $\ell = H_a$.

Let $V_p(M) = 0$. Backward recursion of this finite increasing chain gives remaining expected calls $V_p(a)$. Our objective is

$$F_p = \frac{1 + V_p(0)}{n} = \frac{J(p)}{n}, \tag{2}$$

where the one accounts for prefill. Length-one and length-two paths remain in the average, although their gradients are zero. The denominator is the fixed target length rather than a method-dependent accepted-token count.

For $M > 0$, let $d_p(a)$ be the probability of visiting nonterminal anchor $a \in \{0, \dots, M-1\}$ and let $\bar{S}_{\text{visit}}(p) = \sum_{a=0}^{M-1} d_p(a) S_p(a) / \sum_{a=0}^{M-1} d_p(a)$. Every execution advances exactly M post-prefill positions, so an accounting identity gives

$$J(p) = 1 + \frac{M}{1 + \bar{S}_{\text{visit}}(p)}. \tag{3}$$

This is a ratio of expectations for one fixed response. Because the head also changes which anchors are visited, better predictions at fixed anchors need not improve this on-policy average.

What is held fixed. The sampled-policy theory and empirical objective fix maximal-prefix commitment, proposal count, and unit call cost. The separate deterministic construction states its argmax rule explicitly. Full-prefix commitment matters: if a controller may truncate or discard drafts for free, a stronger drafter can emulate a weaker prefix law under the conditions in Appendix J. The main comparisons fix a canonical full-prefix FP32 target path and optimize its expected call count; Appendix Q marks the separate native-BF16 stratum. The objective concerns computational work rather than task-answer accuracy.

The training comparison. Both objectives evaluate every active anchor. The local-prefix objective minimizes $-\sum_a S_p(a)/n$, while the whole-response objective minimizes F_p. Their gradients differ in visit weights and continuation credit, not in which probability rows are available. The exact gradient formula and estimator controls are in Appendix I.

Five distinct objects. The analysis keeps five objects apart: a *scalar local loss*, which aggregates probabilities; *coordinate dominance*, $q_{a,j} \geq p_{a,j}$ everywhere; a *feasible shared update*; an *actual Adam trajectory*, which also depends on moments and the optimizer clock; and *native latency*, which additionally prices drafting, verification and hardware. The theorems concern the second object, while the training test measures the fourth.

3 Local Improvements and Verification Work

A better draft can change where verification restarts. The decisive question is whether those new destinations have cheaper continuations.

Intuition. Accepting more drafts can skip a cheap continuation and visit an expensive one.

Theorem 1 (Same-rank local dominance and work separation)**.** *For $K \geq 2$, $n \geq 1$ periods and target length $N = n(K+2)+1$, there are a frozen binary target with a three-dimensional representation and two K-head shared readouts in the same exact joint-rank-two pre-unembedding factor stratum. Every head's target-match probability is nondecreasing at every anchor, yet fixed-K argmax proposals with prefill and a $1+L$ verifier step require*

$$C_{\text{old}} = 1 + 2n, \qquad C_{\text{new}} = 1 + n(K+1).$$

Thus local dominance permits a $\Theta(K)$ call-count ratio with unchanged per-call cost and factor bills. Finite-logit independent sampling admits the same construction with $|\mathbb{E}C_i - C_i| \leq 1/8$ at both endpoints.

Appendix C gives the proof for this constructed representation; costs count target calls.

A sharp safe operating region. Let $S_p(a) = \sum_{k=1}^{H_a} \prod_{i=1}^{k} p_{a,i}$ be the expected accepted prefix at anchor a, with H_a the remaining-cap limit.

Intuition. A safe baseline cannot become more costly if no proposal's match probability decreases.

Theorem 2 (Sharp safe-baseline threshold)**.** *For independent sampled proposals on a fixed target path with unit call cost, $S_p(a) \leq 1$ everywhere guarantees $J(q) \leq J(p)$ for* every *coordinatewise improvement $q \geq p$, even if S_q exceeds one. One is the largest uniform threshold: for every $\gamma > 1$ and $K \geq 2$, finite-logit rank-two readouts with all $S < \gamma$ admit a harmful root improvement.*

Proof sketch. Let $V_p(a)$ be the expected remaining calls and $D_p(a) = V_p(a) - V_p(a+1)$. Backward induction in $D_p(a) = 1 - \sum_k (\prod_{i=1}^{k} p_{a,i}) D_p(a+k)$ gives $0 \le D_p(a) \le 1$. Improving a prefix therefore shifts probability only toward continuations with nonnegative savings. The full proof and sharpness construction are in Appendix H.3; its floor is fixed in K for fixed $\gamma > 1$. Above-one acceptance is necessary, not sufficient, for harmful credit. For fixed positive costs, Appendix D gives the weighted sufficient condition $\sum_k A_{p,a,k} c_{a+k} \le c_a$.

One safety structure, two improvement sets. Write $A_{a,k} = \Pr(L_a \ge k)$ for *joint* prefix survival, not a product of marginals for dependent drafts. Cross-restart nesting means $A_{a,k} \le A_{a+1,k-1}$ for $2 \le k \le H_a$. It is sufficient even at high acceptance. An autoregressive drafter satisfies it when its conditional law at a target prefix is independent of the restart anchor; sequential dependence alone is insufficient. Define the risks $R_{\mathrm{up}}(p) = \max_{q \ge p}\{J(q) - J(p)\}$ and $\mathcal{R}_{\mathrm{surv}}(A) = \sup_{C \ge A}\{J(C) - J(A)\}$, respectively, over independently adjustable conditional and valid survival fields, and the suffix credit

$$\mathcal{C}_{a,j} = \sum_{k=j}^{H_a} \left(\prod_{\ell=j+1}^{k} p_{a,\ell} \right) D_p(a+k).$$

On the fixed target path with $M \ge 2$, $K \ge 1$, $H_a = \min\{K, M-a-1\}$, unit costs and $0 < p_{a,j} < 1$ at every active coordinate, the results fit together as

$$\begin{aligned} &[S_p \le 1 \text{ everywhere}] \text{ or } [\text{cross-restart nesting}] \\ &\quad \Longrightarrow D_p(b) \ge 0 \quad (1 \le b < M) \\ &\quad \Longleftrightarrow \mathcal{R}_{\mathrm{surv}}(A) = 0 \quad \Longleftrightarrow \quad \text{maximal commitment is posterior-optimal} \\ &\quad \Longrightarrow R_{\mathrm{up}}(p) = 0 \quad \Longleftrightarrow \quad \mathcal{C}_{a,j} \ge 0 \text{ for all active } (a,j). \end{aligned} \tag{4}$$

Here the posterior controller knows the complete future field and, after observing L_a, may commit any $0, \dots, L_a$ matched drafts at unchanged call cost. Optimal pre-draw caps with nondecreasing draft costs are a separate comparison (Appendix E). No legal destination comparison uses $D_p(0)$. Survival improvements can lower a later conditional probability, so the reverse safety implication need not hold, and risk and posterior policy gain need not be numerically equal.

Why the cap matters. Put $K_\star = \min\{K, M-1\}$. Since the last proposal exposes $\mathcal{C}_{a,H_a} = D_p(a+H_a)$, every destination $K_\star \le b < M$ is tested directly. Conditional safety is therefore equivalent to these tail differences being nonnegative and credits with $a + j < K_\star$ being nonnegative. If conditional safety holds but survival safety fails, negative destination differences lie only at $1, \dots, K_\star - 1$. Appendix D.6 proves the equivalences and cap result using classical policy improvement (Agarwal et al., 2021; Bhandari & Russo, 2024); boundary probabilities require the exact recursion below.

Quantifying a failure of nesting. *Intuition.* The extra work needed to repair cross-anchor inconsistency bounds how harmful further local improvements can be.

Theorem 3 (Sharp risk bound from prefix nesting)**.** *For any valid survival field A on a fixed path, put $B_{a,k} = \min_{0 \le j < k} A_{a+j,k-j}$. This is its greatest nested minorant. Let $d_A(a)$ be the probability of visiting nonterminal anchor a and $D_B(a) = V_B(a) - V_B(a+1)$. Then*

$$\sup_{C \ge A}\{J(C) - J(A)\} \le J(B) - J(A) = \sum_{a=0}^{M-1} d_A(a) \sum_{k=1}^{H_a} (A_{a,k} - B_{a,k}) D_B(a+k). \tag{5}$$

The supremum is over valid survival fields. The bound takes $O(MK)$ arithmetic operations, is attained by a four-anchor family, and is smallest among bounds obtained from nested minorants of A.

Proof sketch. The recurrence $B_{a,k} = \min\{A_{a,k}, B_{a+1,k-1}\}$ preserves legal survival rows and enforces nesting. Every $C \ge A$ costs no more than the safe baseline B, so $J(C) \le J(B)$. The finite-chain performance-difference identity gives the equality. Appendix D.5 proves sharpness and an adjacent-defect bound. Stochastic monotonicity is classical (Daley, 1968; Puterman, 1994); the specialization quantifies the work consequence of nesting defects.

An exact safety test. The largest harmful conditional improvement is computable exactly. Set $U(M) = 0$. For $a = M-1, \dots, 0$, put $F_{H_a+1} = U(a + H_a + 1)$ and recurse over $j = H_a, \dots, 1$:

$$\begin{aligned} \Psi_{a,j} &= (1 - p_{a,j})U(a + j) + p_{a,j}F_{j+1}, \\ F_j &= \max\{\Psi_{a,j}, F_{j+1}\}, \qquad U(a) = 1 + F_1, \\ R_{\mathrm{up}}(p) &= U(0) - V_p(0). \end{aligned} \tag{6}$$

For an empty row, use $U(a) = 1+U(a+1)$. Choose $q_{a,j} = p_{a,j}$ if $\Psi_{a,j} \geq F_{j+1}$, and $q_{a,j} = 1$ otherwise. Affine inner objectives and independent rows let this backward pass attain every continuation maximum (Appendix D).

Zero risk certifies every coordinatewise improvement; positive risk supplies a harmful free-field change. The backward pass applies rectangular robust dynamic programming (Iyengar, 2005) to the prefix chain. Once the field is available, evaluation and witness replay take $O(MK)$ arithmetic operations. Finite logits approach saturated maximizing coordinates by continuity. Section 4.2 applies the test to 10,752 fitted fields.

Shared updates. Zero free-field risk protects every reachable subset. For a finite set of paths with Ω, W and biases fixed, Appendix G bounds risk from one norm-bounded $\Delta\Gamma$ in its stored gauge and constructs positive rectangular but zero common-update risk. On one path with unbounded updates, full row rank and exposed target columns suffice for free-field equality (Appendix F). The fitted linear heads instantiate the fixed-path model.

Safe fields and training steps. Coupled parameters can improve some probabilities at the expense of others. A rank-one example has all $S \leq 3/4$ but $\langle \nabla J, \nabla(-\sum_a S_p(a)) \rangle = -377/64$: local-loss SGD increases work inside the safe region (Appendix J). Even arbitrarily long finite local windows can conceal opposite useful root updates (Appendix H). General finite-lookahead limitations are established (Efroni et al., 2018; Xu & Gan, 2026).

4 Controlled Experiments

The strongest comparison matches anchors, head class and evaluation budget, separating continuation credit from the benefit of evaluating more anchors.

4.1 Ten seeds on a prospective prompt cohort

We retain two frozen backbones, SmolLM3-3B and OLMo-2-1B, two corpora, GSM8K and MBPP (Cobbe et al., 2021; Austin et al., 2021), and both fixed learning rates, 3×10^{-5} and 10^{-4}. All-anchor local and whole-response training use rank-108 heads and 128 clipped Adam updates of eight paths. Ten training seeds give 80 heads: 24 historical heads and 56 new fits. All are evaluated on 128 new prompts per corpus, without choosing heads, rates or seeds from these outcomes. The new prompts follow a registered exclusion rule.

The primary endpoint averages calls per token over 32 independent-proposal replays with common random numbers. This sampler makes target-token softmax mass a match probability; argmax proposals are a separate endpoint. The fixed target responses contain at most 64 tokens and use at most four proposals. Prefill is included. Proposal repetitions reduce simulation error and are not counted as seeds or prompts. The registered analysis treats seeds and prompts as crossed random factors, uses a Satterthwaite t approximation, and applies Holm correction to all eight comparisons. Table 1 gives all eight primary contrasts; Figure 2 also displays the argmax endpoint.

In all four SmolLM3-3B comparisons, whole-response training needs 0.423% to 0.894% more sampled calls than local training, and each has the same direction in all ten seeds. Whole-response training wins none of the eight comparisons, and analytic expected calls give the same four-versus-zero split. Independent recomputation preserves every primary decision; Appendix A gives full values, seed effects and a hardware replication.

Corpus	Rate	Difference	95% interval	p_{Holm}
SmolLM3-3B				
GSM8K	3×10^{-5}	$+2.788$	$[+1.613, +3.962]$	2.8×10^{-8}
GSM8K	10^{-4}	$+2.603$	$[+1.216, +3.989]$	6.0×10^{-6}
MBPP	3×10^{-5}	$+5.121$	$[+3.376, +6.865]$	3.6×10^{-10}
MBPP	10^{-4}	$+4.278$	$[+2.255, +6.302]$	3.7×10^{-7}
OLMo-2-1B				
GSM8K	3×10^{-5}	$+0.680$	$[-0.620, +1.980]$	0.418
GSM8K	10^{-4}	-0.740	$[-2.530, +1.050]$	0.488
MBPP	3×10^{-5}	$+1.125$	$[-0.237, +2.488]$	0.087
MBPP	10^{-4}	$+0.144$	$[-1.862, +2.149]$	0.840

Table 1: Ten-seed primary comparison, in calls per 1,000 tokens. Differences are whole-response minus local verifier calls, including prefill, so positive values favor local training; each row uses ten seeds and 128 new prompts. Intervals are approximate familywise 95% Bonferroni intervals from crossed seed-and-prompt inference, and Holm tests use all eight contrasts at level 0.05.

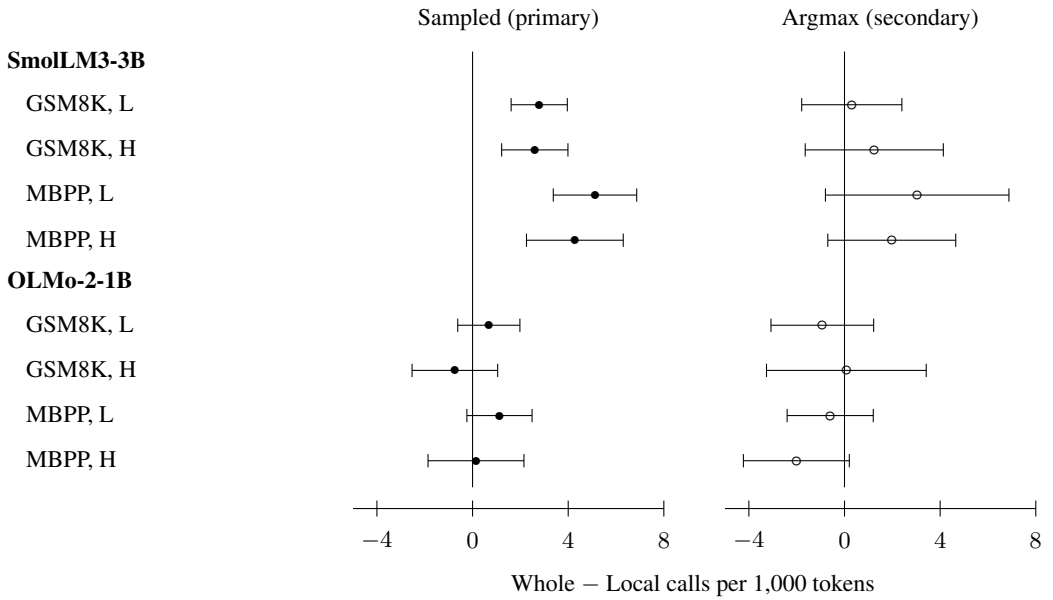


Figure 2: Ten-seed prospective comparison for sampled (primary) and argmax proposals. Positive whole-response minus local calls favor local training; each row uses ten training seeds and 128 new prompts. Bars are approximate familywise 95% Bonferroni intervals from the crossed seed-and-prompt analysis, and L and H denote learning rates 3×10^{-5} and 10^{-4}.

4.2 How often the exact test certifies safety

On 10,752 saved Qwen3.5-0.8B probability fields from 42 method configurations, the exact recursion certifies 370 fields, or 3.44%, against every coordinatewise improvement. The uniform threshold alone certifies two. For each of the remaining 10,382 fields the recursion returns a maximizing free-field improvement; the largest increase is 0.0371 calls per token. On the same complete fields, mean conditional risk is 0.2345 calls against 33.2905 baseline calls, a ratio of means of 0.70%. Survival-order risk averages 0.3935 calls and the nested-minorant bound 0.6368; exactly one field is nested. The full-information posterior controller saves 0.2240 calls on average (Appendix D).

All saved active probabilities are strictly interior. Table 2 shows all three safety classes. For $K = 4$, the 126 disagreements have negative destination differences only in the first three positions, as the cap result requires; the double-unsafe fields contain 56,112 later negative differences

Risk class	Fields	Prompt clusters (overlapping)	$D_p(b) < 0$ $b = 1, 2, 3$	$b \geq 4$
Both risks zero	244	69	0	0
Only conditional risk zero	126	27	152	0
Both risks positive	10,382	508	7,689	56,112

Table 2: Where the two safety notions differ, over all 10,752 fields (42 configurations, 512 prompt clusters). In the middle class, all 152 negative destination differences lie at $b = 1, 2, 3$. Columns count fields, overlapping prompt clusters or destination differences.

(Appendix D.6). Full-vocabulary path-specific constructions pass both numerical dominance and work-increase checks for 10,266 fields, each with its own path-specific update (Appendix F).

4.3 Why the comparison needs strong local controls

Earlier comparisons explain why anchor coverage matters. Against one sampled local anchor, whole-response training reduces work by 3.11% to 4.97%, but it evaluates about eight times as many vocabulary rows. Its advantage narrows against eight-anchor blocks. In the original three-seed all-anchor comparison, local training is favored in seven of eight analytic-work contrasts under the preregistered Bonferroni adjustment (Appendix P). These earlier results are distinct from the ten-seed test above (Appendices N, O and Q).

A scalar local loss can rank heads differently from verification work. In 3,072 paired Qwen3.5-0.8B endpoint records, local loss and verification work disagree in 879, or 28.6%: 812 of 1,721 pairs favoring the locally trained head on local loss, and 67 of 1,351 favoring the whole-response-trained head. The signed continuation-credit identity accounts for all 3,072 differences (Appendix R). Exchanging complete Adam state changes update rankings (Appendix M).

5 Related Work

Local acceptance and complete-response work. Speculative decoding verifies draft prefixes in parallel (Leviathan et al., 2023; Chen et al., 2023). Multi-token drafting heads descend from blockwise parallel decoding (Stern et al., 2018) and include Medusa heads on a frozen backbone (Cai et al., 2024) and multi-token prediction objectives (Gloeckle et al., 2024); EAGLE drafts autoregressively from target features (Li et al., 2024). These methods train for local agreement, which is the objective whose effect on complete-response work we analyze. A systematic study finds that draft latency matters more than the draft model's language-modeling quality for speculative-decoding speed (Yan et al., 2025), and adaptive drafting with reinforcement learning optimizes throughput instead of acceptance length (Zhang et al., 2026); our analysis isolates a different mechanism, the moving verification anchor, at unit call cost. A Markov-chain analysis of speculative decoding establishes the optimality of rejection sampling and the expected number of rejections for fixed models (Yin et al., 2024); we instead compare two drafts under one fixed verification rule. SpecDec++ models complete generation cost and learns an adaptive stop rule (Huang et al., 2025). Block verification goes beyond token-by-token acceptance of the longest matching prefix, and its Theorem 2 optimizes cumulative generated length after every fixed iteration count for fixed target and drafter models (Sun et al., 2025); we instead keep full-prefix commitment fixed and ask which stronger fields remain safe under it. Prefix scheduling compares skip eligibility at equal positionwise metrics (Luo et al., 2026); our counterexample instead raises some of the draft's own match probabilities and lowers none.

Acceptance-aware training. Knowledge distillation aligns drafters with the target model (Zhou et al., 2024). Acceptance-aware objectives use prefix-confidence weights (Wu et al., 2026; An et al., 2026), first-mismatch supervision (Yang & Li, 2026), draft–target overlap (Samarin et al., 2026), a path-acceptance lower bound (Zou et al., 2026), or draft-tree rewards (Hu et al., 2026). Verification-aware training can leave the inference rule unchanged (Gu et al., 2026). Without smoothing, D-PACE's detached coefficients recover the local-prefix gradient up to sign and normalization; that weighting is not our contribution. Sequential-survival and expected-accepted-length losses (Gu

et al., 2026; Xia et al., 2026) likewise differ from pricing the moving verification anchor. GrowMTP trains a detached draft head during reinforcement learning, distinguishes accepted length from step time, and analyzes shared gradients, SGD and Adam (He et al., 2026, Appendix I). Our results concern the safety of specified probability changes, rather than a first acceptance-aware loss or connection between gradients and optimization. The training comparison uses the full all-anchor local objective rather than a sampled-anchor surrogate.

Safety and control. Stochastically monotone chains and the monotonicity of Markov decision values in stochastic orders (Daley, 1968; Puterman, 1994; Müller, 1997), monotone stochastic bounds (Fourneau & Pekergin, 2002), rectangular robust dynamic programming (Iyengar, 2005; Nilim & El Ghaoui, 2005) and policy improvement (Agarwal et al., 2021; Bhandari & Russo, 2024) supply the classical tools: cross-restart nesting is the stochastic monotonicity of the restart chain, and the greatest nested minorant is its maximal monotone lower bound. Our specialized results are the sharp uniform threshold, the sharp work bound from the nested minorant, and the fixed-cap relation between conditional and survival safety. KL certificates address local acceptance and prefix length (Sharma, 2026), whereas our test concerns complete-response upgrades. HedgeSpec studies online drafter selection and discusses EAGLE's relative-position dependence, censoring and off-policy acceptance (Liu et al., 2026b, Appendix B.4); our safety tests instead concern changes to one fixed field. Faster Cascades optimizes deferral across two models (Narasimhan et al., 2025), whereas we fix the verification rule. Online Speculative Decoding adapts drafters to observed queries (Liu et al., 2024).

Architecture and runtime scope. An autoregressive drafter whose conditional law depends only on the visible target prefix satisfies nesting, so position heterogeneity alone does not cause our reversal. Hydra adds sequential dependence among draft heads within a block (Ankner et al., 2024); nesting is a separate, cross-anchor property. Feature-chain, diffusion and tree drafters (Li et al., 2025; Chen et al., 2026; Miao et al., 2024) fall under the theorems whenever their prefix laws meet the stated conditions. Native multi-token modules (DeepSeek-AI, 2024; GLM-5-Team, 2026; Xiaomi MiMo, 2026) differ from the exact-rank readout pair of Theorem 1. Hardware speed further depends on verification cost, horizon censoring, sampler and token counting (Liu et al., 2026c; Wu, 2026; Chordiya, 2026).

6 Conclusion

A better local draft can make a whole response more expensive to verify, because accepting more tokens moves where verification restarts. For a fixed target path, proposal cap, full-prefix commitment and unit call cost, we constructed same-rank readouts with this behavior, identified the largest uniform safety threshold and a nesting condition with a sharp risk bound, and gave an exact test that decides safety against coordinatewise increases outside both regions. In a matched ten-seed comparison, whole-response training won none of eight comparisons against a local objective covering every anchor. Draft-head changes should be judged by their whole-response verification work, with the safety conditions and the exact test showing when a local improvement is guaranteed to be safe.

7 Limitations

The theory fixes a greedy target path, the proposal cap, maximal-prefix commitment and unit call cost. Its sampled proposals differ from argmax drafting and from stochastic-target rejection sampling, and free truncation of accepted drafts changes the comparison. Average acceptance below one does not imply $S \leq 1$ everywhere. The recursion and local safety test specialize rectangular robust dynamic programming and policy improvement. The test, the cap result and the posterior controller use the complete future field, whose acquisition cost is excluded; the posterior controller is not an online controller, and the cap result locates negative destination differences rather than all positive risk. The nesting bound can exceed the risk by a diverging factor even away from probability boundaries, and free-field risk can be loose relative to risk realizable by shared parameters. The finite-window result restricts observations and is not a lower bound for arbitrary queries.

The safety results do not guarantee useful shared-parameter updates. Improving a scalar loss need not yield coordinate dominance, a dominating field need not come from a feasible shared update,

field safety does not order training steps, and unit-call work need not preserve native-latency rankings. Fitted factors need not stay on the exact-rank stratum of the constructions. The shared-update bound fixes Ω, W and biases, so it does not certify the full head. Path-specific constructions do not establish one update shared across prompts, and none of three fixed joint attempts establishes a dominating work-increasing update. A prespecified Γ-only construction retained a zero update in 95 of 96 configurations, and the 96th configuration is unavailable, so its norm-matched comparison was uninformative (Appendix G). A two-prompt common ray leaves its native precision check unresolved and does not improve evaluation work (Appendix G.1). No pair among the 3,072 fitted pairs improves every probability coordinate, so the experiments do not instantiate the premise of Theorem 1, and the signed accounting does not isolate a causal training mechanism.

Prospective effects are small and span two backbones: whole-response training needs 0.423% to 0.894% more sampled calls on SmolLM3-3B, the four OLMo-2-1B comparisons are unresolved, and with argmax proposals neither objective is favored in any of the eight. In the earlier three-seed development comparison, the eighth contrast (OLMo-2-1B, GSM8K, rate 10^{-4}) is unresolved under the preregistered Bonferroni adjustment but favors whole-response training under a Holm sensitivity analysis (adjusted $p = 0.0093$), and argmax favors it in one contrast. There, local updates are larger despite matched learning rates, leaving effective step size as an alternative explanation. Additional development families are mixed and not pooled, an occasional complete-gradient correction worsens work in all four registered comparisons, and Adam-state component interventions identify no generally better optimizer. Risk-based selection among 24 heads changes six of eight choices but establishes no work reduction: six adjusted intervals contain zero and two identical choices give exact zero. A secondary safe-path-count rule increases work for SmolLM3/MBPP at the higher rate by 0.008320 calls per token, interval $[0.002579, 0.014065]$ (Appendix B). Selection reuses the ten-seed cohort, so the two prospective tests are not independent confirmations; the new prompts come from the same corpora and are not verified exclusions from pretraining, and earlier tests use development prompts. A secondary sign statistic is sensitive to roundoff (Appendix A). The 10,752 fields are dependent through recurring prompts, checkpoints and histories, and 768 of them duplicate other fields exactly, so their counts describe this fitted population rather than a universal frequency of safety. Experiments use frozen backbones, linear readouts, responses of at most 64 tokens and at most four proposals. Native autoregressive, EAGLE and DFlash drafters and serving latency are not measured, and equality with other key-value-cache or batched execution paths is not claimed.

AI ASSISTANCE AND RESPONSIBILITY

Large language model assistants were used as research and writing tools: for literature search, discussing ideas, writing code, checking derivations and editing the text. The authors directed the work, verified every proof, result, number and citation, and take full responsibility for the paper and its supplementary material.

REPRODUCIBILITY AND ETHICS

The accompanying materials contain scoped analysis implementations, saved results and reconstruction instructions. For the 96-cell common-update attempt, the anonymous saved-result attachment reconstructs tables but not the original solver, head fitting or full-vocabulary forward evaluation. The separate mathematical-example package reconstructs exact finite-prefix examples and synthetic affine-softmax checks, not trained-head experiments. An additional anonymous scalar-data package reproduces the complete 10,752-field class counts, means and ratio-of-means summaries. It does not reconstruct neural probabilities or model fitting. Replaying saved results does not test a serving implementation. The study uses public datasets and model checkpoints and involves no human subjects.

REFERENCES

01.AI. Yi-1.5: Official model release. Official model repository, fixed revision, 2024. Fixed publisher source revision, accessed September 14, 2026.

Alekh Agarwal, Sham M. Kakade, Jason D. Lee, and Gaurav Mahajan. On the theory of policy gradient methods: Optimality, approximation, and distribution shift. *Journal of Machine Learning Research*, 22(98):1–76, 2021. URL https://jmlr.org/papers/v22/19-736.html.

Christopher Aicher, Nicholas J. Foti, and Emily B. Fox. Adaptively truncating backpropagation through time to control gradient bias. In *Proceedings of the 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pp. 799–808. PMLR, 2020. URL https://proceedings.mlr.press/v115/aicher20a.html.

Zihao An, Taichi Liu, Ziqiong Liu, Dong Li, Ruofeng Liu, and Emad Barsoum. PARD-2: Target-aligned parallel draft model for dual-mode speculative decoding, 2026. URL https://arxiv.org/abs/2605.08632v1.

Zachary Ankner, Rishab Parthasarathy, Aniruddha Nrusimha, Christopher Rinard, Jonathan Ragan-Kelley, and William Brandon. Hydra: Sequentially-dependent draft heads for Medusa decoding. In *First Conference on Language Modeling (COLM)*, 2024. URL https://openreview.net/forum?id=FbhjirzvJG.

Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, and Charles Sutton. Program synthesis with large language models, 2021.

Francis Bach. Self-concordant analysis for logistic regression. *arXiv preprint arXiv:0910.4627*, 2009. URL https://arxiv.org/abs/0910.4627.

Alex Beatson and Ryan P. Adams. Efficient optimization of loops and limits with randomized telescoping sums. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 534–543. PMLR, 2019. URL https://proceedings.mlr.press/v97/beatson19a.html.

George Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962. doi: 10.1080/01621459.1962.10482149. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1962.10482149.

Jalaj Bhandari and Daniel Russo. Global optimality guarantees for policy gradient methods. *Operations Research*, 72(5):1906–1927, 2024. doi: 10.1287/opre.2021.0014. URL https://pubsonline.informs.org/doi/10.1287/opre.2021.0014.

Shrenik Bhansali and Larry Heck. Draft, verify, and improve: Toward training-aware speculative decoding. arXiv preprint arXiv:2510.05421, 2025. URL https://arxiv.org/abs/2510.05421.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. doi: 10.1017/CBO9780511804441. URL https://web.stanford.edu/~boyd/cvxbook/.

Tianle Cai, Yuhong Li, Zhengyang Geng, Hongwu Peng, Jason D. Lee, Deming Chen, and Tri Dao. Medusa: Simple LLM inference acceleration framework with multiple decoding heads. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 5209–5235. PMLR, 2024. URL https://proceedings.mlr.press/v235/cai24b.html.

Charlie Chen, Sebastian Borgeaud, Geoffrey Irving, Jean-Baptiste Lespiau, Laurent Sifre, and John Jumper. Accelerating large language model decoding with speculative sampling, 2023. URL https://arxiv.org/abs/2302.01318v1.

Jian Chen, Yesheng Liang, and Zhijian Liu. DFlash: Block diffusion for flash speculative decoding. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2602.06036v2. arXiv:2602.06036v2.

Param Chordiya. Lossless but not free: An empirical anatomy of speculative decoding on consumer hardware. arXiv preprint arXiv:2607.17283v1, 2026. URL https://arxiv.org/abs/2607.17283v1.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems, 2021.

D. J. Daley. Stochastically monotone Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 10(4):305–317, 1968.

DeepSeek-AI. DeepSeek-V3 technical report, 2024. URL https://arxiv.org/abs/2412.19437.

Yonathan Efroni, Gal Dalal, Bruno Scherrer, and Shie Mannor. Beyond the one-step greedy approach in reinforcement learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1387–1396. PMLR, 2018. URL https://proceedings.mlr.press/v80/efroni18a.html.

Jean-Michel Fourneau and Nihal Pekergin. An algorithmic approach to stochastic bounds. In *Performance Evaluation of Complex Systems: Techniques and Tools*, volume 2459 of *Lecture Notes in Computer Science*, pp. 64–88. Springer, 2002.

GLM-5-Team. GLM-5: From vibe coding to agentic engineering. arXiv preprint arXiv:2602.15763v2, 2026. URL https://arxiv.org/abs/2602.15763v2.

Fabian Gloeckle, Badr Youbi Idrissi, Baptiste Roziere, David Lopez-Paz, and Gabriel Synnaeve. Better & faster large language models via multi-token prediction. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *PMLR*, pp. 15706–15734. PMLR, 2024. URL https://proceedings.mlr.press/v235/gloeckle24a.html.

Geonmo Gu, Byeongho Heo, HeeJae Jun, Yoohoon Kang, Sangmin Lee, Sangdoo Yun, and Dongyoon Han. Verification-aware training for speculative decoding. arXiv preprint arXiv:2608.30135v1, 2026. URL https://arxiv.org/abs/2608.30135v1.

Minghua He, Lingzhe Zhang, Yuan Liu, Xiao Zhou, and Aiwei Liu. GrowMTP: Can RL grow its own draft head? arXiv preprint arXiv:2609.16648, 2026. URL https://arxiv.org/abs/2609.16648.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500830.

Yu-Guan Hsieh, James Thornton, Eugene Ndiaye, Michal Klein, Marco Cuturi, and Pierre Ablin. Careful with that scalpel: Improving gradient surgery with an EMA. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 19085–19100, 2024. URL https://proceedings.mlr.press/v235/hsieh24a.html.

Shijing Hu, Jingyang Li, Zhihui Lu, and Pan Zhou. Bridging draft policy misalignment: Group tree optimization for speculative decoding. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/b6e67ae290635d0874c4cb43ba2a2cfb-Abstract-Conference.html.

Kaixuan Huang, Xudong Guo, and Mengdi Wang. SpecDec++: Boosting speculative decoding via adaptive candidate lengths. In *Conference on Language Modeling*, 2025. URL https://arxiv.org/abs/2405.19715. arXiv:2405.19715v3.

IBM Granite Team. Granite-4.1-3B: Model card. Hugging Face model repository, 2026a. URL https://huggingface.co/ibm-granite/granite-4.1-3b. Accessed September 16, 2026; exact experimental revision in the artifact.

IBM Granite Team. Granite-4.2-8B: Model card. Hugging Face model repository, fixed revision, 2026b. Fixed checkpoint revision; accessed September 21, 2026.

IBM Granite Team. Granite-4.2-3B: Model card. Hugging Face model repository, fixed revision, 2026c. Fixed checkpoint revision; accessed September 21, 2026.

Garud N. Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2): 257–280, 2005. doi: 10.1287/moor.1040.0129. URL https://doi.org/10.1287/moor.1040.0129.

Pierre E. Jacob, John O'Leary, and Yves F. Atchadé. Unbiased markov chain monte carlo with couplings. arXiv preprint arXiv:1708.03625v5, 2019. URL https://arxiv.org/abs/1708.03625v5.

Yaniv Leviathan, Matan Kalman, and Yossi Matias. Fast inference from transformers via speculative decoding. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 19274–19286. PMLR, 2023. URL https://proceedings.mlr.press/v202/leviathan23a.html.

Yuhui Li, Fangyun Wei, Chao Zhang, and Hongyang Zhang. EAGLE: Speculative sampling requires rethinking feature uncertainty. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 28935–28948. PMLR, 2024. URL https://proceedings.mlr.press/v235/li24bt.html.

Yuhui Li, Fangyun Wei, Chao Zhang, and Hongyang Zhang. EAGLE-3: Scaling up inference acceleration of large language models via training-time test. In *Advances in Neural Information Processing Systems*, volume 38, pp. 136737–136756. Curran Associates, Inc., 2025. doi: 10.52202/085713-4562.

Fengbei Liu, Rachit Saluja, Sunwoo Kwak, Ruibo Wang, Ruining Deng, Heejong Kim, Johannes C. Paetzold, and Mert R. Sabuncu. MAdam: Metric-aware multi-objective Adam. *arXiv preprint arXiv:2606.03904*, 2026a. URL https://arxiv.org/abs/2606.03904v1.

Hongyi Liu, Jiaji Huang, Zhen Jia, Youngsuk Park, and Yu-Xiang Wang. Not-a-bandit: Provably no-regret drafter selection in speculative decoding for LLMs. In *International Conference on Learning Representations*, 2026b. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/20563b8508ba42e1b688d922e926ee26-Abstract-Conference.html.

Runjing Liu, Jeffrey Regier, Nilesh Tripuraneni, Michael Jordan, and Jon McAuliffe. Rao-blackwellized stochastic gradients for discrete distributions. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4023–4031. PMLR, 2019. URL https://proceedings.mlr.press/v97/liu19c.html.

Xiaoxuan Liu, Lanxiang Hu, Peter Bailis, Alvin Cheung, Zhijie Deng, Ion Stoica, and Hao Zhang. Online speculative decoding. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 31131–31146. PMLR, 2024. URL https://proceedings.mlr.press/v235/liu24y.html.

Xiaoxuan Liu, Jiaxiang Yu, Jongseok Park, Ion Stoica, and Alvin Cheung. Speculative decoding: Performance or illusion? In *Proceedings of Machine Learning and Systems*, volume 8, 2026c. URL https://proceedings.mlsys.org/paper_files/paper/2026/hash/554e056fe2b6d9fd27ffcd3367ae1267-Abstract-Conference.html.

Haoxuan Luo, Jameson Sandler, and Ferdinando Fioretto. From positionwise confidence to prefix scheduling: Verifier skipping in speculative decoding. *arXiv preprint arXiv:2608.14787*, 2026. URL https://arxiv.org/abs/2608.14787v1.

Saied Mahdian and Peter W. Glynn. Linear algebraic truncation algorithm with a posteriori error bounds for computing markov chain equilibrium gradients. arXiv preprint arXiv:2501.06266, 2025. URL https://arxiv.org/abs/2501.06266.

Itzik Malkiel and Lior Wolf. MTAdam: Automatic balancing of multiple training loss terms. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 10713–10729, 2021. doi: 10.18653/v1/2021.emnlp-main.837. URL https://aclanthology.org/2021.emnlp-main.837/.

Xupeng Miao, Gabriele Oliaro, Zhihao Zhang, Xinhao Cheng, Zeyu Wang, Zhengxin Zhang, Rae Ying Yee Wong, Alan Zhu, Lijie Yang, Xiaoxiang Shi, Chunan Shi, Zhuoming Chen, Daiyaan Arfeen, Reyna Abhyankar, and Zhihao Jia. SpecInfer: Accelerating large language model serving with tree-based speculative inference and verification. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3*, pp. 932–949, 2024. doi: 10.1145/3620666.3651335. URL https://arxiv.org/abs/2305.09781v4.

Alfred Müller. How does the value function of a Markov decision process depend on the transition probabilities? *Mathematics of Operations Research*, 22(4):872–885, 1997. doi: 10.1287/moor.22.4.872.

Harikrishna Narasimhan, Wittawat Jitkrittum, Ankit Singh Rawat, Seungyeon Kim, Neha Gupta, Aditya Krishna Menon, and Sanjiv Kumar. Faster cascades via speculative decoding. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/6f43166f50f26e8d8f3edc5545b0749f-Abstract-Conference.html.

Arnab Nilim and Laurent El Ghaoui. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.

Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 1994. doi: 10.1002/9780470316887.

Qwen Team. Qwen3.8-27B: Model card. Hugging Face model repository, 2026a. URL https://huggingface.co/Qwen/Qwen3.8-27B. Accessed September 16, 2026; exact experimental revision in the artifact.

Qwen Team. Qwen3.5-0.8B: Model card. Hugging Face model repository, fixed revision, 2026b. Fixed checkpoint revision; accessed September 14, 2026.

Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4334–4343. PMLR, 2018. URL https://proceedings.mlr.press/v80/ren18a.html.

Alexander Samarin, Sergei Krutikov, Anton Shevtsov, Sergei Skvortsov, Filipp Fisin, and Alexander Golubev. LK losses: Direct acceptance rate optimization for speculative decoding. In *International Conference on Machine Learning (ICML)*, 2026. URL https://arxiv.org/abs/2602.23881.

John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 1889–1897. PMLR, 2015. URL https://proceedings.mlr.press/v37/schulman15.html.

Sunder Sethuraman and S. R. S. Varadhan. A martingale proof of Dobrushin's theorem for non-homogeneous markov chains. arXiv preprint arXiv:math/0404231v1, 2004. URL https://arxiv.org/abs/math/0404231v1.

Aaryam Sharma. When is a draft accepted? A theory of acceptance in speculative decoding, 2026. URL https://arxiv.org/abs/2606.30265v1.

Mitchell Stern, Noam Shazeer, and Jakob Uszkoreit. Blockwise parallel decoding for deep autoregressive models. In *Advances in Neural Information Processing Systems*, volume 31, pp. 10086–10095. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/c4127b9194fe8562c64dc0f5bf2c93bc-Abstract.html.

Ziteng Sun, Uri Mendlovic, Yaniv Leviathan, Asaf Aharoni, Jae Hun Ro, Ahmad Beirami, and Ananda Theertha Suresh. Block verification accelerates speculative decoding. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=frsg32u0rO.

Richard S. Sutton, David A. McAllester, Satinder P. Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems*, volume 12, 1999. URL https://proceedings.neurips.cc/paper/1999/hash/464d828b85b0bed98e80ade0a5c43b0f-Abstract.html.

Philip Thomas, Georgios Theocharous, and Mohammad Ghavamzadeh. High confidence policy improvement. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2380–2388. PMLR, 2015. URL https://proceedings.mlr.press/v37/thomas15.html.

Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1:1–305, 2008. URL https://people.eecs.berkeley.edu/~jordan/tutorials.html.

Ephrem Wu. Ceiling-clipped acceptance histograms indicate stranded speed-up in block-diffusion speculative decoding. arXiv preprint arXiv:2608.30427v1, 2026. URL https://arxiv.org/abs/2608.30427v1.

Tianyu Wu, Yu Yao, Zhenting Qi, Han Zheng, Zhuohan Wang, Haoran Ma, Lawrence Liao, Himabindu Lakkaraju, Ju Li, and Yilun Du. D-PACE: Dynamic position-aware cross-entropy for parallel speculative drafting, 2026. URL https://arxiv.org/abs/2605.18810v1.

Tianhua Xia, Mugilan Ganesan, Yifei Feng, Haiyu Wang, Maximilian Egger, and Sai Qian Zhang. Acceptance-aware draft model training for speculative decoding, 2026. URL https://arxiv.org/abs/2609.24150v1.

Xiaomi LLM Core Team. MiMo-7B-Base: Model card. Hugging Face model repository, 2025. URL https://huggingface.co/XiaomiMiMo/MiMo-7B-Base. Accessed September 16, 2026; exact experimental revision in the artifact.

Xiaomi MiMo. MiMo-V2.5-Pro: Model card. Hugging Face model card, 2026. URL https://huggingface.co/XiaomiMiMo/MiMo-V2.5-Pro. Accessed September 19, 2026.

Jiamin Xu and Kyra Gan. Fast non-episodic finite-horizon RL with K-step lookahead thresholding. arXiv preprint arXiv:2602.00781v1, 2026. URL https://arxiv.org/abs/2602.00781.

Minghao Yan, Saurabh Agarwal, and Shivaram Venkataraman. Decoding speculative decoding. In *Proceedings of NAACL-HLT 2025*, 2025.

Tianjian Yang and Meng Li. Spec-AUF: Accept-until-fail training under train-inference misalignment for masked block drafters, 2026. URL https://arxiv.org/abs/2607.01893v1.

Ming Yin, Minshuo Chen, Kaixuan Huang, and Mengdi Wang. A theoretical perspective for speculative decoding algorithm. In *Advances in Neural Information Processing Systems*, volume 37, pp. 128082–128117, 2024. doi: 10.52202/079017-4067.

Jiebin Zhang, Zhenghan Yu, Liang Wang, Nan Yang, Eugene J. Yu, Zheng Li, Yifan Song, Dawei Zhu, Xingxing Zhang, Furu Wei, and Sujian Li. Learning to draft: Adaptive speculative decoding with reinforcement learning. *arXiv preprint arXiv:2603.01639*, 2026.

Lefan Zhang, Xiaodan Wang, Yanhua Huang, and Ruiwen Xu. Learning harmonized representations for speculative sampling. In *International Conference on Learning Representations (ICLR)*, pp. 35367–35388, 2025. URL https://openreview.net/forum?id=T9u56s7mbk.

Yongchao Zhou, Kaifeng Lyu, Ankit Singh Rawat, Aditya Krishna Menon, Afshin Rostamizadeh, Sanjiv Kumar, Jean-François Kagy, and Rishabh Agarwal. DistillSpec: Improving speculative decoding via knowledge distillation. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=rsY6J3ZaTF.

Xiandong Zou, Jianshu Li, Jing Huang, and Pan Zhou. Variational speculative decoding: Rethinking draft training from token likelihood to sequence acceptance. In *International Conference on Machine Learning*, 2026. doi: 10.48550/arXiv.2602.05774. URL https://arxiv.org/abs/2602.05774v6. Version compared: arXiv:2602.05774v6.

Appendix roadmap. Appendix A reports the complete ten-seed prospective comparison, and Appendix B the risk-based head selection on the same prompt cohort. The proofs follow: same-rank separation (C); prefix survival, free-field risk and fixed call costs (D), including prefix nesting and its sharp bound (D.5) and conditional cap safety (D.6); shared reachability and cap control (E); one-path realizability (F); the finite-budget risk of one shared update (G), with finite common steps (G.1); and the local-information boundary (H). Training objectives and actual updates (I), finite continuation information (J), actual-update bounds (K), heterogeneous query windows (L) and paired actual updates with optimizer-state interventions (M) come next. The development experiments cover the earlier training controls (N), all training objectives (O), the all-anchor control (P) and further backbones (Q). Signed endpoint accounting for fitted heads closes the appendix (R).

A Prospective comparison across ten training seeds

Question and design. Does the whole-response objective reduce verification work when compared with a local objective that uses the same active anchors? We prospectively evaluated the fixed paired training recipe on a separately registered cohort of 128 GSM8K and 128 MBPP prompts. The design comprises two frozen backbones, two fixed learning rates, two objectives, and ten training seeds, giving 80 heads. Seeds zero through two reuse the earlier fitted heads without modification; the remaining seven seeds contribute 56 new fits. All 80 heads are evaluated on the new cohort, separately from the earlier three-seed development results. No learning rate, head, or seed is selected using these evaluation outcomes. The prompts are new to this experiment under its registered exclusion rule, not new corpora or verified exclusions from backbone pretraining. The exact registered backbones are HuggingFaceTB/SmolLM3-3B and allenai/OLMo-2-0425-1B; these links bind the corresponding revision-specific model cards.

What is measured. We use cached greedy target responses of at most 64 tokens, a maximum proposal length of four, and maximal-prefix commitment. The primary endpoint is the number of verifier calls per token, including prefill, averaged over 32 sampled-proposal replays per head and prompt. The two objectives use common random numbers. The analytic expectation of this sampler and deterministic argmax proposals are reported separately. These are exact offline executions of the specified proposal and commitment rules, not measured serving latency. Repeated proposal draws are not treated as independent prompts or training seeds.

Inference. Each of the eight backbone, corpus, and learning-rate contrasts compares whole-response minus local calls per token. The registered primary analysis models training seeds and prompts as crossed random factors, uses the two-way analysis-of-variance estimate of the mean's variance and a Satterthwaite t approximation, and applies two-sided Holm correction to the eight contrasts at level 0.05. Tables 3–5 report the same analysis for every endpoint, with approximate familywise 95% Bonferroni intervals. Multiplicity is controlled within each endpoint, not jointly across all three; only sampled calls determine the primary conclusion. The seed-sign counts describe consistency but are not an additional significance test.

Result and scope. For sampled proposals, local training requires fewer calls in four of eight contrasts, all four belonging to SmolLM3-3B. The relative increase from local to whole-response training in these contrasts is 0.423% to 0.894%, and all ten seeds have the same direction. None of the four OLMo-2-1B contrasts separates after correction, and no contrast favors whole-response training. The analytic endpoint gives the same four-versus-zero split; under argmax proposals neither objective has a separated advantage in any of the eight contrasts. Bonferroni gives the same counts as Holm for all three endpoints. Thus the prospective result supports a local-training advantage in a subset of the sampled-proposal comparisons, not a universal optimization advantage, an advantage for every decoding rule, or equivalence in the unresolved contrasts. The results concern trained readout heads on two frozen backbones, not end-to-end backbone training. V100 training and evaluation form the preregistered primary route. The separate L40S training replication is reported below; it shares the three historical seeds and is not ten additional independent seeds.

Secondary inference and numerical sensitivity. Both registered bootstrap procedures use 20,000 resamples. The seed-and-prompt bootstrap retains the primary four-versus-zero split for sampled and analytic calls and finds no argmax separation. The prompt-only bootstrap, which conditions on

Corpus	Rate	Local	Whole	$10^3\Delta$	95% CI for $10^3\Delta$	Rel. (%)	p_{Holm}	$+/-/0$
SmolLM3-3B								
GSM8K	3×10^{-5}	0.6249	0.6277	+2.788	$[+1.613, +3.962]$	+0.446	2.8×10^{-8}	10/0/0
GSM8K	10^{-4}	0.6148	0.6174	+2.603	$[+1.216, +3.989]$	+0.423	6.0×10^{-6}	10/0/0
MBPP	3×10^{-5}	0.5726	0.5777	+5.121	$[+3.376, +6.865]$	+0.894	3.6×10^{-10}	10/0/0
MBPP	10^{-4}	0.5607	0.5650	+4.278	$[+2.255, +6.302]$	+0.763	3.7×10^{-7}	10/0/0
OLMo-2-1B								
GSM8K	3×10^{-5}	0.5706	0.5713	+0.680	$[-0.620, +1.980]$	+0.119	0.418	7/3/0
GSM8K	10^{-4}	0.5589	0.5582	-0.740	$[-2.530, +1.050]$	-0.132	0.488	3/7/0
MBPP	3×10^{-5}	0.5993	0.6004	+1.125	$[-0.237, +2.488]$	+0.188	0.087	9/1/0
MBPP	10^{-4}	0.5865	0.5866	+0.144	$[-1.862, +2.149]$	+0.025	0.840	4/6/0

Table 3: Sampled proposals on the prospective cohort. Each row uses ten seeds and 128 prompts. Local and Whole are verifier calls per token. Δ = Whole − Local, so positive values favor the all-anchor local objective. The difference and its approximate familywise 95% Bonferroni interval are multiplied by 10^3. Relative change is $100\Delta/\text{Local}$. Holm correction uses all eight contrasts within this endpoint. The last column counts positive, negative, and zero per-seed differences; it is descriptive.

Corpus	Rate	Local	Whole	$10^3\Delta$	95% CI for $10^3\Delta$	Rel. (%)	p_{Holm}	$+/-/0$
SmolLM3-3B								
GSM8K	3×10^{-5}	0.6251	0.6279	+2.788	$[+1.599, +3.976]$	+0.446	3.4×10^{-8}	10/0/0
GSM8K	10^{-4}	0.6149	0.6175	+2.593	$[+1.205, +3.982]$	+0.422	5.5×10^{-6}	10/0/0
MBPP	3×10^{-5}	0.5724	0.5777	+5.222	$[+3.421, +7.023]$	+0.912	2.0×10^{-9}	10/0/0
MBPP	10^{-4}	0.5606	0.5649	+4.353	$[+2.367, +6.339]$	+0.776	1.3×10^{-7}	10/0/0
OLMo-2-1B								
GSM8K	3×10^{-5}	0.5707	0.5712	+0.497	$[-0.835, +1.830]$	+0.087	0.862	6/4/0
GSM8K	10^{-4}	0.5588	0.5582	-0.604	$[-2.489, +1.281]$	-0.108	0.862	3/7/0
MBPP	3×10^{-5}	0.5995	0.6006	+1.100	$[-0.245, +2.445]$	+0.183	0.092	9/1/0
MBPP	10^{-4}	0.5865	0.5864	-0.091	$[-2.093, +1.911]$	-0.016	0.898	4/6/0

Table 4: Analytic expected calls on the prospective cohort. Each row uses ten seeds and 128 prompts. Local and Whole are verifier calls per token. Δ = Whole − Local, so positive values favor the all-anchor local objective. The difference and its approximate familywise 95% Bonferroni interval are multiplied by 10^3. Relative change is $100\Delta/\text{Local}$. Holm correction uses all eight contrasts within this endpoint. The last column counts positive, negative, and zero per-seed differences; it is descriptive.

the ten fitted seeds, yields five local advantages for sampled and analytic calls and one for argmax; these are secondary results and do not replace the primary analysis. Any zero bootstrap p value in the machine-readable results is a finite-resample estimate, not a claim of an exactly zero tail probability. Independent recomputation reproduces all primary Holm decisions, with unadjusted p discrepancies below 4.45×10^{-16} and primary interval-end discrepancies below 3.47×10^{-18}. Its first full numerical comparison did not pass: the bootstrap sign-count statistic is discontinuous at zero, and changing floating-point summation order moved near-zero resampled differences across zero. Consequently, secondary unadjusted and Holm-adjusted p values differed by as much as 0.0041 and 0.0048, respectively, although bootstrap interval ends differed by at most 3.30×10^{-17}. Replaying the original summation order reproduced all 48 original secondary p values; every sign disagreement lay within the independently checked roundoff envelope, and none changed a Holm decision. We retain the original values and the failed numerical comparison, disclose the explanation rather than relax its tolerance, and provide all 48 paired secondary results and seed-level effects in the accompanying machine-readable tables. Table 6 summarizes the sensitivity.

Hardware replication. Repeating the seven new training seeds on L40S, while retaining the same three historical seeds and evaluating all heads on V100, gives the same primary counts: four local advantages and no whole-response advantage under sampled proposals and analytic expectation, and

Corpus	Rate	Local	Whole	$10^3\Delta$	95% CI for $10^3\Delta$	Rel. (%)	p_{Holm}	$+/-/0$
SmolLM3-3B								
GSM8K	3×10^{-5}	0.5919	0.5922	+0.305	$[-1.790, +2.400]$	+0.052	1.000	5/5/0
GSM8K	10^{-4}	0.5839	0.5851	+1.245	$[-1.643, +4.133]$	+0.213	0.909	7/3/0
MBPP	3×10^{-5}	0.5432	0.5463	+3.040	$[-0.798, +6.877]$	+0.560	0.161	8/2/0
MBPP	10^{-4}	0.5343	0.5363	+1.978	$[-0.698, +4.653]$	+0.370	0.216	9/1/0
OLMo-2-1B								
GSM8K	3×10^{-5}	0.5399	0.5389	-0.928	$[-3.075, +1.219]$	-0.172	0.909	4/6/0
GSM8K	10^{-4}	0.5323	0.5324	+0.078	$[-3.262, +3.418]$	+0.015	1.000	5/5/0
MBPP	3×10^{-5}	0.5608	0.5602	-0.598	$[-2.400, +1.203]$	-0.107	0.981	3/6/1
MBPP	10^{-4}	0.5519	0.5499	-2.014	$[-4.229, +0.201]$	-0.365	0.098	0/10/0

Table 5: Argmax proposals on the prospective cohort. Each row uses ten seeds and 128 prompts. Local and Whole are verifier calls per token. Δ = Whole − Local, so positive values favor the all-anchor local objective. The difference and its approximate familywise 95% Bonferroni interval are multiplied by 10^3. Relative change is $100\Delta/\mathrm{Local}$. Holm correction uses all eight contrasts within this endpoint. The last column counts positive, negative, and zero per-seed differences; it is descriptive.

Bootstrap	Endpoint	Local	Whole	Max. $\lvert\Delta p\rvert$	Max. $\lvert\Delta p_{\mathrm{Holm}}\rvert$
Seed-and-prompt bootstrap	Sampled	4/8	0/8	0.0000	0.0000
Seed-and-prompt bootstrap	Analytic	4/8	0/8	0.0000	0.0000
Seed-and-prompt bootstrap	Argmax	0/8	0/8	0.0002	0.0000
Prompt-only bootstrap	Sampled	5/8	0/8	0.0002	0.0002
Prompt-only bootstrap	Analytic	5/8	0/8	0.0000	0.0000
Prompt-only bootstrap	Argmax	1/8	0/8	0.0041	0.0048

Table 6: Secondary inference and summation-order sensitivity. Counts use the original registered bootstrap results. The final columns compare their original and independently regrouped summations. All six families retain the same Holm decisions. The prompt-only procedure conditions on the ten fitted seeds; it does not include between-seed uncertainty.

no separated argmax contrast. Both secondary procedures also give the same counts as the primary route. We retain this as a hardware replication, not another independent seed cohort or a basis for selecting hardware. The accompanying tables include all 24 endpoint contrasts and all 48 secondary comparisons for this route. Its original independent comparison also failed on secondary sign-count p values: maximum unadjusted and adjusted discrepancies were 0.0059 and 0.0064. Applying the same summation-order diagnosis reproduced all original secondary values and explained every sign disagreement within the same roundoff envelope; no Holm decision changed. We preserve that failed comparison and the original values. Agreement of decision counts does not imply byte-identical training or numerical interchangeability of all statistics.

B IMPROVEMENT-RISK SCREENING FOR HEAD SELECTION

The exact risk certificate answers a robustness question: how much can verification work increase if some draft-match probabilities rise and none falls? It does not rank the original work of two different fitted heads. We tested whether using it to select a head nevertheless reduces work on new prompts. This preregistered comparison found no such reduction.

Selections fixed before evaluation. For each of SmolLM3-3B and OLMo-2-1B, each of GSM8K and MBPP, and each learning rate, the choice set contains six existing heads: all-anchor local and whole-response training at three seeds. There are eight decisions and 24 distinct heads, since both corpora use the same fitted heads. The rates are 3×10^{-5} and 10^{-4}. No head is refitted or selected using the prospective outcomes.

On 128 development prompts per corpus, the primary rule selects the head with the lowest mean $R_{\uparrow}(p)/L$, where L is the response length. The comparison rule minimizes the mean all-anchor local loss $-\sum_a S_a/L$, not token cross-entropy. A secondary rule maximizes the number of paths for which $D_p(a) \geq 0$ at every anchor. Risk ties are broken by local loss; count ties by risk and then local loss. The final tie order is local before whole-response training, then increasing seed. Development risks and signs use exact rational arithmetic on the stored binary64 probabilities. None of the six candidates in any decision has nonnegative D_p at every anchor on all 128 development paths.

Prospective endpoint and uncertainty. All 24 heads are evaluated on the same 128 new prompts per corpus, shared with the ten-seed study in Appendix A. The selections were frozen before this shared cohort was evaluated; these two analyses are not confirmations on independent prompt cohorts. Each path uses 32 independent-proposal replicates with common uniforms across heads, draft depth four, and maximal-prefix commitment. The primary endpoint is sampled verifier calls per response token, including prefill. Intervals use 20,000 paired prompt-bootstrap draws and the preregistered Bonferroni quantiles 0.003125 and 0.996875 over the eight comparisons within each endpoint. They are conditional on the selected fitted heads: they resample neither training seeds nor the development selection.

All eight primary comparisons. Risk selection changes six choices, but none of their adjusted sampled-work intervals excludes zero (Table 7). The remaining two choices are identical and therefore have exactly zero differences. The analytic expected-work comparison has the same conclusion. This establishes neither superiority nor equivalence. A certificate against harmful coordinatewise improvements is not, on this evidence, a rule for choosing the head with lower current work.

Model	Corpus	Rate	Loss choice	Risk choice	$10^3\Delta$	Adjusted interval
SmolLM3-3B	GSM8K	3×10^{-5}	Local 0	Local 2	$+1.091$	$[-4.507, +6.905]$
SmolLM3-3B	GSM8K	10^{-4}	Local 1	Local 0	-0.965	$[-5.993, +4.284]$
SmolLM3-3B	MBPP	3×10^{-5}	Local 2	Local 0	$+0.820$	$[-4.623, +6.496]$
SmolLM3-3B	MBPP	10^{-4}	Local 2	Local 0	$+2.598$	$[-2.558, +7.912]$
OLMo-2-1B	GSM8K	3×10^{-5}	Local 1	Local 1	$+0.000$	$[+0.000, +0.000]$
OLMo-2-1B	GSM8K	10^{-4}	Local 2	Local 0	$+0.143$	$[-4.468, +4.845]$
OLMo-2-1B	MBPP	3×10^{-5}	Local 2	Local 0	-2.899	$[-9.024, +2.813]$
OLMo-2-1B	MBPP	10^{-4}	Local 0	Local 0	$+0.000$	$[+0.000, +0.000]$

Table 7: Primary comparison of risk selection with local-loss selection. Local and Whole identify the training objective; the following number is the fixed training seed. Positive Δ means the risk-selected head uses more sampled verifier calls per token. Differences and adjusted intervals are multiplied by 10^3. Exact zero rows select the same head; other intervals containing zero indicate unresolved differences.

Complete secondary comparisons. The safe-path-count rule changes four choices. It reduces sampled work in none and increases it for SmolLM3 on MBPP at the higher rate: 0.008320 calls per token, interval $[0.002579, 0.014065]$ (Table 8). Expected work agrees: 0.007774 with interval $[0.002141, 0.013213]$ in this cell. On new-prompt improvement risk, the primary rule separates no comparison; the count rule increases risk for OLMo on GSM8K at the higher rate by 0.005781 per

token, interval $[0.003643, 0.008045]$. The descriptive argmax comparisons separate no primary-rule choices. These results retain every cell and do not replace sampled proposals with argmax. The executable supplement contains all 56 comparisons across seven endpoints, including the zero rows and all secondary risk intervals.

Model	Corpus	Rate	Loss choice	Count choice	$10^3\Delta$	Adjusted interval
SmolLM3-3B	GSM8K	3×10^{-5}	Local 0	Local 0	$+0.000$	$[+0.000, +0.000]$
SmolLM3-3B	GSM8K	10^{-4}	Local 1	Local 0	-0.965	$[-5.993, +4.284]$
SmolLM3-3B	MBPP	3×10^{-5}	Local 2	Local 1	$+3.105$	$[-2.750, +8.913]$
SmolLM3-3B	MBPP	10^{-4}	Local 2	Whole 1	$+8.320$	$[+2.579, +14.065]$
OLMo-2-1B	GSM8K	3×10^{-5}	Local 1	Local 1	$+0.000$	$[+0.000, +0.000]$
OLMo-2-1B	GSM8K	10^{-4}	Local 2	Whole 0	-0.829	$[-6.054, +4.301]$
OLMo-2-1B	MBPP	3×10^{-5}	Local 2	Local 2	$+0.000$	$[+0.000, +0.000]$
OLMo-2-1B	MBPP	10^{-4}	Local 0	Local 0	$+0.000$	$[+0.000, +0.000]$

Table 8: Secondary comparison of safe-path-count selection with local-loss selection. Conventions and the eight-comparison adjustment are as in Table 7. One interval excludes zero unfavorably; no interval supports a reduction in sampled verification work.

Independent reconstruction and execution provenance. A separately written implementation reconstructs all 56 contrasts and all development choices. It uses forward occupancy for expected work and explicit row-corner enumeration for risk. Every choice and interval decision agrees; the largest per-prompt discrepancy is 1.12×10^{-15} after upward rounding, within the fixed 10^{-9} tolerance. All 48 head-by-corpus Monte Carlo checks are within four estimated standard errors of analytic work. The original evaluation ran in a debugging queue. An unchanged replay then ran on a normal-queue V100: all 196,608 sampled call counts, 6,144 argmax counts, analytic calls and active log probabilities agree exactly. This replay preserves the original provenance and does not create a second untouched cohort. An earlier normal-queue attempt failed on output-directory validation before evaluating any head; only the equivalent filesystem-path check was repaired.

The anonymous supplement supplies numeric probability fields and raw call counts, but no text, token identifiers, hidden states or weights. Clean extraction reconstructs all 56 contrasts and the choices from the published development summaries. This export does not reconstruct exact-rational development geometry, retrain a model or demonstrate serving latency.

C LOCAL DOMINANCE AND WHOLE-RESPONSE WORK WITHIN ONE EXACT-RANK STRATUM

This appendix proves Theorem 1. The two policies have the same target, output length, number of proposals, factor rank, parameter bill, and unit cost per target call. More work is therefore not explained by a larger or more expensive drafter: it results from the different anchors visited after accepting drafts. The construction uses a three-dimensional frozen representation, not measured Qwen states. The deterministic argmax counterexample and the differentiable sampled-policy expectation are distinct statements.

C.1 THE FINITE-CAP LOOP AND A PERTURBATION LEMMA

The verification contract. The current target token is committed before each call; the call adds L matching drafts and exactly one target token, advancing $L + 1$ positions. The same-rank, local-information and safe-region results, and all whole-response training comparisons, use this same convention. There is no additional in-call token beyond the $L + 1$ advance.

Fix a target-greedy path of $N = M + 1$ emitted tokens. After the prefill call, the first target bonus is already emitted and the current anchor is $a = 0$. At anchor a, the representation follows exactly a generated tokens and the current bonus has index a. Head $j \in \{1, \dots, K\}$ proposes the token at index $a + j$. Its target-match probability is $p_{a,j}$. In the sampled policy, heads draw independently, conditional on this fixed anchor. The verifier commits the matching prefix and one target bonus. The exact finite-cap convention is

$$H_a = \min\{K, M - a - 1\}, \qquad a_{\text{next}} = a + 1 + L_a, \qquad 0 \le L_a \le H_a. \tag{7}$$

The terminal anchor is M; anchor $M - 1$ has no usable draft but still costs one call. All K proposals are offered even in a clipped final round. Thus $\sum_{\text{rounds}}(1 + L_a) = M$ exactly, not merely in expectation. The construction assumes the accepted-prefix representation is independent of the unused draft tail and the target's verification chunking.

Put $A_{a,0} = 1$ and $A_{a,k} = \prod_{j=1}^{k} p_{a,j}$. The transition probabilities for L_a are $A_{a,l}(1 - p_{a,l+1})$ for $0 \le l < H_a$ and A_{a,H_a} for $l = H_a$. Let $V_p(a)$ be expected remaining target calls, excluding the already performed prefill. Finite-horizon backward evaluation gives

$$\begin{aligned} V_p(M) &= 0, \\ V_p(a) &= 1 + \sum_{l=0}^{H_a - 1} A_{a,l}(1 - p_{a,l+1}) V_p(a + l + 1) + A_{a,H_a} V_p(a + H_a + 1), \\ J(p) &= 1 + V_p(0). \end{aligned} \tag{8}$$

The empty sum at $H_a = 0$ is zero. This is exact for the stated independent sampled proposer, not an approximation to an argmax-proposal expectation. It is an ordinary acyclic Bellman recursion, not a new dynamic-programming theorem or a strict contraction assertion in the undiscounted sup norm.

Intuition. Only comparisons that determine the accepted prefix can change a fixed-cost verification trace.

Lemma 1 (Finite-cap probability perturbation)**.** *For two probability fields p, q in the loop above, if $\max_{a,j} |p_{a,j} - q_{a,j}| \le \eta$, then*

$$|J(p) - J(q)| \le M^2 \eta. \tag{9}$$

Proof. Use a common independent uniform variable for each anchor–head coordinate. Until the first differing acceptance decision, both policies visit the same anchors. A round uses L_a successful comparisons and at most one failed comparison to determine its transition. Hence the entire trace needs at most $\sum(1 + L_a) = M$ relevant queries. Every newly queried coordinate has conditional disagreement probability at most η, because its uniform variable is independent of the preceding decisions. A union bound gives trace-disagreement probability at most $M\eta$. The total call count, including prefill, ranges between 1 and $M + 1$, with range at most M. Multiplying these two bounds proves the result. The sampled but ignored tail coordinates do not enter this argument: they affect neither the next anchor nor the declared unit call cost. Proposal-dependent costs or a numerically chunk-dependent target state would require a different argument.

C.2 PROOF OF THE SAME-STRATUM SEPARATION

Proof of Theorem 1. **Binary reference fields.** Let $K \geq 2$, $n \geq 1$, $P = K + 2$, $M = nP$, and $N = M + 1$. At every anchor with $a \bmod P = 1$, set all K OLD match bits to one; set every other bit to zero. NEW changes only the first head at phase-zero anchors from zero to one. Every head's match bit is nondecreasing, so every local prefix survival event is nondecreasing as well. In one period, OLD visits

$$jP \longrightarrow jP + 1 \longrightarrow (j + 1)P,$$

whereas NEW visits

$$jP \longrightarrow jP + 2 \longrightarrow jP + 3 \longrightarrow \cdots \longrightarrow (j + 1)P.$$

For example, $K = 4, n = 1$ gives OLD $0 \to 1 \to 6$ versus NEW $0 \to 2 \to 3 \to 4 \to 5 \to 6$: including prefill, three versus six calls. The two patterns require respectively two and $K + 1$ verification calls. The last period lands exactly at M under Equation (7). Therefore the binary call counts, including the same prefill, are

$$C_{\text{old}} = 1 + 2n, \qquad C_{\text{new}} = 1 + n(K + 1). \tag{10}$$

Strictly interior match fields. Set $\epsilon = 1/(8M^2)$. Replace every zero and one above by ϵ and $1 - \epsilon$. At phase zero, additionally give head two probability $\epsilon/(2 - \epsilon)$ in both policies. All probabilities are strictly between zero and one and within ϵ of their binary reference. The only OLD–NEW difference remains the first head at phase zero, increasing from ϵ to $1 - \epsilon$. Thus every local prefix survival probability is nondecreasing. Lemma 1 implies

$$\begin{aligned} &|J_{\text{old}} - C_{\text{old}}| \leq \tfrac{1}{8}, && |J_{\text{new}} - C_{\text{new}}| \leq \tfrac{1}{8}, \\ &J_{\text{new}} - J_{\text{old}} \geq n(K - 1) - \tfrac{1}{4} > 0, && \frac{J_{\text{new}}}{J_{\text{old}}} \geq \frac{1 + n(K + 1) - 1/8}{1 + 2n + 1/8}. \end{aligned} \tag{11}$$

In particular $n = 1$ gives the lower ratio $(8K + 15)/25$. The interior probability floor shrinks with M; no fixed-ϵ asymptotic claim is made.

The same exact-rank-two realization. Take three-dimensional row representations $h_a = e_0^\top$ at phase zero, $e_1^\top$ at phase one, and $e_2^\top$ otherwise. A frozen representation map may use prefix position modulo P and ignore token identities; this defines the same state at each accepted prefix regardless of draft tails. These are three representation types, not a claim that the complete position-dependent target is a three-state Markov machine. With $c = (1, 1, 1)^\top$ and fixed binary unembedding $W = [0, c]$, the target always selects token one, since its logit difference is one.

Define

$$\begin{aligned} &\lambda = \log \frac{1 - \epsilon}{\epsilon}, && \delta = -\log 2, && v = c/3, \\ &u = (-\lambda, +\lambda, -\lambda)^\top - c, && \Omega = [u, e_0], && \Gamma_t = \begin{bmatrix} 1 \\ \beta_t \end{bmatrix} v^\top, \\ &R_t = \Omega\Gamma_t, && z_t(h) = hW + (hR_t)W. \end{aligned} \tag{12}$$

Set $\beta_2 = \delta$ and $\beta_t = 0$ for $t \geq 3$ in both policies. OLD has $\beta_1 = 0$ and NEW has $\beta_1 = 2\lambda$. Since $v^\top c = 1$, the three logit differences of head t are the coordinates of $(-\lambda, +\lambda, -\lambda)^\top + \beta_t e_0$. These produce exactly the interior fields above, including $\sigma(-\lambda - \log 2) = \epsilon/(2 - \epsilon)$.

The two columns of Ω are independent because u has a nonzero third coordinate. The stacked loading $[\Gamma_1 \ \cdots \ \Gamma_K]$ has row rank two because $(1, \beta_1)^\top$ and $(1, \delta)^\top$ are independent whenever $\beta_1 \geq 0$ and $\delta < 0$. Thus both endpoints, and every point along $\beta_1 \in [0, 2\lambda]$, have

$$\operatorname{rank}(\Omega) = \operatorname{rank}([\Gamma_1 \ \cdots \ \Gamma_K]) = \operatorname{rank}([R_1 \ \cdots \ R_K]) = 2.$$

Each individual residual R_t has rank one. The shared factor count is $6(K+1)$ at both endpoints and the exact factor-stratum quotient dimension is $2(3+3K-2) = 6K+2$. These are pre-unembedding factor quantities, with fixed zero biases. They are not a claim that the binary unembedding introduces no observable-law kernel. Nor are the corresponding rank-one isolated and rank-two shared classes capacity matched; this comparison stays within one shared class. For $n = 1$ the largest factor entries are $O(\log K)$, because $\lambda = \log(8(K + 2)^2 - 1)$, so every finite instance uses finite logits.

Argmax endpoints and the order of separation. Every low endpoint probability is below $1/2$, and every high endpoint probability is above $1/2$. Argmax proposals from these same finite-logit heads therefore reproduce the binary fields exactly and obey Equation (10), without perturbation error. This deterministic statement is separate from Equation (11) for independent samples.

For any fixed-K policy in the declared loop, each call advances between one and $K+1$ anchors. Thus, for any two such policies,

$$\frac{\mathbb{E}C_{\text{new}}}{\mathbb{E}C_{\text{old}}} \leq \frac{M+1}{1+\lceil M/(K+1)\rceil} < K+1. \tag{13}$$

Together with the $n=1$ linear lower bound, this yields worst-case order $\Theta(K)$. The upper bound is a general advancement bound, not a claimed optimal constant under local dominance.

C.3 Signed continuation credit and an opposing update direction

The next identities specialize ordinary performance-difference and policy-gradient arguments (Sutton et al., 1999; Schulman et al., 2015). They are not new general sequential-decision results. Let $P_p(a,b)$ be the transition probability under field p, and let $d_p(a)$ be the probability of visiting anchor a. Acyclicity makes $d_p(a)$ also its expected number of visits. Costs below are fixed and independent of the head parameters.

Intuition. A useful accepted prefix is one that also leaves a cheap remaining response, weighted by how often its anchor is reached.

Proposition 1 (Exact finite-horizon continuation coefficients)**.** *For the independent sampled loop and two fields p, q,*

$$J(q) - J(p) = \sum_{a=0}^{M-1} d_q(a) \sum_b \big(P_q(a,b) - P_p(a,b)\big) V_p(b). \tag{14}$$

At an interior field, writing $\Delta_{a,k} = V_p(a+k+1) - V_p(a+k)$, the exact log-probability derivative is

$$g_{a,j} := \frac{\partial J}{\partial \log p_{a,j}} = d_p(a) \sum_{k=j}^{H_a} A_{a,k}\Delta_{a,k}, \qquad j \leq H_a, \tag{15}$$

and is zero for an unused tail coordinate. Therefore any differentiable shared parameterization satisfies $\nabla_\theta J = \sum_{a,j} g_{a,j} \nabla_\theta \log p_{a,j}$.

Proof. Sum $1 + \mathbb{E}_q[V_p(a_{\text{next}}) \mid a] - V_p(a)$ along a q-trace. The value differences telescope from $V_p(0)$ to $V_p(M) = 0$; the common prefill cancels. Substitution of the Bellman equality for p gives Equation (14). Summation by parts in Equation (8) gives

$$V_p(a) = 1 + V_p(a+1) + \sum_{k=1}^{H_a} A_{a,k}\Delta_{a,k}.$$

Coordinate $p_{a,j}$ affects no future row of this acyclic probability field. Its local value derivative propagates to the initial value with coefficient $d_p(a)$. Differentiating each prefix product with respect to $\log p_{a,j}$ therefore yields Equation (15). The formula contains no division by p and has a continuous extension to boundary fields. Finally apply the chain rule over all coordinates affected by θ. Effects of sharing across future anchors are included in that full sum, not discarded by the coordinatewise derivation.

For $n=1$ in the constructed rank-two path, changing β_1 changes only $p_{0,1}$. Since a coordinate is queried at most once, J is affine in this probability with all other coordinates fixed. Its slope is

$$\frac{\partial J}{\partial p_{0,1}} = \frac{J_{\text{new}} - J_{\text{old}}}{1-2\epsilon} > 0, \qquad \frac{dp_{0,1}}{d\beta_1} = p_{0,1}(1-p_{0,1}) > 0. \tag{16}$$

Thus expected work increases throughout this valid same-stratum path, while target CE and negative local expected-prefix length have negative directional derivatives. D-PACE's stopped-gradient

coefficient is likewise negative because its smoothed prefix weights are positive (Wu et al., 2026). The conflict concerns one feasible direction, not antiparallel full parameter-gradient vectors. It does not say that reducing target probability generally improves a real drafter. For comparison, $K = 1$ has $D_a = V_p(a) - V_p(a+1) = 1 - p_{a,1}D_{a+1} \in [0,1]$, starting from $D_{M-1} = 1$; all probability-coordinate derivatives are nonpositive. The separation therefore requires at least two proposed tokens in this unit-cost model.

Mathematical checks and scope. The accompanying implementation checks $K \in \{1,2,3,4,8,16,32,64\}$ and $n \in \{1,2,8\}$, giving 24 cases. It passes 579 checks, including exact Fraction expectations and inequalities, an independent binary counter, 47 exact gradient coordinates, and four performance-difference identities. Its 1,800 small-matrix rank checks have maximum probability reconstruction error 1.11×10^{-16}; the analytical argument above, not a numerical rank tolerance, proves rank for all $K \geq 2$ separation cases; $K = 1$ is the separate monotonicity boundary. As one constructed illustration, $K = 64, n = 8$ gives argmax counts 17 versus 521, and sampled expectations 17.00768688 versus 520.99934538. These are toy-theorem calculations, not observed language-model deployment results. The source bundle retains the complete numerical outputs and their executable definitions.

Nearest-work boundary. SpecDec++ already formulates full-response cost-to-go for adaptive stopping and trains an acceptance predictor (Huang et al., 2025). D-PACE already derives prefix-product weighting with a local continuation term (Wu et al., 2026); here $\Delta_{a,k}$ instead compares the remaining complete response after landing at different anchors. DVI already combines a frozen backbone, low-rank draft-head training, distillation warm-up, and online acceptance feedback (Bhansali & Heck, 2025). None is alleged to claim the universal local-monotonicity statement excluded by this construction. The added separation is same-cost and same-exact-rank, rather than a larger-drafter cost tradeoff. Priority for this exact construction is not established by an exhaustive literature search. Whether its signed credit helps finite-budget fitting on actual target representations is the separate empirical question in Section 4; the theorem alone neither explains fitted optimizer behavior nor guarantees a method gain.

D Prefix Survival, Free-Field Risk and Fixed Call Costs

The unit-call theorem has two distinct extensions. Within-call dependence does not invalidate safety when the ordered quantities are joint prefix survivals. Unequal fixed call costs change the sufficient condition and its sharp uniform threshold. A separate exact risk calculation asks whether any independently adjustable improvement could be harmful; it does not establish that shared neural parameters can realize that improvement.

D.1 Conditional prefixes, not token marginals

Keep the finite path and $1 + L$ transition of Equation (7). There are M transient anchors and absorbing anchor M. Let $c_a > 0$ be a fixed call cost, independent of the proposal field, and let the prefill cost also be fixed. A valid survival row satisfies

$$1 = A_{p,a,0} \geq A_{p,a,1} \geq \cdots \geq A_{p,a,H_a} \geq 0, \qquad A_{p,a,H_a+1} = 0.$$

Here $A_{p,a,k}$ is the probability that the first k drafts all match the fixed target prefix. Then

$$P_p(a, a+1+l) = A_{p,a,l} - A_{p,a,l+1}, \quad 0 \leq l \leq H_a,$$
$$V_p(M) = 0, \qquad V_p(a) = c_a + \sum_{l=0}^{H_a} P_p(a, a+1+l) V_p(a+1+l). \tag{17}$$

The last anchor has $H_a = 0$ and still incurs its call cost. For $M = 0$ only the prefill remains. An anchor must specify the conditional future law and cost: persistent hidden drafter state requires an augmented state. Averaging random target trajectories does not by itself produce this fixed-path model.

With independent parallel proposals, $A_{p,a,k} = \prod_{j=1}^{k} p_{a,j}$ uses individual match probabilities. With autoregressive or correlated proposals, the same product uses match probabilities conditional on all

earlier drafts matching, not their unconditional token marginals. Conditional prefix factorization is already used in prefix scheduling (Luo et al., 2026); our extension concerns the safety condition for the present strict verification rule.

Set $D_p(a) = V_p(a) - V_p(a+1)$, and let $d_q(a)$ be the probability of visiting anchor a under q. Summation by parts and the performance-difference identity give

$$D_p(a) = c_a - \sum_{k=1}^{H_a} A_{p,a,k} D_p(a+k), \tag{18}$$

$$J(q) - J(p) = -\sum_a d_q(a) \sum_{k=1}^{H_a} (A_{q,a,k} - A_{p,a,k}) D_p(a+k). \tag{19}$$

For completeness, subtract the two Bellman equations with V_p held fixed in the changed transition term. Resolving the resulting forward triangular system weights that term by $d_q(a)$. Summation by parts within each row produces Equation (19). This is a specialization of the usual identity, not a new general policy theorem. For unit costs, $\sum_k A_{p,a,k} \le 1$ implies $0 \le D_p(a) \le 1$ by backward induction. Thus $A_q \ge A_p$ implies $J(q) \le J(p)$, without requiring independent proposals or a bound on $\sum_k A_{q,a,k}$.

Why marginal improvement is insufficient. For two proposals, change the joint match distribution over $(11, 10, 01, 00)$ from $(1/2, 0, 0, 1/2)$ to $(1/5, 2/5, 2/5, 0)$. Both marginals rise from $1/2$ to $3/5$, but the second prefix survival falls from $1/2$ to $1/5$. Take $M = 3$, $H = (2, 1, 0)$, zero acceptance at anchor one and unit costs. Continuation work rises from 2 to $11/5$, although the original expected prefix is at most one at every anchor. The failed premise is prefix-survival dominance, not independence. Lossy acceptance, changed output laws, field-dependent costs, changing proposal counts or unresolved native drafter state remain outside this result. It does not map a complete EAGLE or other architecture into the earlier factor-rank construction.

D.2 Exact upward risk with independently adjustable probabilities

Suppose each active conditional probability can independently range over $q_{a,j} \in [p_{a,j}, 1]$. Define

$$R_\uparrow(p) = \max_{q \in [p,1]} J(q) - J(p).$$

This rectangular problem is a relaxation of a shared readout's feasible image. Set $U(M) = 0$ and work backwards over anchors. At an empty row use $U(a) = c_a + U(a+1)$. Otherwise put $F_{H_a+1} = U(a + H_a + 1)$ and, for $j = H_a, \dots, 1$, compute

$$\begin{aligned} F_j &= \max\{(1 - p_{a,j}) U(a+j) + p_{a,j} F_{j+1},\ F_{j+1}\}, \\ U(a) &= c_a + F_1. \end{aligned} \tag{20}$$

Record $q_{a,j} = p_{a,j}$ when the first branch wins, including ties, and $q_{a,j} = 1$ otherwise.

Intuition. Independent row choices let one backward calculation find the most harmful field that lowers no match probability.

Proposition 2 (Exact free-field upward risk)**.** *The recorded single field attains every continuation maximum $U(a)$. Consequently $R_\uparrow(p) = U(0) - V_p(0) \ge 0$. Evaluation and witness replay each require $O(MK)$ arithmetic operations; storing the field and values requires $O(MK + M)$ space.*

Proof. Given that the first $j-1$ drafts match, the next continuation is $(1-q_{a,j})U(a+j) + q_{a,j} F_{j+1}$. An affine function on the allowed interval attains its maximum at an endpoint, proving the inner recursion. Backward induction over anchors is valid because each later row is chosen independently, and one already selected later field attains all required $U(b)$. Acyclicity excludes repeated visits with conflicting choices. The recorded choices therefore form one global maximizing witness. Counting the active entries gives the stated arithmetic and storage bounds.

This is classical rectangular robust dynamic programming (Iyengar, 2005) specialized to an acyclic prefix chain, not a new general algorithm. The arithmetic bound is not a rational bit-complexity

bound or a bound on acquiring model probabilities. Teacher generation, fitting, full-vocabulary evaluation and transfers remain charged. Positive risk proves vulnerability only in the free field, not realizability by the registered shared factors. Zero risk certifies this entire relaxation and therefore also any subset of it, under exactly the same fixed-path cost model.

For interior p, a maximizing coordinate at one generally requires an infinite logit. Replacing each such coordinate by $p_{a,j} + (1-\delta)(1-p_{a,j})$, with $\delta \downarrow 0$, preserves dominance and converges in value by polynomial continuity. Hence the same supremum applies to finite free-field logits, and strictly positive risk has a finite-logit harmful witness. This does not construct shared-rank factors.

D.3 A CONDITIONAL WEIGHTED SAFETY BOUND

Intuition. With unequal fixed call costs, the relevant safety budget weights accepted prefixes by the costs they can skip.

Proposition 3 (Weighted safe baseline)**.** *Assume the fixed positive costs and legal survival model above. If*

$$\sum_{k=1}^{H_a} A_{p,a,k} c_{a+k} \le c_a \quad \textit{at every anchor } a, \tag{21}$$

then $0 \le D_p(a) \le c_a$*, and every survival-dominating* q *satisfies* $J(q) \le J(p)$*. Under the same condition, define* $b_a = c_a - \sum_k A_{p,a,k} c_{a+k} \ge 0$*. The finite improvement obeys*

$$J(p) - J(q) \ge \sum_a d_q(a) \sum_k (A_{q,a,k} - A_{p,a,k}) b_{a+k} \ge 0. \tag{22}$$

Proof. Start at the final anchor and assume $0 \le D_p(b) \le c_b$ at later anchors. Equation (18) then gives $b_a \le D_p(a) \le c_a$. Condition (21) makes the lower bound nonnegative, closing the induction. Substitute these bounds into Equation (19), whose survival differences and visit weights are nonnegative.

The margin is not an unconditional lower bound. For $K = 1$, $M = 4$, costs $(1, 1, 1/4, 1)$ and active probabilities $(1/2, 1/2, 1/2)$, the true differences are $(7/16, 9/8, -1/4, 1)$. The root expression $c_0 - A_{p,0,1} c_1 = 1/2$ exceeds the actual $7/16$: the condition fails at anchor two and $D_p(1) > c_1$ invalidates the substitution. Both the finite-gain conclusion and the substituted margins require Equation (21) at every anchor.

Sharp scalar threshold and its exceptions. If $c_{\min} \le c_a \le c_{\max}$, the uniform condition $S_p(a) = \sum_k A_{p,a,k} \le r = c_{\min}/c_{\max}$ implies Equation (21). For $r < 1$ this scalar threshold is sharp even for $K = 1$. Normalize $c_{\max} = 1$. Given $\gamma > r$, put $s = (r + \min\{\gamma, 1\})/2$, choose $M = 3$, costs $(1, r, 1)$, and set $p_{1,1} = s$, $0 < p_{0,1} < s$. All expected prefixes are below γ, but $D_p(1) = r - s < 0$ and $\partial J/\partial p_{0,1} = s - r > 0$. A sufficiently small finite root increase stays below γ and is harmful. For fixed $K \ge 2$ the same three-anchor path has root cap two. Choose $p_{0,2} = (s - r)/2$ and $p_{0,1} = \min\{s/4, 1/4\}$. Then $S_p(0) < s$ and the harmful root derivative is $(s - r)/2 > 0$. All active probabilities are interior. For $r = 1$, the earlier unit-cost sharpness result applies when $K \ge 2$. Constant costs with $K = 1$ are the exception: $S \le 1$ is automatic and no above-one unsafe family exists.

The weighted condition can certify fields outside the uniform r bound. For example, $K = 1$ and nonincreasing costs imply it automatically. Conversely $S \le 1$ alone need not ensure safety under unequal costs. These are fixed-cost mathematical statements, not latency measurements. Costs depending on a candidate, KV layout, precision, batching or scheduling must be represented explicitly before applying the proposition.

Supporting arithmetic checks. An exact-rational calculation compares the risk recursion with all 4,700 corners of 150 small fields, with zero maximum residual. It also checks the weighted boundary, its strict witnesses, and the correlated-marginal counterexample. A separate exact example checks the failed unconditional margin above. These finite checks support implementation and algebra; they are neither empirical prevalence measurements nor substitutes for the proofs.

D.4 Complete diagnostic on published probability fields

We evaluate the rectangular risk on all 10,752 previously published Qwen3.5-0.8B field records from 42 fitted method cells across the single-anchor, block and all-anchor studies. The first two retain their original selected rates; the last retains both rates. No field is selected by its acceptance or risk. These are opened development outputs, not a new model or prompt sample. The numerical object is the exact rational interpretation of the binary64 probabilities obtained from the saved log probabilities, not unrounded model probabilities. Each path retains its original length and prefill convention.

The threshold certifies 2 records; exact risk is zero on 370, including those 2, and positive on 10,382. The maximum relaxed risk is 0.037134 calls/token. Exact payload comparison identifies 9,984 distinct fields and 39 distinct complete cell payloads: 768 fields occur twice. Prompts recur only within their disjoint cohorts; checkpoints and training histories add dependence. These denominators cannot be interpreted as independent trials or diagnostic accuracy.

We also compare 3,072 paired final-checkpoint records on Qwen3.5-0.8B. The readout with the better scalar local objective is also the costlier one to verify in 879 of them, 28.6 percent: in 812 of the 1,721 records where the local-trained readout is better locally, and in 67 of the 1,351, 5.0 percent, where the whole-response readout is. Work worsens in 976 records overall. The widest of the 67 separates the two quantities by 9.69×10^{-4} in local objective and 1.22×10^{-2} calls per token, on a GSM8K prompt at the third of the three seeds; its largest single coordinate probability changes are $+0.766$ and -0.820. Neither direction of coordinatewise dominance occurs in any pair. Thus the absence of an observed dominance-with-harm pair does not empirically confirm the theorem: its dominance premise is absent. These compare two trained endpoints, not one model immediately before and after an update.

An independent implementation replays all baseline values with maximum floating discrepancy 8.527×10^{-14}, exactly recomputes all 370 zero-risk cases, and verifies attainment and optimality of all 28 serialized worst witnesses. It does not independently reoptimize every positive-risk field. The main readout takes 938.014 seconds and the independent replay 13.573 seconds on CPU; acquiring fields, fitting and deployment are not included. All fields, repeated-field identities, paired differences, group summaries and witnesses accompany the results.

D.5 Prefix Nesting and a Sharp Work-Risk Bound

This subsection retains the fixed target path, unit call cost and maximal-prefix commitment rule of Appendix D. Let $M \geq 1$ be the absorbing anchor, $H_a = \min\{K, M - a - 1\}$ for $0 \leq a < M$, and

$$A_{a,0} = 1 \geq A_{a,1} \geq \cdots \geq A_{a,H_a} \geq 0, \qquad A_{a,H_a+1} = 0.$$

Here $A_{a,k} = \Pr(L_a \geq k)$ is the *joint* probability that the first k proposals match. Correlated proposals require their joint survival, not a product of unconditional marginal probabilities. Write $Y_a = a + 1 + L_a$, $V_A(M) = 0$, $V_A(a) = 1 + \mathbb{E}[V_A(Y_a)]$, and $D_A(a) = V_A(a) - V_A(a+1)$. A fixed prefill cost cancels from every work contrast below. The case $M = 0$ has no subsequent work.

An architectural sufficient condition. *Intuition.* If starting one token later never moves the next verified anchor backwards in distribution, less remaining text cannot require more expected calls.

Proposition 4 (Prefix-nesting safety)**.** *Suppose the baseline survival field satisfies*

$$A_{a,k} \leq A_{a+1,k-1}, \qquad 0 \leq a < M, \quad 2 \leq k \leq H_a. \tag{23}$$

Then $0 \leq D_A(a) \leq 1$ *at every nonterminal anchor. For every valid survival field* $C \geq A$*, the same verification rule satisfies* $J(C) \leq J(A)$*. The improved field* C *need not satisfy* (23)*.*

Proof. At threshold $a+1+k$, the tails of Y_a and Y_{a+1} are respectively $A_{a,k}$ and $A_{a+1,k-1}$. For $k = 1$ the latter is one; comparisons beyond the first row's cap are automatic. Thus Y_a is stochastically no larger than Y_{a+1}. Starting with $V_A(M-1) = 1 \geq V_A(M) = 0$, backward induction gives $\mathbb{E}[V_A(Y_a)] \geq \mathbb{E}[V_A(Y_{a+1})]$, hence $D_A(a) \geq 0$. Equation (18), with unit costs, then gives $D_A(a) \leq 1$. Equation (19) proves the last claim because every changed survival and every baseline D_A are nonnegative.

The argument uses classical stochastic monotonicity and Bellman comparison (Puterman, 1994). For a fixed autoregressive drafter whose conditional match law at a target prefix is independent of the anchor from which that prefix was reached, $A_{a,k} = r_{a+1}A_{a+1,k-1}$, so the condition holds even when r_i varies with position. Within-block dependence alone does not imply this cross-anchor consistency: changing restart features or head indices can change the conditional law.

Nesting is sufficient, not necessary. It is also incomparable with the earlier sufficient condition $\sum_k A_{a,k} \le 1$. With $K = 2, M = 4$, constant individual probabilities $9/10$ satisfy nesting but give $\sum_k A_{a,k} = 171/100$ in each full row. Conversely, individual rows $((1/2, 1/2), (1/16, 1/2), (1/4), ())$ satisfy the sum condition everywhere but violate nesting at $(a, k) = (0, 2)$.

How much risk can a nesting defect create? Define the largest nested field below A by

$$B_{a,k} = \min_{0 \le j < k} A_{a+j,k-j} = \min\{A_{a,k}, B_{a+1,k-1}\}, \qquad B_{a,1} = A_{a,1}. \tag{24}$$

Set $B_{a,0} = 1$ and out-of-cap tails to zero. The recursive expression applies for $k \ge 2$. Let $d_A(a)$ denote the probability of visiting nonterminal anchor a from zero, and define

$$\mathcal{R}_{\mathrm{surv}}(A) = \sup_{\substack{C \ge A \\ C \text{ valid survival field}}} \{J(C) - J(A)\}.$$

This set is larger than the independent conditional-probability box defining $R_\uparrow(p)$: increasing each conditional probability increases every prefix survival, but the converse need not hold. Neither relaxation requires one shared neural update to realize all row changes, and B itself need not be realizable by the shared head. Such rectangular relaxations have established dynamic-programming antecedents (Iyengar, 2005).

Intuition. Lower the inconsistent survivals just enough to recover a safe baseline. The extra work caused by this correction bounds how much any subsequent local improvement could harm the original field.

Theorem 4 (Sharp nested-minorant bound)**.** *The field B in* (24) *is a valid survival field and the greatest nested minorant of A. For independent conditional probabilities p inducing A,*

$$\begin{aligned} 0 \le R_\uparrow(p) \le \mathcal{R}_{\mathrm{surv}}(A) &\le J(B) - J(A) \\ &= \sum_{a=0}^{M-1} d_A(a) \sum_{k=1}^{H_a} (A_{a,k} - B_{a,k}) D_B(a+k) \\ &\le \sum_{a=0}^{M-1} d_A(a) \sum_{k=1}^{H_a} (A_{a,k} - B_{a,k}). \end{aligned} \tag{25}$$

The minorant bound and its displayed sum require $O(MK)$ arithmetic given the complete survival field. Among bounds $J(F) - J(A)$ obtained from nested minorants $F \le A$, B gives the smallest bound. The coefficient one is sharp, including asymptotically among fields with all individual probabilities strictly between zero and one.

Proof. Clearly $0 \le B \le A$. Each diagonal term used for $B_{a,k+1}$ is no larger than the corresponding term for $B_{a,k}$, and the former minimum has an additional term. Thus $B_{a,k+1} \le B_{a,k}$, proving legal row survival. The recurrence gives nesting. If $F \le A$ is nested, repeatedly applying its nesting inequality gives $F_{a,k} \le F_{a+j,k-j} \le A_{a+j,k-j}$ for every $j < k$, so $F \le B$.

Every admissible $C \ge A$ also dominates B. Proposition 4 therefore gives $J(C) \le J(B)$. The performance-difference identity, with B as baseline and A as the changed field, gives the equality in (25). Only nonterminal anchors are summed. Since B is nested, $0 \le D_B \le 1$, proving the last inequality. For any nested $F \le A$, maximality gives $F \le B$ and safety of F gives $J(B) \le J(F)$, proving the restricted optimality claim. The backward envelope, Bellman and forward visitation recursions each use $O(MK)$ operations. Sharpness follows from the family below.

For a direct measure of inconsistency, put $e_{a,k} = (A_{a,k} - A_{a+1,k-1})_+$ for $k \geq 2$. Telescoping along a diagonal and replacing signed increments by positive parts gives

$$A_{a,k} - B_{a,k} \leq \sum_{j=0}^{k-2} e_{a+j,k-j}, \qquad \mathcal{R}_{\mathrm{surv}}(A) \leq \sum_{a=0}^{M-1} d_A(a) \sum_{k=2}^{H_a} \sum_{j=0}^{k-2} e_{a+j,k-j}. \tag{26}$$

The value difference in (25) can be tighter than this defect-only bound. These are structural certificates, not an improvement on exact risk maximization over the same fully observed field. Evaluating a neural model to obtain that field remains a separate cost. The displayed backward recurrence attains the stated arithmetic complexity; the direct diagonal-minimum reference implementation takes $O(MK^2)$.

Sharpness and interior approximation. Take $K = 2, M = 4$ and survival rows $A = ((0,0),(1,w+e),(w),())$, where $0 \leq w < w+e \leq 1$. Only $B_{1,2} = w$ changes. Remaining work is $V_A(0) = 3 - w - e$ and $V_B(0) = 3 - w$. Raising the first individual probability at anchor zero to one while leaving its second zero gives remaining work $3 - w$. The common prefill cancels. Hence both risks and every bound in (25)–(26) equal e. For $0 < w < w + e < 1$, instead use individual rows $((\epsilon,\epsilon),(1-\epsilon,w+e),(w),())$. For sufficiently small $\epsilon > 0$, the bound is $(1-\epsilon)(e-\epsilon(w+e))$. Raising only the first entry to $1-\epsilon$ increases work by $(1-2\epsilon)(e-\epsilon(2+e-w))$. Their ratio tends to one, with every active individual probability and its updated value in $(0,1)$.

Upward risk is not posterior commitment regret. For individual rows $((1/2,1/2),(4/5,1/2),(1/5),())$, maximal-prefix commitment gives $V = (129/50, 44/25, 9/5, 1, 0)$ and $R_\uparrow = 0$. Writing the five individual entries as x, y, u, v, w, its work polynomial is $4 - w - x(1 + y(1-w)) - (1-x)u(1-w+v)$; over the upward probability box its maximum is attained at the original five values. After observing a realized prefix, however, a controller may discard a matched token. Its recurrence is $W(a) = 1 + \mathbb{E}[\min_{a+1 \leq b \leq a+1+L_a} W(b)]$. When $L_0 = 1$, advancing to anchor one is cheaper than advancing to two, since $44/25 < 9/5$. This yields $W(0) = 257/100$, strictly below $129/50$. The two risks optimize different choices; allowing posterior truncation changes the operational question.

Complete fixed-field calculation. We apply the nested minorant and four commitment rules to all 10,752 previously published fields, retaining the same 42 fitted configurations and 512 actual prompts in two disjoint cohorts. Only one field satisfies exact nesting. Conditional-coordinate risk is zero for 370 fields, whereas the larger survival-order risk is zero for 244. Thus nesting itself is rare here; its quantitative bound remains applicable to the other fields. All rows satisfy the exact risk ordering in (25). Exact arithmetic refers to the rational representation of the saved binary64 probabilities, not to unrounded neural laws.

Table 9: Complete-field descriptive means, in verification calls, not latency. Risks cancel the common prefill. Policies know the fixed target path and its entire probability field. Repeated prompts and fitted heads make these dependent descriptive records, not independent trials.

Quantity	Mean over 10,752 fields
Conditional-coordinate risk $R_\uparrow$	0.234508
Survival-order risk $\mathcal{R}_{\mathrm{surv}}$	0.393501
Nested-minorant work bound $J(B) - J(A)$	0.636830
Visit-weighted deletion bound	1.252141
Adjacent-defect bound	2.523133
Default maximum-prefix calls	33.290543
Optimal posterior-commitment calls	33.066546
Fixed greedy-commitment calls	33.091435
Optimal pre-draw-cap calls	33.134603

The posterior policy reduces mean work by 0.223997 calls, while the ordinary pre-draw-cap control reduces it by 0.155939. These are different information and action sets, not attainable speedups from a learned controller. The greedy rule, which maximizes committed length plus the next anchor's expected accepted length, has lower mean work but can increase work: its signed change

has 95th nearest-rank quantile $+0.005584$ and maximum $+0.404244$ calls. Mean proposal slots are 123.914166 for the default, 123.130634 for posterior commitment, 123.208012 for the greedy rule and 112.972035 for the pre-draw cap. Proposal slots do not measure vocabulary-projection cost. The full tables retain all fields, per-token values and distributions; zero-risk ratio denominators remain undefined. Among the 10,382 positive conditional-risk fields, the per-field bound-to-risk ratio has median 3.53960, 95th quantile 390.29135, mean 4.92495×10^{20} and maximum 5.11308×10^{24}. The 370 zero-risk fields retain undefined ratios. These are statistics of individual ratios, not ratios of group means. Very small positive risks make the mean enormous: worst-case sharpness does not make this bound a uniformly accurate risk estimate on fitted heads. Independent reconstruction checks every exact-value digest and the complete grouped summaries. Primary and independent calculations take 287.149 and 290.048 seconds respectively, excluding field acquisition, training and native serving. This calculation measures the looseness of a structural bound and idealized policy headroom, not shared-head reachability or optimization benefit.

D.6 Conditional Safety and the Position of Cap Effects

The exact risk recursion also admits a sign characterization. This is policy improvement on an expanded prefix chain, not a new general local-to-global optimization principle (Agarwal et al., 2021; Bhandari & Russo, 2024). Its specialization locates the possible disagreement between conditional and survival safety under a fixed proposal cap.

Retain the fixed target path, maximal-prefix commitment and fixed positive costs c_a of Appendix D. For now allow any fixed $0 \le H_a \le M - a - 1$. Write $A_{a,j} = \prod_{\ell=1}^{j} p_{a,\ell}$, with $A_{a,0} = 1$ and $A_{a,H_a+1} = 0$. The factors are probabilities conditional on the earlier drafts matching. For dependent drafts they are not unconditional marginals. Let V_p, D_p and d_p be the remaining work, adjacent value difference and nonterminal visitation probability already defined. Total work is $J(p) = c_{\mathrm{prefill}} + V_p(0)$; fixed prefill cancels from every contrast below. Neither law-dependent costs nor a changed verifier is included.

The exact sign of an individual change. Define the suffix credit

$$\mathcal{C}_{a,j} = \sum_{k=j}^{H_a} \left(\prod_{\ell=j+1}^{k} p_{a,\ell} \right) D_p(a+k), \qquad 1 \le j \le H_a. \tag{27}$$

The empty product is one. Hence $\mathcal{C}_{a,H_a} = D_p(a+H_a)$ and, for $j < H_a$, $\mathcal{C}_{a,j} = D_p(a+j) + p_{a,j+1}\mathcal{C}_{a,j+1}$. Define baseline suffix values by

$$F_{H_a+1} = V_p(a+H_a+1), \qquad F_j = (1-p_{a,j})V_p(a+j) + p_{a,j}F_{j+1}.$$

Backward induction gives $\mathcal{C}_{a,j} = V_p(a+j) - F_{j+1}$. A row cannot affect the probability of reaching itself, or any later baseline continuation, because the chain is acyclic. Therefore, with every other coordinate fixed and $0 \le t \le 1 - p_{a,j}$,

$$\begin{aligned} \frac{\partial J}{\partial p_{a,j}} &= -d_p(a) A_{a,j-1} \mathcal{C}_{a,j}, \\ J(p + te_{a,j}) - J(p) &= -t\, d_p(a) A_{a,j-1} \mathcal{C}_{a,j}. \end{aligned} \tag{28}$$

The second identity is exact, not a first-order approximation. A negative credit gives a finite harmful increase when visitation, prefix probability and upward slack are positive. It is a free-coordinate witness, not a construction of a shared neural update.

Intuition. If no individual conditional increase can raise work, the sequential structure prevents several such increases from causing hidden harm.

Proposition 5 (Conditional safety from suffix credits)**.** *Suppose $0 < p_{a,j} < 1$ at every active coordinate. Then*

$$\begin{aligned} R_{\uparrow}(p) = 0 \quad &\iff \quad \frac{\partial J}{\partial p_{a,j}} \le 0 \text{ for all active } (a,j) \\ &\iff \quad \mathcal{C}_{a,j} \ge 0 \text{ for all active } (a,j). \end{aligned} \tag{29}$$

This holds also for positive unequal fixed costs and variable fixed caps. The test requires $O(M + \sum_a H_a)$ arithmetic after the complete field is available, as does the existing exact risk calculation.

Proof. Strict support gives positive visitation at every anchor by successive zero-match outcomes, positive probability for every active prefix, and upward slack at every coordinate. Equation (28) then proves both necessity and the equivalence of the two sign conditions.

For sufficiency, suppose all suffix credits are nonnegative. Induct backwards over anchors in the exact maximizing recursion (20), assuming every later maximum equals its baseline value V_p. Within the current row, start at the last head and assume the optimized suffix after head j equals F_{j+1}. The continuation under $q_{a,j} \in [p_{a,j}, 1]$ is $(1 - q_{a,j})V_p(a + j) + q_{a,j}F_{j+1}$, whose slope is $-\mathcal{C}_{a,j} \le 0$. The baseline lower endpoint maximizes it and leaves the suffix equal to F_j. Reverse induction over heads and then anchors proves that baseline work is the global maximum over the upward box. Empty rows simply add their fixed call cost. Computing all credits uses their backward recurrence and the Bellman values.

Equivalently, at each internal state (a, j) one may keep the baseline Bernoulli transition or force acceptance. Mixing these actions with probability $\alpha_{a,j}$ yields $q_{a,j} = p_{a,j} + (1 - p_{a,j})\alpha_{a,j}$. Charge the call cost only at entry to its anchor. This gives the ordinary finite-state policy improvement interpretation: the force-accept advantage for work is $-(1 - p_{a,j})\mathcal{C}_{a,j}$. The result does not follow from multiaffinity alone and does not establish benign optimization in shared parameter space.

Comparing two upgrade sets. For variable caps define the available comparison indices $\mathcal{I} = \{a + j : 1 \le j \le H_a\}$. Under strict support,

$$\mathcal{R}_{\mathrm{surv}}(A) = 0 \quad \Longleftrightarrow \quad D_p(b) \ge 0 \quad \text{for every } b \in \mathcal{I}. \tag{30}$$

Nonnegative differences suffice by the survival performance-difference identity. If $D_p(a + j) < 0$, increasing only $A_{a,j}$ slightly is legal because adjacent survival inequalities are strict. Its exact work increase is $-d_p(a)\epsilon D_p(a + j) > 0$, proving necessity. Such a survival change may lower a later conditional probability; it need not belong to the box defining $R_\uparrow$.

The same destination condition characterizes maximal commitment's optimality within the following full-information class: after observing L_a, a controller may commit any $0, \dots, L_a$ matched drafts, paying the unchanged call cost. If all available adjacent values decrease, maximal advance minimizes continuation and satisfies the optimal Bellman equation. If $D_p(a + j) < 0$, the positive-probability event $L_a = j$ allows commitment of $j - 1$ rather than j drafts and a strictly cheaper continuation. This equivalence does not supply future laws to an online controller. No available destination comparison uses $D_p(0)$.

Intuition. Every full row exposes its last destination directly; only destinations near the start can hide an unfavorable difference inside a weighted suffix.

Corollary 1 (Fixed-cap localization)**.** *Let $M \ge 2$, $K \ge 1$, $H_a = \min\{K, M - a - 1\}$ and $K_\star = \min\{K, M - 1\}$. Under the premises of Proposition 5, conditional safety is equivalent to*

$$D_p(b) \ge 0 \quad (K_\star \le b \le M - 1), \qquad \mathcal{C}_{a,j} \ge 0 \quad (a + j < K_\star). \tag{31}$$

There are $K_\star(K_\star - 1)/2$ early credit tests. If $R_\uparrow(p) = 0$ but $\mathcal{R}_{\mathrm{surv}}(A) > 0$, every negative available destination difference lies among $1, \dots, K_\star - 1$.

Proof. The exposed last-head indices $a + H_a$ form precisely $\{K_\star, \dots, M - 1\}$. Since $\mathcal{C}_{a,H_a} = D_p(a + H_a)$, conditional safety implies all displayed tail signs. Conversely, under these signs every credit with $a + j \ge K_\star$ is a nonnegative weighted sum of tail differences. The remaining credits are exactly those displayed. For each $b = 1, \dots, K_\star - 1$, there are b active pairs with $a + j = b$; their count sums to $K_\star(K_\star - 1)/2$. The final statement follows from (30).

This localizes the possible negative differences when conditional risk is zero. It does not localize all positive risk gaps, bound their magnitude, or imply that risk per token vanishes with response length. For $M = 0$ or $M = 1$ there is no active coordinate and both risks vanish. With arbitrary variable caps, conditional safety still requires every suffix-credit test; the exposed endpoint set $\{a + H_a : H_a \ge 1\}$ gives necessary destination signs but cannot replace those tests.

Table 10: Fixed-cap safety tests under strict support, with $K_\star = \min\{K, M-1\}$. The split locates possible negative destination differences when conditional risk is zero, not the magnitude of risk gaps.

Destination index	Conditional safety test	Interpretation
$1 \le b < K_\star$	All credits with $a + j = b$	Suffixes may offset $D_p(b) < 0$
$K_\star \le b < M$	$D_p(b) \ge 0$	A last head exposes each sign

Why the support premise is substantive. At probability zero, a negative credit may be hidden behind zero prefix mass, and a joint change can expose harm invisible to all baseline partials. At probability one, a negative credit may have no upward slack. An unreachable anchor can also carry an irrelevant credit. Thus neither necessity of every credit sign nor the local-gradient/global equivalence extends automatically to boundary fields. Nonnegative credits remain sufficient by the same backward proof, but boundary fields otherwise require the full exact risk recursion. Rounding stored probabilities to zero or one does not preserve the strict-support premise. Finally, these results characterize independently adjustable conditional fields, not a common budgeted parameter update, Adam training, or native serving latency.

Sharp absolute bounds need not estimate small risks accurately. The looseness of the nested-minorant bound is not restricted to probabilities near zero or one. Take unit costs, $M = 4$, $K = 2$, and conditional rows

$$p = \big((1/2, 1/20 - \eta), (4/5, 1/2), (1/5), ()\big), \qquad 0 < \eta < 1/40.$$

Every active probability lies in $[1/40, 4/5]$. Backward substitution gives $V_p(3) = 1$, $V_p(2) = 9/5$, $V_p(1) = 44/25$ and $V_p(0) = 69/25 + 2\eta/5$. The only negative conditional credit is $\mathcal{C}_{0,1} = -4\eta/5$. All later rows maximize work at their baseline values, so raising only $p_{0,1}$ to one gives $R_\uparrow(p) = 2\eta/5$. In the larger survival set, raising only $A_{0,1}$ to one while keeping $A_{0,2}$ fixed gives $\mathcal{R}_{\mathrm{surv}}(A) = 1/50$. The greatest nested minorant changes only $A_{1,2} = 2/5$ to $1/5$; its work increment is $J(B) - J(A) = 1/10$. Consequently,

$$\frac{J(B) - J(A)}{R_\uparrow(p)} = \frac{1}{4\eta} \longrightarrow \infty. \tag{32}$$

Thus matching worst-case examples do not provide a uniform relative-error guarantee for small positive risks. The minorant and the exact sign test explain structure; neither reduces the cost of obtaining the full field.

For clarity, on one fixed path with unit call costs, let R_Γ maximize work increase over updates with $\|\Delta\Gamma\|_F \le \rho$, holding Ω, W and biases fixed, and requiring every active conditional probability to be nondecreasing. The attainable image is a subset of the conditional box. Hence

$$0 \le R_\Gamma \le R_\uparrow(p) \le \mathcal{R}_{\mathrm{surv}}(A) \le J(B) - J(A).$$

The zero update is feasible. All quantities here use the identical baseline, path and cost convention. This inclusion does not establish a nonzero shared witness, the tightness of the shared bound, or improvement from training or deployment.

Complete fitted-field anatomy. We applied this characterization retrospectively to all 10,752 fields from the 42 fitted configurations in Appendix D.5. All 2,596,320 active saved probabilities lie strictly between zero and one; none rounded to a probability boundary. The suffix-credit test identifies the same 370 conditionally safe fields as exact risk, while 244 also satisfy survival safety and posterior maximal-commitment optimality. The remaining 126 fields contain 152 negative destination differences, all at indices 1, 2 or 3 and none at indices 4 or later, as predicted for $K = 4$. These fields involve 27 prompt clusters, not 126 independent prompts. Across the complete family, the 512 prompt clusters belong to two disjoint cohorts; configurations reuse prompts within their cohort. Independent arithmetic reproduces every field's value, credit and risk digest and all grouped counts. This is a description conditional on the fitted heads, not new confirmation, a causal training explanation, or evidence that a shared update attains the free-field risk.

E SHARED-HEAD REACHABILITY AND STRICT CAP CONTROL

The rectangular risk need not equal the risk of genuine shared-parameter changes. We give an explicit, nonempty separation and its sharp boundary in a small tied-representation class. We then state a distinct policy comparison: an optimal choice among all shorter caps can remove a fixed-cap reversal under stated cost and information assumptions.

E.1 A SHARP BOUNDARY WITH A SHARED PROBABILITY

Take four transient anchors, unit costs, maximum cap two, and rows

$$p_0 = (x, y), \qquad p_1 = (u, v), \qquad p_2 = (x), \qquad p_3 = ().$$

The two occurrences of x are tied. Backward evaluation gives

$$\begin{aligned} &V_3 = 1, \quad V_2 = 2 - x, \quad V_1 = 3 - u - uv - x + ux, \\ &V_0 = 1 + (1 - x)V_1 + x(1 - y)V_2 + xy. \end{aligned} \tag{33}$$

Prefill is fixed and cancels from comparisons.

Intuition. A harmful change to one probability can become safe when the same parameter must also improve a later prediction.

Proposition 6 (Sharp safety threshold in the tied four-anchor class)**.** *If $x \geq 1/2$, every coordinate-wise improvement of (x, y, u, v) in $[0, 1]^4$ preserves or decreases V_0. The uniform threshold $1/2$ is sharp for this layout: for each $0 < x < 1/2$ there are interior y, u, v and a strictly positive finite simultaneous increase of the two tied probabilities that raises work. Both a nonconstant safe path and every constructed harmful path below the threshold admit finite multiclass realizations in one exact-rank-one factor class of Equation* (1)*, including vocabulary biases, with a centered unembedding of full row rank.*

Proof. The four derivatives of Equation (33) are

$$\begin{aligned} &\partial_x V_0 = 2u + uv - 2 - y + 2x(y - u), && \partial_y V_0 = -x(1 - x), \\ &\partial_u V_0 = -(1 - x)(1 + v - x), && \partial_v V_0 = -(1 - x)u. \end{aligned}$$

The last three are nonpositive. Bounding $v \leq 1$, if $y \geq u$ then $x \leq 1$ gives $\partial_x V_0 \leq u + y - 2 \leq 0$. If $y < u$, then $x \geq 1/2$ gives $\partial_x V_0 \leq 2u - 2 \leq 0$. Integrate these derivatives one coordinate at a time inside the upper rectangle to obtain finite-change safety.

For sharpness, fix $0 < x < 1/2$ and put $d = 1 - 2x$, $e = d/10$, $u = v = 1 - e$, $y = e$ and $h = d/16$. The tied derivative is $D = d + (4x - 5)e + e^2 \geq d/2$. Since the second derivative in x is exactly $2(y - u)$,

$$V_0(x + h, y, u, v) - V_0(x, y, u, v) = hD + h^2(y - u) \geq 7d^2/256 > 0.$$

Moreover $x + h < 1/2$ and all active probabilities remain interior. Any smaller proposed uniform threshold permits an x above it and below $1/2$. The factor realizations below complete the proof.

Positive free-field risk but zero reachable upward risk. At $(x, y, u, v) = (1/2, 1/2, 9/10, 9/10)$, untie only the first occurrence of x. Its derivative is $V_2 - V_1 - y(V_2 - V_3) = 1/100$. Increasing that coordinate to one raises continuation cost by $1/200$, and every smaller strict increase is harmful. The free-field risk is therefore positive. In contrast, every dominating change that respects the tie is safe by Proposition 6.

Multiclass shared-factor realization. Let vocabulary size $V \geq 3$, representation dimension two, $h_0 = h_2 = h_3 = (1, 0)$ and $h_1 = (0, 1)$. Every target is the last vocabulary token. Fix $0 < \delta < 1$. Let its column of W be $(1, 1)^\top$, one non-target column be $(0, \delta)^\top$, and the remaining $V - 2$ columns be zero. The native target is strictly preferred at both representation types. Relative to a zero column, the two-column determinant has magnitude δ; thus the centered unembedding has row rank two and the hidden-row-to-softmax map is injective. These are constructed representations, not measured states. Write $S_0 = V - 1$, $S_1 = V - 2 + e^\delta$ and $\ell(t) = \log(t/(1 - t))$. For $1/2 \leq t < 1$, set

$$\Omega(t) = \begin{pmatrix} \log S_0 + \ell(t) - 1 \\ \log(9S_1) - 1 \end{pmatrix}, \quad \Gamma_1 = \Gamma_2 = (1, 0), \quad b_1 = b_2 = 0.$$

Only the first residual coordinate changes, so non-target log-sums remain $\log S_0$ and $\log S_1$. The active probabilities are $x = y = t$, $u = v = 9/10$. Each strict increase of t strictly decreases work because $\partial_y V_0 < 0$ while $\partial_x V_0 \le 0$. All factors and logits are finite for $t < 1$. The second entry of Ω is positive, and the concatenated Γ is nonzero, preserving exact rank one.

For arbitrary factor and bias changes, equality of h_0, h_2 and their first-head target still forces $p_{0,1} = p_{2,1}$. Hence every realizable dominating field is in the safe tied domain, not merely the one-parameter path displayed above. Every positive-radius neighborhood of the baseline contains nonzero finite improving changes. A radius smaller than half the minimum baseline factor norm preserves both nonzero factors. Bias coordinates remain free in that neighborhood; reachable upward risk is zero there even though the enclosing independent rectangle has positive risk.

For the harmful path used in sharpness, put $D = \log(S_1/S_0) > 0$. For $t \in [x, x + h]$, use

$$\Omega(t) = (\log S_0 + \ell(t) - 1, \log S_1 + \ell(u) - 1)^\top,$$
$$\Gamma_1 = (1, 0), \quad b_1 = 0, \qquad \Gamma_2 = \alpha(t)(1, 0), \quad \alpha(t) = \frac{\ell(v) - \ell(y) + D}{\ell(u) - \ell(t) + D}.$$

Set the target component of b_2 to $\ell(y) + \log S_0 - 1 - \alpha(t)\Omega_0(t)$ and every other component to zero. Head one realizes t, u; the second head's target log-odds are $\ell(y)$ and $\ell(y) - D + \alpha(t)[\ell(u) - \ell(t) + D] = \ell(v)$. Since $t < 1/2 < u$ and $u > 9/10$, the denominator and $\alpha(t)$ are positive, and the second entry of $\Omega(t)$ is positive. The full path has finite factors and biases, exact rank one, and nondecreasing active probabilities.

The constructions retain repeated representations despite the injective centered unembedding. Factor-gauge and finite-design/bias ambiguities remain: this is not full parameter-quotient injectivity, a characterization for arbitrary backbones, or a method for computing reachable risk on distinct measured representations. Representation tying is an established source of functional equality; the contribution here is this explicit finite-work separation, its multiclass realization and its sharp threshold in the stated layout. The original same-rank harmful construction remains a counterexample to any claim that sharing is always protective.

E.2 An optimal cap can avoid a fixed-cap reversal

At every anchor, allow each deterministic proposal cap $k = 0, \ldots, H_a$. Conditional probabilities are cap-consistent: choosing a longer cap does not change the earlier draft probabilities. Call cost $c_a(k) > 0$ is deterministic, independent of the probability field, and nondecreasing in k; it may include nonzero draft cost. Future states depend only on the verified prefix, not on unused drafts or verification chunking. Policies use the current anchor and the supplied complete field. The cap is selected before drawing the proposal tokens; proposal randomness is fresh and cannot be inspected before that choice. The result is not an oracle lower bound for policies that observe a random draft's content and then select its cap without conditioning the state.

Intuition. An optimal cap need not use an extra proposal when accepting it leads to a more expensive continuation.

Proposition 7 (Monotonicity of optimal strict cap control)**.** *Under these premises, minimum expected total cost over the cap policies is nonincreasing under coordinatewise improvement of the conditional match probabilities. Randomizing the chosen cap cannot improve the optimum.*

Proof. Fix continuation costs w, a coordinate j and a cap $k \ge j$. Let F_k denote the expected cost of choosing cap k, and let $B_{j,k}$ be the later continuation conditional on the first j drafts matching. Then $B_{j,k}$ is independent of p_j, and

$$F_k - F_{j-1} = c_a(k) - c_a(j-1) + A_{j-1} p_j [B_{j,k} - w_{a+j}].$$

If the bracket is positive, cap $j - 1$ weakly dominates k for every p_j. Otherwise F_k is nonincreasing in p_j. Smaller caps do not depend on it. Thus the row minimum is nonincreasing in p_j after discarding dominated actions. It is also nondecreasing in each continuation cost because each action is an average with nonnegative transition probabilities. Backward induction proves the claim. A randomized cap averages deterministic action values and cannot beat their minimum.

The reference recurrence minimizes

$$F_k = c_a(k) + w_{a+1} + \sum_{i=1}^{k} A_{a,i}(w_{a+i+1} - w_{a+i})$$

and breaks ties toward the smaller cap. Prefix sums permit $O(MK)$ arithmetic once the full field and costs are supplied. The saved reference implementation separately recomputes each cap's row sum and therefore costs $O(MK^2)$; its fixed results are not relabelled as an optimized implementation. This is classical Bellman minimization and action dominance, not a new general planning algorithm. SpecDec++ already formulates adaptive draft lengths through whole-response cost (Huang et al., 2025); the free-emulation result in Appendix J concerns constant costs and free randomized truncation. Here the extra premise is explicitly nondecreasing cap cost, not free acquisition of future information.

That premise is substantive. With four anchors, cap two and old rows $((0,0),(1,1),(0),())$, improve only the root first probability to one. If cap zero costs 100 at the first three anchors while every positive cap costs one, and the final call costs one, optimal old continuation cost is two and new cost is three. Nonmonotone costs violate the premise.

Full-field optimal control is not automatically implementable online: future target continuations and their probability fields have acquisition costs. The result does not certify an acceptance-aware learned controller, measured latency, or amortization. A useful method must respect those information rights and outperform a same-information control after charging its complete work.

F Probability changes a readout can realize

The probability box used to define verification risk need not coincide with the changes available to a shared model. The following result gives a sufficient condition for equality on one fixed response path. Its quantifiers do not assert that one update realizes prescribed changes on several paths.

Readout and sampling semantics. Fix active hidden states $h_a \in \mathbb{R}^d$, output columns $w_v \in \mathbb{R}^d$, factors $\Omega \in \mathbb{R}^{d\times r}$ and $\Gamma_j \in \mathbb{R}^{r\times d}$, and biases $b_{j,v}$. The logits are

$$z_{a,j,v} = (h_a^\top + h_a^\top \Omega \Gamma_j) w_v + b_{j,v}.$$

The target token $y_{a,j}$ is fixed at each active coordinate. Its softmax probability is a match probability for a proposal sampled from this law and checked against that target, not an argmax match rate. Applying the work recursion additionally assumes the stated independent proposal draws and maximal-prefix commitment. Only Γ_j changes below; hidden states, Ω, output columns, and biases remain fixed. Factor shapes constrain rank to be at most r, not necessarily exactly r.

Intuition. Independent control of a path's hidden rows lets a shared readout realize any interior target-probability field on that path.

Proposition 8 (Realizability on one fixed path)**.** *Let H contain the m active hidden-state rows, and let $U = H\Omega$. Suppose U has row rank m and every active target satisfies*

$$\delta_y := \min_{v\neq y} w_y^\top (w_y - w_v) > 0.$$

For every assignment $q_{a,j} \in (0,1)$ of target probabilities on this path, there exist finite matrices Γ'_j realizing that assignment, with all other quantities fixed. The matrices may depend on the path and the full assignment. The exposure condition is sufficient, not necessary.

Proof. At one coordinate, adding $tw_y^\top$ to the residual row changes the target log-odds to

$$f_y(t) = z_y + t\|w_y\|^2 - \log \sum_{v\neq y} \exp(z_v + t w_y^\top w_v).$$

Its derivative is a convex combination of $w_y^\top(w_y - w_v)$, so $f_y'(t) \geq \delta_y > 0$. The function tends to $-\infty$ and $+\infty$ at the respective ends of the real line. Thus each requested log-odds determines a unique finite $t_{a,j}$. Let row a of X_j be $t_{a,j} w_{y_{a,j}}^\top$, using zero for unused rows. Full row rank gives the right inverse $U^+ = U^\top (UU^\top)^{-1}$, with $UU^+ = I_m$. The update $\Gamma'_j = \Gamma_j + U^+ X_j$ realizes all requested residual changes on this path.

Risk equality is a supremum statement. For the original field p, continuity of finite-path work implies

$$\sup_{\Gamma' :\, p(\Gamma') \geq p} \{J(p(\Gamma')) - J(p)\} = \max_{q \in [p,1]} \{J(q) - J(p)\} = R_{\uparrow}(p)$$

under Proposition 8. The upper bound is set inclusion. For the reverse bound, approach a maximizing corner by interior dominating fields and apply the proposition. Probabilities equal to one are generally approached rather than attained with finite softmax parameters. If the maximum is positive, a finite dominating update with positive work increase exists. Continuity also allows every otherwise unchanged coordinate to increase strictly while preserving a positive work increase. This is an existence result with unrestricted update size, not an optimizer guarantee.

A quantitative construction bound. The proof also gives

$$\left(\sum_j \|\Gamma'_j - \Gamma_j\|_F^2 \right)^{1/2} \leq \|U^+\|_2 \left[\sum_{(a,j)} \frac{\|w_{y_{a,j}}\|^2}{\delta_{y_{a,j}}^2} \left| \operatorname{logit} q_{a,j} - \operatorname{logit} p_{a,j} \right|^2 \right]^{1/2}.$$

Indeed, $f'_y \geq \delta_y$ bounds $|t_{a,j}|$ by the absolute log-odds change divided by δ_y, and $\|U^+ X_j\|_F \leq \|U^+\|_2 \|X_j\|_F$ completes the bound. Small singular values or exposure margins can therefore require large updates. A bounded log-odds box is not a parameter-norm ball: risk equality for the former does not imply equality under an arbitrary norm budget. The displayed inequality supplies a conservative sufficient condition for a log-odds box to fit a specified norm budget.

One update for several paths. The proposition states $\forall P \, \forall q_P \, \exists \Gamma'_P$. It does not state $\forall (q_P)_P \, \exists \Gamma' \, \forall P$. Stacking all active hidden-state rows gives a sufficient joint condition: if $H_{\mathrm{all}}\Omega$ has full row rank and all target exposure conditions hold, the same proof gives one common update. Separate pathwise rank tests do not imply that condition.

For a simple counterexample, take $d = r = 1$, $h = \Omega = 1$, $\Gamma = -1$, zero biases, and vocabulary columns $w_+ = 1$, $w_- = -1$. Two paths have the same hidden row but opposite targets. Each individual matrix $U = [1]$ has full row rank, and each target has exposure margin 2. Both baseline target probabilities are $1/2$. Each path individually can reach $3/4$, but one common update cannot make both targets more likely than $1/2$, because their probabilities sum to one. Averaging pathwise witnesses does not establish a witness for one shared model.

Numerical scope. Exact rational work calculations on probabilities reconstructed from stored log probabilities concern that declared field. A new floating-point forward pass need not reproduce it exactly. Modular full rank certifies rational rank of the stored tensor product; floating singular-value checks and full-vocabulary head replays are separate numerical evidence. The existing replay envelopes do not include a validated directed-interval bound for every library operation. Their agreement cannot be interpreted as an unconditional guarantee for native serving. Positive work gain without coordinatewise dominance does not establish a dominating joint update, and failure of a sufficient numerical test does not prove that no such update exists.

Complete development evaluation. Table 11 summarizes the complete 42-cell Qwen3.5-0.8B family, with 256 stored probability fields per cell. Cells share prompts, checkpoints and training histories, so these counts are not independent trials. All 10,752 fields pass the recorded numerical path-rank and exposure checks. For every positive-risk field, a finite path-specific construction is retained; 10,266 satisfy both implemented dominance and work-increase envelope checks. Each construction may use a different update. These results therefore do not establish a single update for the whole cohort or a useful parameter-norm bound.

The complete recomputation covers every exact field and every predetermined targeted construction, together with the two fixed joint configurations and both independent constructions. It retains the original full-head evaluations outside the targeted set; it is not another forward evaluation of every original head record. The target selection is development-informed, not a prospective confirmation design.

Table 11: Complete pathwise development evidence. Free-field risks use the stored probability fields; finite readout changes use full-vocabulary floating-point evaluation. Numerical envelope checks are not validated interval certificates.

Evidence object	Count
Stored probability fields	10,752
Zero free-field risk	370
Positive free-field risk	10,382
Finite path-specific constructions	10,382
Constructions passing both numerical checks	10,266
Fields checked by independent arithmetic	256

Table 12: All fixed joint attempts. The scale s specifies the target log-odds profile defined in the text. None establishes a dominating work-increasing update shared across prompts.

Training objective	Scale s	Iterations	Outcome
Adaptive single-anchor	0.01	9	Converged; dominance not established
Adaptive single-anchor	0.1	5	Did not converge
All-anchor local	0.01	5	Did not converge

Joint attempts retain the shared-update requirement. The two fixed Qwen3.5-0.8B heads use adaptive single-anchor training at learning rate 10^{-4} and all-anchor local training at 3×10^{-5}, both with training seed 1. Each is tested with one common update on its own 256 recorded paths from GSM8K and MBPP. For the mean calls per token $F(p) = 256^{-1} \sum_P J_P(p_P)/L_P$, define $c_i = \partial F / \partial \operatorname{logit} p_i$, $c_i^+ = \max(c_i, 0)$ and $c_i^- = \max(-c_i, 0)$ at the baseline head. The requested probability profile is

$$\operatorname{logit} q_i^\star = \operatorname{logit} p_i + s \left(\epsilon + \frac{c_i^+}{\max_\ell c_\ell^+} \right),$$
$$\epsilon = \min\left\{ 10^{-3}, \frac{0.1 \sum_i (c_i^+)^2}{\max_i c_i^+ \ \max(\sum_i c_i^-, 10^{-300})} \right\}.$$

Here s scales the target log-odds profile, not log probabilities; its largest requested log-odds increase is $s(1+\epsilon)$, not s. The positive-credit maximum is nonzero for these configurations, and the profile is fixed before the iterative solve. Table 12 retains every attempted scale. For the only converged attempt, the work-increase lower envelope is 1.6668678×10^{-6} calls per token, but the minimum probability margin is $-6.8652602 \times 10^{-9}$. A positive work increase does not compensate for failed coordinatewise dominance. Neither envelope is a validated directed-interval bound, and no attempt establishes the intended dominating joint witness. The stopping rule ends a configuration after failed convergence; no additional step or iteration is selected afterward.

Independent comparison and retained discrepancies. Exact rational digests, identities and discrete decisions require exact equality. Floating comparisons use the fixed tolerance $|x_{\text{new}} - x_{\text{old}}| \leq 10^{-10} + 10^{-8}|x_{\text{old}}|$. The complete replay retains two failed comparisons. First, the historical record of the construction using $\eta = 5 \times 10^{-10}$ omitted this setting from its metadata. Its shared numerical values agree, but the registered schema comparison fails. The construction with $\eta = 10^{-8}$ is the matching-target independent comparison; the smaller setting is retained as a distinct construction, not substituted for it.

Second, independent arithmetic on the fixed 256-field subset agrees on all exact-risk and discrete decisions, but one of 247 finite head targets has a work-increase discrepancy of 1.3459263×10^{-9}, exceeding the fixed tolerance. The two values are 0.0005539909591456056 and 0.0005539923050719153. No tolerance is relaxed and no failed comparison is replaced. Consequently, exact-field reproducibility is supported, but complete floating agreement and a cohort-common dominance witness are not established.

G FINITE-BUDGET RISK OF ONE SHARED READOUT UPDATE

The rectangular certificate allows each probability to change independently. Here one update must act on every prompt. Fix N positive-length target paths, their hidden rows, Ω, W and biases in (1). Only the stacked vector $x = \mathrm{vec}(\Delta\Gamma_1, \dots, \Delta\Gamma_K)$ changes. Write $c_P = 1/(Nn_P)$, and index active coordinates by $i = (P, a, j)$. With $u_i = h_a\Omega$, define

$$\ell_i(x) = \log \mathrm{softmax}(z_i(0) + u_i \Delta\Gamma_j W)_{y_i}, \qquad \delta_i(x) = \ell_i(x) - \ell_i(0).$$

For finite $\rho \geq 0$, the shared risk is

$$R_\rho^{\mathrm{shared}} = \sup_{\|x\|_2 \leq \rho,\ \delta_i(x) \geq 0\ \forall i} \sum_P c_P [J_P(p(x)) - J_P(p(0))]. \tag{34}$$

The finite logits give positive mass to every vocabulary item. The constraint set is convex because each ℓ_i is concave; J need not be convex. This norm depends on the fixed factor coordinates, not just the represented function. A zero risk here does not certify changes to Ω or biases. Empty coordinate sums and maxima are zero, giving zero risk when no path has an active proposal. Prefill cancels in every difference.

The coordinates that can cause harm. Let $D_P(a) = V_P(a) - V_P(a+1)$ and $A_{P,a,k} = \prod_{j \leq k} p_{P,a,j}$ refer to the baseline. Choose fixed, uniform bounds on the whole feasible ball, for example

$$t_i = \min\{-\log p_i, \rho\|u_i\|_2 \max_v \|W_{y_i} - W_v\|_2\}.$$

Here W_v denotes a vocabulary column. This bounds a log-probability increase by bounding every target-relative logit perturbation, and requires the complete vocabulary. Put $T_{P,a,k} = \sum_{j \leq k} t_{P,a,j}$, $E(T) = (e^T - 1)/T$ with $E(0) = 1$, and

$$b_{P,a,j} = c_P \sum_{k=j}^{H_a} [-D_P(a+k)]_+ A_{P,a,k} E(T_{P,a,k}). \tag{35}$$

Only prefixes preceding a negative continuation difference contribute.

Intuition. Shared gradients can cancel the directions that would make independent probability changes harmful.

Theorem 5 (Selective shared-update bound)**.** *Let $g_i = \nabla\ell_i(0)$ and $C_i = \mathrm{Cov}_{v\sim\pi_i}(\nabla z_{i,v})$, where π_i is the full baseline categorical law. For any fixed weights $w_i \geq b_i$, put $r_w = \sum_i w_i g_i$ and $F_w = \sum_i w_i C_i$. Let $B = \rho \max_i \|u_i\|_2 \,\mathrm{diam}\{W_v\}_v$ and $\alpha(B) = (B - 1 + e^{-B})/B^2$, with $\alpha(0) = 1/2$. Then*

$$0 \leq R_\rho^{\mathrm{shared}} \leq \min\left\{ \sum_P c_P R_{\mathrm{up}}(p_P),\ \rho\|r_w\|_2,\ \frac{r_w^T F_w^\dagger r_w}{4\alpha(B)} \right\}. \tag{36}$$

In particular, exact $r_w = 0$ certifies zero shared risk. The weights may vanish outside the harmful-prefix support; not all probabilities must remain fixed.

Proof. The performance-difference identity gives

$$J_P(q) - J_P(p) = -\sum_a d_q(a) \sum_k [A_{q,a,k} - A_{p,a,k}] D_P(a+k),$$

where $0 \leq d_q(a) \leq 1$ is the visit probability in the increasing chain. For feasible x, drop nonpositive terms and use $e^s - 1 \leq E(T)s$ for $0 \leq s \leq T$. Collecting coefficients yields $\sum_P c_P \Delta J_P \leq \sum_i b_i \delta_i(x) \leq \sum_i w_i \delta_i(x)$. The exact exponential-family identity is

$$\delta_i(x) = g_i^T x - \mathrm{KL}(\pi_i \| \pi_i(x)).$$

For a logit perturbation with oscillation at most B, its tilted law obeys $\pi_{i,t}(v) \geq e^{-tB}\pi_i(v)$. The minimum-over-centers formula for variance therefore gives $\mathrm{Var}_{\pi_{i,t}}(f) \geq e^{-tB}\,\mathrm{Var}_{\pi_i}(f)$. Integrating the log-partition Hessian with weight $1 - t$ proves $\mathrm{KL}(\pi_i\|\pi_i(x)) \geq \alpha(B) x^T C_i x$. Thus the work difference is at most $r_w^T x - \alpha(B) x^T F_w x$. If $v \in \ker F_w$, every supported coordinate has constant logit perturbation over the full vocabulary. Hence $g_i^T v = 0$ there, and $r_w \in \mathrm{range}(F_w)$. Completing the square gives the pseudoinverse term; the norm inequality gives the norm term. Set inclusion gives the rectangular term, and the zero update gives the lower bound.

Table 13: Complete trained-head shared-update attempt. Each configuration has five registered roles: constrained, norm-matched local, norm-matched whole-response, fixed random and zero. Role records are not independent solver attempts. All available candidates have exact stored squared norm zero. The missing configuration remains unavailable rather than a success or a zero-result observation.

Object	Registered	Available	Observed outcome
Common-update constructions	96	95	95 residual failures
Construction-role records	480	475	475 zero-norm updates
Evaluation baselines	96	95	95 floating diagnostics
Evaluation-role records	480	475	475 saved-value replays agree

An upper bound from any trial point. For fixed $w \geq b$, $S_w(x) = \sum_i w_i \delta_i(x)$ is concave. Its supporting hyperplane at any finite x_0 gives

$$R_\rho^{\text{shared}} \leq S_w(x_0) - \nabla S_w(x_0)^T x_0 + \rho \|\nabla S_w(x_0)\|_2.$$

The trial point need not satisfy dominance. A solver's objective value alone is instead a lower bound on the corresponding maximization, not a safety certificate. Floating residuals, softmax underflow and discarded eigenvalues cannot be treated as exact cancellation or a rigorous pseudoinverse bound. The norm expression needs full-vocabulary gradient accumulation, but not a dense parameter-sized covariance matrix. Its practical cost and tightness on trained heads have not been established.

A strict gap on two target-greedy paths. Let $s = \log(9)/2$, $d = 2$, $r = 1$, $K = 2$, $\Omega = (1, 0)^T$, $W_+ = (0, 1)^T$, $W_- = (0, -1)^T$, and $\Gamma_1 = \Gamma_2 = (0, 1)$ with zero biases. Use the target paths $(-, +, +, +, +)$ and $(-, -, +, +, +)$ and hidden states

$$h_0 = (1, -1), \quad h_1^A = (0, s), \quad h_1^B = (2s, -s), \quad h_2 = (-2s, s), \quad h_3 = h_4 = (0, s).$$

The native target logits $h_a W$ select token y_a strictly, whereas head j at anchor a predicts y_{a+j}. Both draft fields are $[(1/2, 1/2), (9/10, 9/10), (1/10), ()]$. Their values are $(473/200, 32/25, 19/10, 1, 0)$, so only $D(1) = -31/50$ is negative. The only harmful-prefix coordinates are therefore the two first root proposals. Their target probabilities are complementary; common dominance fixes both at $1/2$. Equal weights cancel their opposite gradients, and Theorem 5 gives zero shared risk for every finite budget. Nonzero feasible changes to the second head remain possible. In contrast, independently raising each root's first probability to one gives $17/200$ extra calls per path, or $17/1000$ after averaging and normalizing by path length. An arbitrarily small positive increase already harms both fields. This is a constructed common-update obstruction, not a frequency claim about pretrained representations or a guarantee for full-factor updates.

Relation to standard tools. Supporting hyperplanes and weighted scalarization are classical convex analysis (Boyd & Vandenberghe, 2004); the KL identity and covariance Hessian are standard exponential-family geometry (Wainwright & Jordan, 2008). Exponential Hessian comparisons also occur in logistic-loss analysis (Bach, 2009). The specific content here is the verifier's selective weights (35), their common-update risk bound, and the strict relaxation gap. These results do not establish a generally superior optimizer, a useful real-head bound, or a deployment speedup.

A complete trained-head attempt. We fixed all 24 trained rank-108 heads, both GSM8K and MBPP, and relative factor budgets $0.001\|\Gamma\|_F$ and $0.01\|\Gamma\|_F$ in each checkpoint's stored gauge. Each of the 96 configurations used 16 paths to construct one common $\Delta\Gamma$ and 16 disjoint paths to assess the unchanged candidate. These are development paths, not a previously untouched cohort. The procedure used the full vocabulary, independent categorical draft samples, $K = 4$, and unit verifier calls including prefill. It held Ω, W and biases fixed; the result does not address their joint updates, Adam training or native decoding latency.

The 95 distinct constructions each failed the prespecified conjugate-gradient residual criterion. Norm matching therefore makes the local, whole-response and random controls zero as well; the zero control adds no informative contrast. Using saved target log-probabilities from the original full-vocabulary evaluation, an independent implementation recomputed the expected calls and

rectangular-risk diagnostics for all 95 available baselines and 475 evaluation roles with no disagreement at absolute and relative tolerance 10^{-10}. The remaining configuration and five roles are unavailable, so numerical coverage is incomplete. This comparison checks the saved arithmetic, not a feasible nonzero direction or a rigorous probability interval. Neither a harmful common update nor a proof that no such update exists follows from this solver failure.

The 96 construction tasks, 96 evaluation tasks and two collection jobs sum to 3,701,456 allocated CPU-seconds. Their allocation intervals overlap; the sum is neither elapsed study time, actual CPU use, billing, nor monetary cost. Head training, model acquisition, result extraction and independent checks are outside that total. These costs, all failures and the single missing configuration remain in the saved-result package.

G.1 Finite Steps on One Common Readout Direction

A nonzero direction that improves every linearized target probability need not yield a useful finite update. This subsection makes that distinction explicit for fixed Ω, biases and vocabulary projection, with only Γ changing. The statements apply to a finite full-support softmax law at each active coordinate. They do not assert reachability by Adam or a benefit over trained local or whole-response objectives.

Classical curvature bounds on a fixed ray. At coordinate i, let π_i be the baseline vocabulary law and s_{iw} the logit slope along one common parameter direction. Set $m_i = s_{iy_i} - \mathbb{E}_{\pi_i} s_i$, $R_i = \max_w s_{iw} - \min_w s_{iw}$, $b_i = \max_w (s_{iw} - \mathbb{E}_{\pi_i} s_i)$ and $\sigma_i^2 = \mathrm{Var}_{\pi_i}(s_i)$. For a step $t \geq 0$ the exact target log-gain is

$$\delta_i(t) = \log \pi_i(t)_{y_i} - \log \pi_{iy_i} = t m_i - \mathrm{KL}(\pi_i \| \pi_i(t)). \tag{37}$$

Classical exponential bounds (Hoeffding, 1963; Bennett, 1962) imply

$$\mathrm{KL}(\pi_i \| \pi_i(t)) \leq t^2 R_i^2/8, \tag{38}$$

$$\mathrm{KL}(\pi_i \| \pi_i(t)) \leq \frac{\sigma_i^2 t^2}{2(1 - t b_i/3)} \quad (0 \leq t b_i < 3). \tag{39}$$

Thus, when every nonconstant coordinate has $m_i > 0$, either common limit

$$t \leq \min_i \frac{4 m_i}{R_i^2}, \qquad \text{or} \qquad t \leq \min_i \frac{m_i}{\sigma_i^2 + m_i b_i/3} \tag{40}$$

ensures $\delta_i(t) \geq t m_i/2$ at every nonconstant coordinate. The minima run only over nonconstant coordinates; if there are none, every law is unchanged for every step. An external step cap preserves the parameter budget. This application does not introduce a new concentration inequality.

For completeness, the first bound is the centered bounded-variable log-moment bound. For the second, put $X = s - \mathbb{E}_\pi s$ and $b = \max X$. The function $(e^u - 1 - u)/u^2 = \int_0^1 (1 - r) e^{ru}\, dr$ is increasing, so for $X \leq b$, $e^{tX} - 1 - tX \leq X^2 (e^{tb} - 1 - tb)/b^2$. Taking expectations and using $\log(1+z) \leq z$ gives $\mathrm{KL} \leq (\sigma^2/b^2)(e^{tb} - 1 - tb)$. The series inequality $e^u - 1 - u \leq u^2/[2(1 - u/3)]$ for $0 \leq u < 3$ follows from $k! \geq 2\, 3^{k-2}$ for $k \geq 2$. Substitution gives (40). Full support makes $\sigma^2 = 0$ or $b = 0$ equivalent to constant slopes: the entire law is then unchanged. A nonconstant row with $m \leq 0$ strictly loses target probability for every positive step. Numerically near-zero values do not establish this exact neutral case. The variance limit is larger than the range limit only when $\sigma^2 + mb/3 < R^2/4$; there is no uniform improvement.

Complete two-prompt construction. We test one fixed SmolLM3-3B head with rank 108, four proposal horizons and all 128,256 vocabulary entries. Two 64-token construction prompts supply 484 active coordinates. Eight disjoint, previously opened development prompts supply 1,936 further coordinates. None of the latter determines the direction or step. The radius is $0.01 \|\Gamma\|_F = 0.392426$ in the stored factor coordinates. A bounded minimum-norm common-margin solve and an ordinary cone projection have explicit primal, dual, complementarity and dual-gap checks; no direction is selected from evaluation outcomes. The target-CE, expected-prefix, whole-work and fixed-random controls use the full radius; the projection retains its smaller norm. The comparison is therefore not norm-matched and does not establish training efficacy.

The computed common ray receives a range step of 2.39053×10^{-8}, with achieved norm 9.38105×10^{-9}. Its minimum construction log-gain is -4.26×10^{-14} in double precision and -1.42×10^{-6} under the original split-single-precision logit computation. It fails the fixed numerical gain requirement. In particular, reproducing the small double-precision effect is not a precision-resolved common improvement. The projection satisfies its linear numerical gate but its minimum finite construction gain is -7.78629, illustrating the curvature gap.

Table 14: One-step changes in expected calls per token on the complete opened development sample. Negative means less work. Construction and evaluation contain two and eight prompts respectively. Different attained norms preclude a matched-budget method comparison.

Direction and step	Norm	Construction	Evaluation
Common margin, range	9.38×10^{-9}	-6.95×10^{-10}	$+1.01 \times 10^{-11}$
Target CE	0.392426	-0.007081	$+0.011107$
Expected prefix	0.392426	-0.047877	$+0.001585$
Whole work	0.392426	-0.040040	$+0.000260$
Linear cone projection	0.360882	-0.049288	-0.000291
Fixed random	0.392426	$+0.000294$	$+0.000084$
Zero	0	0	0

All nonzero ordinary controls lower at least one construction probability. The native-arithmetic common candidate changes construction work by $+8.65 \times 10^{-11}$ calls/token, opposite to its tiny double-precision sign. No value is rounded into a successful gate. Full-vocabulary independent reconstruction covers all seven roles, both precisions and all ten prompts, but does not supply directed interval error bounds. Primary computation takes 46.518 seconds and independent reconstruction 8.102 seconds; these are head-analysis costs, not serving latency. The earlier 96-configuration attempt remains a separate failure, not repaired by this development study.

A separately registered variance-sensitive step. After observing that result, we froze a second development calculation on the identical saved ray and complete two-plus-eight prompt set. It changes only the curvature bound, with no direction search, row selection, tuning or new model. Target-centered long-double moment sums use every vocabulary entry; their inputs are binary64 logit products, not interval enclosures. All 484 construction moment screens pass. The variance limit gives $t = 0.141079$ and norm 0.055363. Recomputing the range limit with the same stable moments gives 2.39053×10^{-8}; the larger step therefore does not arise simply from changing first-moment accumulation.

The minimum construction gains are 1.44519×10^{-7} in both arithmetics. Nevertheless, the fixed split-single-precision gate requires a gain above 1.44289×10^{-4}, ten times the largest observed cross-precision gain discrepancy plus 10^{-12}. The result remains numerically unresolved under that unchanged rule. Stable accumulation and independent agreement do not supply a directed error certificate.

Construction work changes by -0.00405809 calls/token, but mean work on all eight development evaluation prompts changes by $+0.00009164$. On those prompts, target CE falls by 0.00047273 while expected-prefix loss rises by 0.00021951; the minimum coordinate gain is -0.360295. Thus a finite step that improves the computed construction coordinates does not transfer coordinatewise, and lower mean CE does not guarantee lower work. The seven earlier controls remain fixed descriptive comparators, not newly tuned competitors. Independent raw-input reconstruction checks all moments, the scalar step and all ten endpoints under both precisions. This isolates a range-bound bottleneck, but establishes neither a precision-certified common improvement nor a superior training method.

H A SHARP BOUNDARY FOR LOCAL TRAINING INFORMATION

The same-stratum construction in Appendix C shows that locally better proposals can require more verification calls. The following result asks a different question: can an arbitrarily long but finite local window determine even the correct sign of one proposal update? We retain the fixed target

path, independent sampled heads, finite-cap convention, and constant unit call cost of Equation (7). The result extends that mathematical construction; it is not a reinterpretation of a fitted-model result.

H.1 One proposal is a genuine exceptional case

Intuition. With only one proposal, skipping one position cannot bypass a longer cheap verification jump.

Proposition 9 (One-proposal monotonicity under unit call cost)**.** *For $K = 1$, every coordinatewise increase of the target-match probability field weakly decreases J. This statement concerns fixed unit verification costs; it need not hold for state-dependent costs.*

Proof. Write $D_a = V(a) - V(a+1)$, with $D_{M-1} = 1$. For $a \le M-2$,

$$D_a = 1 - p_{a,1} D_{a+1}.$$

Backward induction gives $0 \le D_a \le 1$. The derivative with respect to an active probability coordinate is $\partial J / \partial p_{a,1} = -d_p(a) D_{a+1} \le 0$, where $d_p(a)$ is the visit probability. Integrating one coordinate at a time proves the finite coordinatewise assertion, including boundary probabilities by continuity. For the cost qualification, take $M = 3$, costs $(1, 1, 3)$, and $p_{1,1} = 1$. Then $V(1) = 1$ and $V(2) = 3$, so increasing $p_{0,1}$ increases the total cost with slope two. This counterexample changes the cost assumption and is not a counterexample to the proposition.

H.2 Finite local windows do not identify the update sign

A local observation at anchor zero consists of the target path, representation rows, match probabilities, and visit probabilities at anchors $0, \dots, L$, together with K, M, and the unit call cost. It does not include the remote probability field, its continuation values, or a global parameter description from which those remote probabilities can be recovered. We ask only about changing $p_{0,1}$ while keeping every other field coordinate fixed. This is an information restriction on a coordinate-update rule, not a claim about every simultaneous neural-parameter update.

Intuition. Two identical local stretches can lead to opposite remote verification patterns, so the same extra accepted token can either save or add work.

Theorem 6 (Constant-margin nonidentifiability in one exact-rank class)**.** *For every $K \ge 2$ and $L \ge 1$, there are two strictly interior fields with $M = L + K + 3$ and the same local observation at anchors $0, \dots, L$, whose derivatives $G = \partial J / \partial p_{0,1}$ have opposite signs. Each has magnitude at least $K - 5/4 \ge 3/4$. Both fields and their root-coordinate updates belong to the same shared-factor exact-rank-two class with $p = m = 5$, raw factor count $10(K+1)$, and factor-quotient dimension $10K + 6$. At $p_{0,1} = 1/2$, the corresponding root-logit derivatives have magnitude at least $(K - 5/4)/4$.*

The probabilities lie in $[\epsilon, 1-\epsilon]$, with $\epsilon = 1/(16M^2)$, and the finite logits grow as $O(\log M)$. Thus the sign margin does not vanish with L, but the probability floor does: this is not a uniform fixed-noise or fixed-logit-bound claim.

Proof. **Two binary suffixes.** Put $B = L + 1$ and $M = B + K + 2$. At anchor zero, set $p_{0,1} = t \in (0, 1)$ and $p_{0,j} = 0$ for $j \ge 2$. At every anchor $1 \le a < B$, set $(p_{a,1}, \dots, p_{a,K}) = (1, 0, \dots, 0)$. This common segment always advances by two. Only the following suffix differs:

	$a = B$	$a = B+1$	$B+2 \le a < M$
A	$(1, \dots, 1)$	$(0, \dots, 0)$	$(0, \dots, 0)$
B	$(1, 0, \dots, 0)$	$(1, \dots, 1)$	$(0, \dots, 0)$.

In world A, the jump from B reaches $M-1$, whereas all later anchors advance by one. Consequently $V(B) = 2$, $V(B+1) = K+1$, and $D_B = 1 - K$. In world B, the jump from $B+1$ reaches M, while $V(B+2) = K$, giving $V(B) = K+1$, $V(B+1) = 1$, and $D_B = K$. These jumps obey the finite-cap convention, including the final no-draft call.

At each common anchor $1 \le a < B$, $D_a = 1 - D_{a+1}$. The map $d \mapsto 1 - d$ exchanges $1 - K$ and K. Thus D_1^{A} and D_1^{B} are these two numbers in an order determined by the parity of L. Because

$V(0) = 1 + (1-t)V(1) + tV(2)$, the two binary-reference slopes are $G_0 = -D_1$, with opposite signs and magnitudes $K-1$ and K. This assertion is affine in t, not merely infinitesimal.

Interior fields and the derivative error. Replace every displayed zero or one by ϵ or $1-\epsilon$, respectively, keeping $t = 1/2$. Until two coupled traces disagree, a round uses at most its accepted-prefix length plus one relevant independent uniform comparison. The sum of these quantities is at most M over a complete response. Ignored proposal tails do not enter this count. A first-disagreement union bound and the remaining-call range at most M give, at every suffix anchor s,

$$|V_\epsilon(s) - V_0(s)| \le M^2\epsilon.$$

This is the coupling argument of Lemma 1, applied from s; using M rather than $M-s$ only loosens its bound. Differentiating the root row, whose heads $j \ge 2$ all have probability ϵ, yields

$$G_\epsilon = \sum_{k=1}^{K} \epsilon^{k-1}\big(V_\epsilon(k+1) - V_\epsilon(k)\big).$$

The $k = 1$ difference changes by at most $2M^2\epsilon$. Every remaining value difference has magnitude at most M, so

$$|G_\epsilon - G_0| \le 2M^2\epsilon + \frac{M\epsilon}{1-\epsilon} \le 4M^2\epsilon = \frac{1}{4}.$$

Both signs therefore persist with magnitude at least $K - 5/4$. All rows through L are exactly equal between the two interior worlds. Acyclicity implies that their visit probabilities at these anchors are exactly equal as well: a later row cannot alter an earlier visit.

A fixed five-dimensional exact-rank realization. Use five frozen one-hot representation types: anchor zero, the common body, anchor B, anchor $B+1$, and the remaining tail. All five types occur because $L \ge 1$. Write $c = (1,1,1,1,1)^\top$ and choose the fixed binary unembedding $W = [0, c]$, whose target always selects token one. Let $\ell_1 \in \mathbb{R}^5$ be the first-head logit differences of one interior world on those five types, and let ℓ_2 be the common logit differences of all its heads $j \ge 2$. Define

$$u_i = \ell_i - c, \quad v = c/5, \quad \Omega = [u_1, u_2], \quad \Gamma_1 = \begin{bmatrix}1\\0\end{bmatrix} v^\top, \quad \Gamma_j = \begin{bmatrix}0\\1\end{bmatrix} v^\top \ (j \ge 2).$$

Then $R_j = \Omega\Gamma_j$ and the residual readout $hW + (hR_j)W$ has exactly the declared logit difference, since $v^\top c = 1$. Put $\lambda = \log((1-\epsilon)/\epsilon) > 0$. On the common body and the final tail, u_2 has the same nonzero entry $-\lambda - 1$, while u_1 has distinct entries $\lambda - 1$ and $-\lambda - 1$. Hence the two columns of Ω are independent, for every $t \in (0,1)$. The stacked loading $[\Gamma_1 \ \cdots \ \Gamma_K]$ also has row rank two. Both factor ranks and the concatenated residual rank are therefore exactly two throughout the root-coordinate change. Each individual residual has rank one. Changing only the first entry of u_1 changes only the root first-head logit, because the root has its own representation type. Raw factor count is $2(5+5K)$ and quotient dimension is $2(5+5K-2)$. These are pre-unembedding factor quantities, not a claim that binary unembedding has no observable-law kernel.

Finally, the root-logit derivative is $t(1-t)G_\epsilon$, equal to $G_\epsilon/4$ at $t = 1/2$. The magnitude claim follows. All non-root logits equal $\pm\lambda$, where $\lambda = \log(16M^2 - 1)$, so the construction uses finite logits of order $\log M$.

What the information obstruction rules out. A deterministic rule based only on the declared local observation must choose the same root-coordinate displacement in the two worlds. Since J is affine in that probability when the future field is fixed, every nonzero such displacement increases J in one world. Under this worst-case guarantee, the only common safe displacement is zero. A randomized rule cannot have a nonzero mean displacement that weakly improves expected work in both worlds. Zero-mean randomization is not excluded, and neither are updates using remote information or general simultaneous parameter changes. The result is not a sample-complexity lower bound, an optimizer-convergence claim, or evidence that this obstruction explains the fitted-model results.

Increment and prior-work boundary. Nonlocal value information and multi-step policy improvement are standard sequential-decision concepts (Efroni et al., 2018); truncated lookahead can already have a linear worst-case gap in general MDPs (Xu & Gan, 2026). The latter paper's

lookahead depth is not our number K of simultaneous proposals. SpecDec++ already formulates whole-response decoding cost as an MDP (Huang et al., 2025). The restricted increment here is the one-versus-two-proposal boundary, arbitrary finite-window sign nonidentifiability with a nonvanishing margin, and its common finite-dimensional exact-rank realization. It does not invent Bellman evaluation, policy gradients, delayed credit, or the general failure of short-horizon planning. Its constructive implication is a requirement on training information: local acceptance statistics alone do not determine the required continuation-sensitive coordinate credit. Whether estimating that information is useful on real backbones remains an empirical question.

H.3 A SHARP, OBSERVABLE OPERATING-REGION BOUNDARY

Mixed fitted-objective development results motivate a separate question: when is the harmful sign excluded altogether? This subsection was developed after those results, rather than preregistered as their predictor. It adds a mathematical operating-region result; it does not turn the observed outcomes into a confirmation of its premises or conclusions. The earlier numerical check covers the preceding finite-window construction, not the additional proof below.

Define the local expected accepted-prefix length

$$S_p(a) = \sum_{k=1}^{H_a} A_{a,k}, \qquad D_p(a) = V_p(a) - V_p(a+1), \quad 0 \le a < M.$$

The finite-cap Bellman identity gives

$$D_p(a) = 1 - \sum_{k=1}^{H_a} A_{a,k} D_p(a+k), \qquad D_p(M-1) = 1. \tag{41}$$

Both the local sum and the value differences refer to the declared independent sampled loop and constant unit target-call cost.

Intuition. If every anchor accepts at most one proposal on average, no hidden continuation can make a later anchor more costly than the immediately preceding anchor.

Theorem 7 (Sharp expected-prefix operating threshold; Theorem 2 with a quantitative margin)**.** *If $S_p(a) \le 1$ for every anchor, then $0 \le D_p(a) \le 1$ for every anchor and every probability-coordinate derivative of J is nonpositive. More strongly, every coordinatewise dominating field $q \ge p$ satisfies*

$$J(q) \le J(p), \tag{42}$$

even if $S_q(a) > 1$ at some anchors. If instead $S_p(a) \le 1 - \beta$ uniformly, for $0 \le \beta \le 1$, then $D_p(a) \ge \beta$ and

$$J(p) - J(q) \ge \beta \sum_{a=0}^{M-1} d_q(a) \sum_{k=1}^{H_a} (A_{q,a,k} - A_{p,a,k}) \ge 0. \tag{43}$$

The threshold one is sharp for a uniform bound on S. For every $\gamma > 1$ and every $K \ge 2$, strictly interior fields exist with $S(a) < \gamma$ at all anchors but a positive root first-head derivative. They have a shared-factor exact-rank-two realization with three frozen representation types and finite logits. The harmful root update can remain strictly inside the same $S < \gamma$ region.

Proof. **Safe operating region and finite improvements.** Backward induction in Equation (41) shows that if all later differences lie in $[0, 1]$, then $1 - S_p(a) \le D_p(a) \le 1$. This proves the asserted interval, and also $D_p(a) \ge \beta$ under the stricter bound. For an active coordinate, the ordinary probability derivative is

$$\frac{\partial J}{\partial p_{a,j}} = -d_p(a) \sum_{k=j}^{H_a} \left(\prod_{\substack{1 \le i \le k \\ i \ne j}} p_{a,i} \right) D_p(a+k) \le 0.$$

This polynomial expression covers zero probabilities without division. For finite improvements, apply the performance-difference identity (14) and sum by parts in each transition row:

$$J(q) - J(p) = -\sum_{a=0}^{M-1} d_q(a) \sum_{k=1}^{H_a} (A_{q,a,k} - A_{p,a,k}) D_p(a+k). \tag{44}$$

Every summand has the stated sign when $q \geq p$, since prefix products are coordinatewise nondecreasing. This proves both finite bounds without imposing a condition on S_q or the intermediate update path. This is a guarantee relative to the safe baseline p; it does not assert monotonicity between two later unsafe fields.

Sharpness with strictly interior probabilities. Fix $\gamma > 1$, $K \geq 2$, and put

$$\delta = \min\{(\gamma - 1)/4, 1/4\}, \qquad \epsilon = \delta/16, \qquad M = K + 2.$$

For $t \in [1/4, 3/4]$, define three row types:

$$p_{0,\cdot} = (t, \epsilon, \ldots, \epsilon), \qquad p_{1,\cdot} = (1-\epsilon, \delta, \ldots, \delta), \qquad p_{a,\cdot} = (\epsilon, \ldots, \epsilon) \quad (2 \leq a < M).$$

Every probability is in $[\epsilon, 1-\epsilon]$, with $0 < \epsilon \leq 1/64$. The root and tail satisfy

$$S(0) \leq \frac{3/4}{1-\epsilon} \leq \frac{16}{21} < 1, \qquad S(a) \leq \eta := \frac{\epsilon}{1-\epsilon} < 1 \quad (a \geq 2).$$

Anchor one has $H_1 = K$ and

$$S(1) = (1-\epsilon) \sum_{h=0}^{K-1} \delta^h.$$

Its lower bound $(1-\epsilon)(1+\delta) > 1$ is strict. Its upper bound satisfies

$$S(1) < \frac{1}{1-\delta} < 1 + 2\delta \leq 1 + \frac{\gamma - 1}{2} < \gamma.$$

Thus every root-coordinate update within the displayed interval retains the strict global $S < \gamma$ bound; only anchor one has $S > 1$.

The safe-region induction on the tail gives $1 - \eta \leq D(a) \leq 1$ for $a \geq 2$. Consequently

$$-D(1) \geq S(1)(1-\eta) - 1.$$

Differentiating only t while holding the future field fixed gives

$$\begin{aligned} G := \frac{\partial J}{\partial t} &= -D(1) - \sum_{k=2}^{K} \epsilon^{k-1} D(k) \\ &\geq S(1) - 1 - (S(1) + 1)\eta \\ &\geq \delta - \epsilon(1+\delta) - \left(1 + \frac{1}{1-\delta}\right) \frac{\epsilon}{1-\epsilon} \geq \delta - 4\epsilon = \frac{3\delta}{4} > 0. \end{aligned}$$

For the last inequality, $\delta \leq 1/4$ and $\epsilon \leq 1/64$ bound the coefficient of ϵ by $5/4 + (7/3)(64/63) < 4$. The derivative is independent of t, so every strictly positive finite root displacement that remains in $[1/4, 3/4]$ increases work. At $t = 1/2$ its logit derivative is at least $3\delta/16$. Unlike the finite-window construction above, the probability floor here depends on γ but not on K; no fixed floor is claimed uniformly as $\gamma \downarrow 1$.

The same rank-two readout class. Use $h = e_0^\top, e_1^\top, e_2^\top$ for the root, anchor one, and tail, respectively, $c = (1,1,1)^\top$, and $W = [0, c]$. Let $\ell_1 = (\operatorname{logit} t, \operatorname{logit}(1-\epsilon), \operatorname{logit} \epsilon)^\top$ and $\ell_2 = (\operatorname{logit} \epsilon, \operatorname{logit} \delta, \operatorname{logit} \epsilon)^\top$. Set

$$\Omega = [\ell_1 - c, \ell_2 - c], \qquad \Gamma_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} (c/3)^\top, \qquad \Gamma_j = \begin{bmatrix} 0 \\ 1 \end{bmatrix} (c/3)^\top \quad (j \geq 2).$$

The residual formula $hW + (h\Omega\Gamma_j)W$ realizes exactly the three row types. The second column of Ω has equal nonzero root and tail entries, whereas the first column has unequal entries there because $t \geq 1/4 > \epsilon$. Hence Ω and the stacked loading both have rank two throughout the root-update interval. So does the concatenated residual. With $p = m = 3$, raw factor count is $6(K+1)$ and factor-quotient dimension is $6K + 2$. All logits are finite; these are factor-class statements rather than claims about the observable-law quotient. This completes the sharpness proof.

Necessary obstruction, not a sufficient predictor. A positive probability-coordinate derivative at anchor a requires some negative $D(b)$ with $b > a$. Choose the rightmost negative difference $D(c)$ at or after that b. All later differences are nonnegative and therefore, by the recursion, at most one. Hence $D(c) \geq 1 - S(c)$, so necessarily $S(c) > 1$. A harmful update thus requires an above-one row somewhere downstream, not necessarily at the updated anchor: the sharpness example has $S(0) < 1$. The converse is false. With $M = 3$, $K = 2$, and all probabilities $9/10$, $S(0) = 171/100 > 1$, but $(D(0), D(1), D(2)) = (1/10, 1/10, 1)$ and every active coordinate derivative is negative. Counting $S > 1$ therefore measures an opportunity for the obstruction, not its occurrence or a prediction of benefit from continuation-aware training.

Scientific increment and use. The Bellman induction and performance-difference identity are standard; the restricted increment is the largest uniform expected-prefix threshold, its entire upward-cone guarantee, and strict-interior same-rank counterexamples above the threshold. Proposition 9 is an immediate $K = 1$ corollary because then $S(a) \leq 1$ automatically. Measured S, D, and signed credit can separate a ruled-out operating region from one in which the mechanism is possible. Their post-outcome measurement does not establish prevalence on new families, causal mediation, objective superiority, serving speed, or a new architecture.

I Training Objectives and Actual Updates

Complete continuation credit changes a gradient, but need not improve the optimizer step produced from that gradient.

What the two objectives differentiate. Let $d_p(a)$ be the probability of visiting anchor a. The exact log-probability credit for whole-response work is

$$G_{a,j} = d_p(a) \sum_{k=j}^{H_a} A_{a,k} \big[V_p(a+k+1) - V_p(a+k) \big]. \tag{45}$$

Its parameter gradient is $g = n^{-1} \sum_{a,j} G_{a,j} \nabla_\theta \log p_{a,j}$. The local-prefix loss instead uses $G^{\mathrm{local}}_{a,j} = -\sum_{k=j}^{H_a} A_{a,k}$ and gradient l. Complete credit accounts for both anchor visits and continuation costs. This is a specialization of the policy-gradient identity, not a new general gradient theorem. The all-anchor local implementation already evaluates the complete probability field and both credit arrays. Acquiring that information is not an additional cost of the whole-response direction in this comparison.

A field gradient is not an optimizer step. Shared factors couple the probability coordinates, and Adam transforms their combined gradient using its first and second moments. A useful objective must therefore improve the actual parameter update, not just its own freely adjustable probability field. We compare two candidates from identical parameters and complete optimizer state, then measure work on both their update batch and a disjoint batch. Separate fixed-parameter state swaps test whether the update-batch comparison changes with optimizer state.

What partial information can guarantee. Conditional window bounds propagate through same-state clipped Adam, and parameter-displacement bounds enclose work after an actual update (Appendices L and K). They specialize classical contraction and stability arguments, require rejection and gradient bounds, and are not descent guarantees. All unqueried suffix effects and candidate-formation costs remain in the comparison; the tested partial bounds decide no pair.

Why the cheaper estimators remain controls. Single-anchor sampling is unbiased before clipping and Adam, but introduces anchor and visit noise. We also tested a block-local estimator X with $\mathbb{E}X = l$ and an occasional complete correction: use X with probability $1 - q$ and $[g - (1-q)l]/q$ otherwise, with $q = 1/8$ fixed before fitting. This is an established variance-reduction construction (Beatson & Adams, 2019; Liu et al., 2019), not a new general operation. Neither unbiasedness before optimization nor a lower computation bill guarantees an effective trainer; the following experiments retain both negative results.

J FINITE CONTINUATION INFORMATION FOR FIXED-CLASS TRAINING

The motivating question is computational: can a sampled training anchor obtain useful continuation credit without evaluating vocabulary logits at every later anchor? Exact calculations check these identities, and GPU numerical examples test actual toy-tensor updates. These are not fitted empirical effectiveness results.

Allowed stopping actions change the comparison. The earlier work-separation theorems hold the proposal count and commitment rule fixed. With free discarding, they do not imply non-monotonicity of an optimal cap policy. For fields $q \geq p$, independently draw an integer cap $C \in \{0, \dots, K\}$ before observing proposals, with $\Pr(C \geq k) = \prod_{j \leq k} p_{a,j}/q_{a,j}$ at anchor a. These survivals decrease in k, so they define a probability law. Then $\Pr(\min(C, L_q) \geq k) = \prod_{j \leq k} p_{a,j}$: the stronger field exactly reproduces the weaker field's accepted-prefix transition. Zero products follow by continuity. An additional independent cap implements any weaker state-conditioned cap policy. With the same unit call costs, the optimal value over these policies is therefore nonincreasing under $q \geq p$. The argument requires a zero-proposal cap, free truncation and cap draws independent of draft outcomes. It concerns neither arbitrary controllers nor fixed learned policies or length-dependent hardware costs.

Model and notation. Retain the independent sampled parallel-head loop, fixed greedy target path, unit verification-call cost, and final no-draft call. There are M active anchors $0, \dots, M-1$, terminal value $V(M) = 0$, and at most K proposals. Write $r = K + 1$, $H_a = \min(K, M - a - 1)$, $A_{a,k} = \prod_{j=1}^{k} p_{a,j}$, and $P_a(j)$ for the probability of advancing j positions. Thus $P_a(j) = A_{a,j-1}(1 - p_{a,j})$ for $j \leq H_a$ and $P_a(H_a + 1) = A_{a,H_a}$. Pad a row to length r with zeros. Values beyond M are zero. The added prefill constant has no derivative.

J.1 BOUNDARY INFLUENCE, INCLUDING UNSAFE ACCEPTED-MASS REGIONS

Fix $0 \leq a < b \leq M$ and evaluate only the transition rows before b. For a start $s < b$, let c_s be the expected number of calls before first reaching $[b, b + K]$, and let q_s be its first-entry distribution on that set. For $s \in [b, b + K]$, put $c_s = 0$ and q_s equal to the corresponding coordinate vector. An independently specified boundary extrapolation $\widehat{v} = (\widehat{V}(b), \dots, \widehat{V}(b + K))$ gives

$$\widehat{V}(s) = c_s + q_s^\top \widehat{v}, \qquad V(s) - \widehat{V}(s) = q_s^\top e, \quad e = (V(b), \dots, V(b + K)) - \widehat{v}.$$

The distributions and pre-exit costs are obtained by an acyclic backward recursion using queried rows only. They are not fitted continuation models.

Intuition. Only a contrast between the ways two starts reach the unseen suffix can affect their difference in remaining work.

Proposition 10 (Exact boundary-credit representation)**.** *For $D_s = V(s) - V(s+1)$, the continuation error is*

$$D_s - \widehat{D}_s = (q_s - q_{s+1})^\top e.$$

At a sampled anchor a, define its unoccupied log-probability credit $C_{a,j} = -\sum_{k=j}^{H_a} A_{a,k} D_{a+k}$. Then

$$C_{a,j} - \widehat{C}_{a,j} = w_{a,j}^\top e, \qquad w_{a,j} = -\sum_{k=j}^{H_a} A_{a,k}(q_{a+k} - q_{a+k+1}), \qquad \mathbf{1}^\top w_{a,j} = 0. \tag{46}$$

Consequently, if $\operatorname{span}(e) \leq R_b$, then

$$|C_{a,j} - \widehat{C}_{a,j}| \leq \tfrac{1}{2} \|w_{a,j}\|_1 R_b. \tag{47}$$

For the relaxed boundary class containing every vector of span at most R_b, this bound is attained. This dual-norm sharpness does not assert that every such boundary vector is realizable by a legal decoding suffix. If instead coordinate bounds $l_z \leq e_z \leq u_z$ are known, its exact box maximum and minimum are obtained by selecting the appropriate endpoint according to the sign of $w_{a,j,z}$. Either valid bound may be used, or their minimum. These conclusions impose no condition $\sum_k A_{a,k} \leq 1$.

Proof. Stop the finite increasing chain at its first entry into the boundary set and condition on that entry. Unit costs before entry give the displayed decomposition. Subtract two adjacent starts and then the weighted credit expressions. Each q_s has total mass one, so every displayed contrast has sum zero. Subtracting the midpoint of the largest and smallest coordinates of e therefore leaves the inner product unchanged and proves the bound. For sharpness choose $e_z = R_b$ on positive coordinates of w and zero on negative coordinates (or reverse these choices). The box claim follows by separating the linear functional coordinate by coordinate.

A usable bound without a global probability floor. The linear extrapolation $\widehat{V}(z) = n_z := \max(M - z, 0)$ is the cost of unit advances in the unqueried suffix. Every legal loop satisfies

$$\lceil n_z/r \rceil \le V(z) \le n_z, \qquad \lceil n_z/r \rceil - n_z \le e_z \le 0.$$

Thus $R_b = \max_{b \le z \le b+K}(n_z - \lceil n_z/r \rceil)$ is valid, as is the looser $M - b$. The box bound is also available without querying a single suffix logit. At $b = M$ all residuals are exactly zero. Merely observing interior probabilities in the queried prefix does not justify a global probability-floor assumption. With no queried continuation rows, the linear extrapolation gives $\widehat{D}_s = 1$ at active positions, namely local-prefix credit; this is a baseline choice, not a newly invented objective.

Stochastic companion form. For a boundary error harmonic under the transition rows, write $E_a = (e_a, \dots, e_{a+K})^\top$. Then

$$E_a = C_a E_{a+1}, \qquad C_a = \begin{bmatrix} P_a(1) & P_a(2) & \cdots & P_a(r) \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}.$$

Every C_a is row stochastic. The span contraction coefficient of a row-stochastic matrix T is $\delta(T) = \frac{1}{2} \max_{i,j} \|T_{i,:} - T_{j,:}\|_1$. The exact adjacent-row contrast can be smaller than this worst-row maximum. Products of C_a recover the same first-entry contrasts as above. This uses an ordinary stochastic-kernel contraction, not an absolute-value contraction of $D_a = 1 - \sum_k A_{a,k} D_{a+k}$, whose row sum may exceed one.

Intuition. For a constant matching probability, one full proposal block erases exactly a fixed fraction of the suffix contrast, even where accepted mass exceeds one.

Theorem 8 (Exact fixed-noise influence law)**.** *For $K \ge 1$, on an uncapped homogeneous stretch with $p_{a,j} = t \in (0,1)$, let C be the companion matrix above and $\pi = (1, t, \dots, t^K)^\top/(1 + t + \cdots + t^K)$. Then*

$$C^r = t^r I + (1 - t^r)\mathbf{1}\pi^\top, \qquad \delta(C^{nr}) = t^{nr} \quad (n \ge 1). \tag{48}$$

In particular, homogeneous continuation-difference errors obey $D_a - \widehat{D}_a = t^r(D_{a+r} - \widehat{D}_{a+r})$ wherever these rows are available. At block-aligned depths L, the relaxed worst-case span-radius contribution is at most $\eta < R$ precisely when $t^L R \le \eta$, giving a depth of order $\log(R/\eta)/(-\log t)$. This is exact worst-case boundary-contrast attenuation; a particular credit may cancel earlier, and it is not an assertion that every fitted parameter-gradient error attains it.

The exponential rate cannot be replaced by finite-window exactness even with a fixed probability floor and a common exact-rank class. For $K = 2$, fix $\epsilon \in (0, 1/2)$, $t = 1 - \epsilon$, and $B \ge 3$ divisible by three. There are two fields in $[\epsilon, 1-\epsilon]$, equal through anchor $B-1$, whose root first-head probability derivatives differ by

$$(1 - 2\epsilon)t^{B-1}. \tag{49}$$

Both fields, and a neighborhood of their root-coordinate changes, have a shared-factor exact-rank-two realization with $p = m = 4$. The result bounds approximation error, not the signs of both derivatives: it does not retain the old shrinking-noise construction's constant opposite-sign margin. The indistinguishability is restricted to queried prefix probability rows. An estimator with only those observations incurs worst-case absolute error at least half the displayed gap. Globally known factors and representations can reveal the remote rows, so this is not a computation lower bound for white-box neural training or a proof of query-optimality of our implementation.

Proof. The homogeneous transition row is $((1-t), t(1-t), \ldots, t^{K-1}(1-t), t^K)$. Direct substitution gives $\pi^\top C = \pi^\top$. The characteristic polynomial factors as

$$(x-1)(x^K + tx^{K-1} + \cdots + t^K).$$

Its roots are the distinct numbers 1 and $t\omega$ for the nontrivial rth roots of unity ω. Hence C^r acts as the identity on the constant eigenspace and as t^r on its complementary invariant space. Its stationary projection is $\mathbf{1}\pi^\top$, proving the matrix identity. Row differences of its nth power are exactly t^{nr} times the corresponding row differences of I, proving the coefficient identity. Alternatively, subtract the neighboring homogeneous recursions to obtain $D_a - \widehat{D}_a = t^r(D_{a+r} - \widehat{D}_{a+r})$ directly.

For the legal lower witness set $M = B + 2$. Every active head before B has probability t. At B, change the first head from ϵ in one world to t in the other; its second head is inactive. The final anchor has zero draft horizon. Thus the difference of the two D_B values is $1 - 2\epsilon$, whereas the difference at $B+1$ is zero. Homogeneous propagation gives $\Delta D_0 = t^B(1-2\epsilon)$. Since $\partial J/\partial p_{0,1} = (D_0 - 1)/t$ when the root second head is fixed at t, Equation (49) follows. The required rows from 0 through $B-1$ all have horizon two.

For the rank realization use four frozen one-hot types: root, common body, boundary B, and final anchor. Put the final anchor's inactive first-head probability at ϵ in both worlds, and its inactive second head at t. With $c = \mathbf{1} \in \mathbb{R}^4$, binary unembedding $W = [0, c]$, and the typewise logit differences ℓ_1, ℓ_2, take $\Omega = [\ell_1 - c, \ell_2 - c]$, $v = c/4$, $\Gamma_1 = [1, 0]^\top v^\top$, and $\Gamma_2 = [0, 1]^\top v^\top$. The residual readout has exactly these logit differences. The second column of Ω is constant while the first distinguishes the final type from the body, so both factor ranks and the concatenated residual rank are two. A sufficiently small change of the first root-type entry preserves this rank. If $\log(t/(1-t)) = 1$, the constant second column in this particular residual parameterization vanishes; replace the fixed unembedding by $W = [0, 2c]$ and use $\Omega = [\ell_1/2 - c, \ell_2/2 - c]$ instead. The same construction then remains nondegenerate. These are pre-unembedding factor ranks, not an observable-law no-extra-kernel claim.

J.2 From credit uncertainty to a shared-parameter training operation

Let $\ell_{a,j}(\theta) = \log p_{a,j}(\theta)$ and retain the exact shared or isolated factor class. For any parameter-space direction u, put $s_{a,j} = \nabla_\theta \ell_{a,j}^\top u$. The conditional directional credit error at an anchor is exactly

$$\left(\sum_j s_{a,j} w_{a,j}\right)^\top e, \qquad |\text{error}| \le \tfrac{1}{2} \left\| \sum_j s_{a,j} w_{a,j} \right\|_1 R_b.$$

This combines the shared-parameter Jacobian *before* taking absolute values. It therefore permits cancellation but never assumes that an actual factor update increases every match probability. Summing the bound with the exact visit weights gives a full-gradient directional error bound. If a full deterministic approximate gradient $\widehat{g}$ and direction u obey $\widehat{g}^\top u + E(u) = -m < 0$, then the true directional derivative is negative. Under an explicitly valid L-smoothness bound along the parameter segment,

$$J(\theta + \alpha u) \le J(\theta) - \alpha m + \tfrac{1}{2} L \alpha^2 \|u\|^2.$$

Thus $0 < \alpha < 2m/(L\|u\|^2)$ decreases that objective. Ordinary Adam, a minibatch estimate, an unverified smoothness constant, or an uncontrolled rank-changing step does not automatically satisfy this claim. The finite bound is standard inexact-gradient calculus; the decoder-specific part is the explicit boundary coefficient propagated through its Jacobian.

Lazy sampled-anchor rule and its actual computational bill. Sample U uniformly from $\{0, \ldots, M-1\}$. Simulate the current increasing chain from zero until it reaches or overshoots U, querying only visited rows; let I_U indicate a reach. Conditional on a reach, query the anchor probabilities, extend its continuation window until the bound in Equation (47) is within a declared tolerance, and apply the vector–Jacobian product

$$\widehat{G} = \frac{M}{Z} I_U \sum_j \widehat{C}_{U,j} \nabla_\theta \ell_{U,j}.$$

Here Z is a declared, parameter-independent normalization: $Z = M$ gives work per active anchor, while $Z = M + 1$ includes the fixed initial emitted token in the response-length denominator. All sampling, occupancy, boundary coefficients and credits are stop-gradient objects in this expression. With exact credit, $\mathbb{E}[G] = \nabla_\theta J/Z$, because $\mathbb{E}[I_U \mid U = a] = d_p(a)$. Approximate credit introduces the explicit conditional bias above. If each coordinate error is at most η, then

$$\|\mathbb{E}[\widehat{G}] - \nabla J/Z\| \le \frac{\eta}{Z} \sum_a d_p(a) \sum_j \|\nabla_\theta \ell_{a,j}\|.$$

This is an expectation/bias statement, not descent of every sampled update. No per-prompt L1 normalization is inserted. For variable lengths and a length-based Z, averaging this estimator optimizes mean normalized conditional work, not mean unnormalized work; the estimand must be frozen explicitly. In particular the implementation's `cap` is $M + 1$, so its required coefficient is $M/\texttt{cap}$, not an implicit factor one. The implementation applies its tolerance η_{impl} *after* this scaling. Its corresponding bound is therefore $\eta_{\text{impl}} M^{-1} \sum_a d_p(a) \sum_j \|\nabla_\theta \ell_{a,j}\|$, not the preceding expression with an unscaled $\eta = \eta_{\text{impl}}$. The case $M = 0$ has zero gradient and is handled separately.

The operation can avoid suffix vocabulary projections and restrict autograd to one sampled anchor. Its bill still includes the forward path needed to sample occupancy, every queried continuation row, duplicate-query caching, and the sampled anchor's backward pass. Merely truncating the cheap scalar dynamic program after already computing all vocabulary logits saves none of that forward work. A tolerance forcing $b = M$, or high sampling variance, can erase the computational advantage. An exact-credit same-anchor control and an exact all-anchor optimizer are both needed to separate approximation bias, sampling variance and optimizer adequacy. No useful fitted outcome follows from the bound alone.

An optional unbiased continuation primitive. Two chains starting at s and $s + 1$ may share a cache of independently sampled row increments. Advance the currently lagging chain until it meets or overtakes the leading chain; stop at their first common anchor. The difference of their pre-meeting call counts has expectation D_s, because the remaining coupled trajectories then coincide. Before meeting their gap is at most K, and the chance that the next at most K lagging increments are all unit advances is at least ϵ^K if the *global* failure floor $P_a(1) \ge \epsilon > 0$ is known. Hence the expected number of row increments until meeting is at most K/ϵ^K (terminal meeting can only shorten it). Independent fresh randomness gives this bound conditional on each failed attempt; no row is revisited before a meeting. This primitive is unbiased but may have high variance and a poor worst-case work bound. It is not the adaptive deterministic-window estimator just defined, and must not be silently substituted into its experiment.

Safety is not equivalence of shared-parameter objectives. Even when every $S_p(a) \le 1$, local-prefix and complete-work gradients can oppose after parameter sharing. Take $K = 2$, $M = 4$, active probabilities $1/4$, and a scalar direction changing only first-head logits at anchors zero and one with derivatives 1 and $-9/8$. Local-prefix loss $-\sum_{a,k} A_{a,k}$ has directional derivative $15/512 > 0$, whereas complete work has derivative $-9/512 < 0$. All active expected prefix lengths are at most $5/16$. The discrepancy is caused by shared-parameter tradeoffs and unequal continuation/visit weights, not by a positive probability-coordinate derivative. A same-rank-two realization follows from the four-type construction above after making the final anchor's inactive first-head probability $3/4$. Consequently the small previously measured mass of positive coordinate credit does not prove that renewal and local training are equivalent. Conversely, this counterexample does not prove that their observed finite-data difference is caused by this mechanism.

Actual all-parameter SGD can disagree, including vocabulary biases. The preceding example concerns a specified tangent direction, not necessarily the full Euclidean gradient of its factorization. A separate construction makes this distinction explicit. Take $K = 2$, $M = 4$, $p = m = 2$, shared rank one, fixed native output map $W = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, and frozen representation rows

$$h_0 = (1, 64), \qquad h_1 = (1, -80), \qquad h_2 = h_3 = (1, 0).$$

The native target logit is one and the other native logit is zero at every anchor, so the fixed target is always token one. Use $\Omega = (-1, 0)^\top$, $\Gamma_1 = \Gamma_2 = (1, 0)$, and *both* trainable binary vocabulary

biases $b_j = (0, 0)$ for each head. The head logits are $(h_a + h_a\Omega\Gamma_j)W + b_j$. Every match probability is $1/2$; every active expected prefix length is at most $3/4$. Both factors and the concatenated residual have rank one.

Write the complete parameter coordinates as

$$(\Omega_0, \Omega_1, \Gamma_{1,0}, \Gamma_{1,1}, \Gamma_{2,0}, \Gamma_{2,1}, b_{1,0}, b_{1,1}, b_{2,0}, b_{2,1}).$$

For local loss $L = -\sum_{a,k} A_{a,k}$ and complete work J, ordinary Euclidean differentiation in *all ten* trainable coordinates gives

$$\nabla L = (-5/4, 8, 1, 0, 1/4, 0, 1, -1, 1/4, -1/4),$$
$$\nabla J = (-9/16, -1, 7/16, 0, 1/8, 0, 7/16, -7/16, 1/8, -1/8).$$

In particular,

$$\langle \nabla J, \nabla L \rangle = -377/64 < 0, \qquad \frac{d}{d\alpha} J(\theta - \alpha \nabla L)\Big|_{\alpha=0} = 377/64 > 0.$$

Continuity supplies sufficiently small positive steps that reduce L but increase J, with nonzero factors and the strict safe operating region preserved. The bias coordinates have not been frozen or quotiented away: their contribution to the gradient inner product is $15/16$, already included above. Fixed positive objective normalizations preserve the sign. This statement concerns the declared Euclidean factor/bias coordinates; arbitrary preconditioners, reparameterizations, or Adam do not inherit it. It is a finite, deliberately constructed instance of general objective-gradient conflict, not a new architectural theory breakthrough, an assertion about rank-108 training, or an explanation of the earlier mixed fitted outcomes.

Prior-work and novelty boundary. Stochastic-kernel contraction, inexact gradients, adaptive truncation of backpropagation, occupancy-weighted gradients, and regenerative/coupled estimators are existing techniques. In particular, Aicher et al.'s adaptive truncated-backpropagation work (Aicher et al., 2020) already trades estimated gradient bias against truncation depth; it cannot be renamed as our contribution. The bounded candidate increment here is the exact independent-decoder companion identity, a legal fixed-floor same-rank lower witness, and its first-entry boundary coefficients used to avoid actual vocabulary-row queries in a fixed-class training comparison. A broad novelty or practical effectiveness claim requires the nearest-work comparison and new experiment; this mathematical section supplies neither on its own. Regenerative equilibrium-gradient truncation bounds (Mahdian & Glynn, 2025) and coupled unbiased estimators (Jacob et al., 2019) are neighboring foundations, not priority claims of this finite-path specialization.

Current implementation and result boundary. The implemented sampler uses aligned chunks of eight head rows, including prefetch and vector–Jacobian recomputation in its physical bill. In the 12-fit GPU implementation test, all six objectives and both learning rates update their parameters; the maximum log-probability replay difference is zero. At each rate the adaptive sampler requests 68 logical rows against 88 for the exact same-anchor control, but both execute 520 physical head rows. Chunking erases the toy query saving, so this is not a runtime improvement. The completed two-model comparison uses two historically opened models, GSM8K and MBPP, three seeds, six objectives and two rates, with all rates chosen on a separate tuning split before evaluation. All 72 fits are complete. The registered bounded estimator fails the joint effectiveness and cost criterion; the exact sequence objective has favorable exploratory contrasts. The subsequent third-model confirmation and full cost accounting are reported in Appendix O. The native target path is stopped at EOS or 64 tokens, including short paths; its cache and unit-call endpoint are not complete native serving latency.

Optimization scope. The fixed-path bounds concern the specified probability field and optimizer map. They establish neither deployed path identity nor a causal explanation of fitted cross-entropy and response-work rankings.

K Bounding an Actual Parameter Update

An unbiased gradient estimate does not imply an unbiased clipped Adam step. We therefore bound the probabilities after the realized parameter change, including every shared factor and vocabulary

bias. This gives a finite-batch sufficient condition for improvement. The first implementation below fails to make that condition useful at partial query budgets.

Why the local-window witnesses are not arbitrary-query lower bounds. If the frozen rows have R known representation types, querying one full vocabulary law per type and head recovers the entire match field in at most RK row queries. Equal representations give equal head laws; known target labels select the required components. Scalar recursion then recovers all continuation credits. The arbitrary-window construction uses five types and the fixed-noise construction four, so these particular families admit $5K$ and $4K$ row-query upper bounds independent of response length. Grouping the coefficients by type also permits grouped vector-Jacobian products, whose arithmetic remains charged. Thus the witnesses separate prefix-local observations, not arbitrary adaptive querying. They do not rule out other hard families. More generally, public parameters and representations permit direct probability computation: charging named oracle calls while allowing equivalent arithmetic for free does not establish a white-box lower bound.

K.1 From factor changes to probability intervals

Keep the representation and unembedding fixed. At anchor a, head j has logits $z_{a,j} = [h_a + h_a\Omega\Gamma_j]W + b_j$. Let subscripts 0 and 1 denote a cached state and an actual proposed state. The latter can include clipping and the complete Adam moment history; the mathematical bound does not approximate that history. Define

$$t_{a,j} = h_a(\Omega_1\Gamma_{1,j} - \Omega_0\Gamma_{0,j}), \qquad \delta b_j = b_{1,j} - b_{0,j}, \tag{50}$$

$$R_W = \max_v \|W_v\|_2, \qquad \beta_{j,y} = \max_v |\delta b_{j,v} - \delta b_{j,y}|, \tag{51}$$

$$B_{a,j} = \|t_{a,j}\|_2(R_W + \|W_y\|_2) + \beta_{j,y}. \tag{52}$$

Here W_v is vocabulary column v and y is the fixed target token for this row. The finite factor difference includes $h_a\Delta\Omega\Delta\Gamma_j$ as well as the two first-order terms. It must not be replaced by the linearized logit change.

Intuition. A bound on how far target-relative logits move gives an interval for the new target probability, even at rows that were not queried.

Proposition 11 (Finite-step target-probability interval)**.** *For finite logits, let $p_{0,a,j}$ and $p_{1,a,j}$ be the old and new target probabilities. With σ the logistic function,*

$$\sigma(\operatorname{logit} p_{0,a,j} - B_{a,j}) \;\le\; p_{1,a,j} \;\le\; \sigma(\operatorname{logit} p_{0,a,j} + B_{a,j}). \tag{53}$$

Proof. Suppress the row indices and write $\delta z_v = tW_v + \delta b_v$. The definition of B gives $|\delta z_v - \delta z_y| \le B$ for every v. For the old non-target conditional law $w_v = p_{0,v}/(1 - p_{0,y})$, $v \ne y$, the exact odds identity is

$$\frac{1 - p_{1,y}}{p_{1,y}} = \frac{1 - p_{0,y}}{p_{0,y}} \sum_{v \ne y} w_v \exp(\delta z_v - \delta z_y). \tag{54}$$

The weighted sum lies in $[e^{-B}, e^{B}]$. Taking logarithms proves the claim.

This construction needs only the old target probability at each row, not storage of its complete law. Obtaining that old probability still costs a vocabulary computation. Computing R_W reads the frozen unembedding; $\beta_{j,y}$ requires a scan of the changed head bias and target-index lookup. Every anchor also incurs latent-space factor products. None of this work is free. A queried row can replace its interval by its new measured probability. Stable log odds and explicit numerical tolerances are needed in floating point; the mathematical statement is not a directed-rounding certificate.

K.2 Propagating intervals without assuming that more acceptance helps

Suppose each new match probability lies in $[\ell_{a,j}, u_{a,j}]$. The resulting rectangular set may contain fields that no single shared head can realize. We use that set as a conservative relaxation, not as the neural model class. Initialize $\underline{V}(M) = \overline{V}(M) = 0$. For either extremum and each anchor in

Table 15: Actual first-step changes and half-block query intervals for the batch mean, in expected calls per token. Each row is one shared-parameter update, not eight independent trials. A negative upper bound would establish improvement without complete querying; none is negative.

Model	Objective	True change	Interval width	Upper change
SmolLM3-3B	All-anchor local	-0.01871	0.36371	0.21342
	Whole response	-0.02552	0.35967	0.20575
OLMo-2-1B	All-anchor local	-0.01463	0.35005	0.22731
	Whole response	-0.02309	0.34289	0.21523

reverse order, set $R = V(a+H_a+1)$. For $j = H_a, \dots, 1$, replace R by the minimum or maximum, respectively, of

$$(1-\ell_{a,j})V(a+j)+\ell_{a,j}R, \qquad (1-u_{a,j})V(a+j)+u_{a,j}R. \tag{55}$$

Finally set $V(a) = 1 + R$. The same formula with no inner step handles $H_a = 0$.

Intuition. Choosing each interval endpoint according to its continuation cost encloses total work, including cases where a higher acceptance probability is harmful.

Proposition 12 (Finite-step work enclosure)**.** *The recurrence computes the exact minimum and maximum of $V(0)$ over the rectangular probability set in $O(MK)$ scalar operations. Consequently, for the actual new head and fixed positive normalization n,*

$$\frac{1+\underline{V}(0)}{n} \;\le\; F(\theta_1) \;\le\; \frac{1+\overline{V}(0)}{n}. \tag{56}$$

The additional one is the prefill call. Exactness concerns the rectangular relaxation, not its shared-parameter subset.

Proof. Conditioned on reaching the jth proposal, rejection moves to anchor $a+j$; acceptance continues to the next decision. The conditional continuation is $(1-p_j)V(a+j)+p_jR$. It is affine in p_j, so an endpoint attains its extremum. Its coefficient on R is nonnegative. Backward induction over proposals therefore gives the exact row extremum. The rows form an acyclic rectangular system: maximizing later rows simultaneously realizes their upper continuation values, and minimizing rows realizes their lower values. Backward induction over anchors completes the proof.

This is a specialization of robust dynamic programming (Iyengar, 2005), not a new general Bellman principle. Simply treating every upper probability as lower work would be incorrect outside the coordinatewise safe region. If the cached state is current, a new upper work bound below its exact old work proves improvement on that finite objective. With an older cache, the new upper bound must instead fall below a valid lower bound for the current state. For a fixed batch, normalize each prompt first and then average the intervals. Neither statement supplies population generalization.

K.3 The first implementation does not decide at partial budgets

We fixed the first four training paths from each corpus for each of two models, giving two eight-prompt batches. Each of the two objectives starts separately at the seed-zero warm checkpoint with empty fine-training Adam moments, takes one update at learning rate 10^{-4} and uses norm-5 clipping. No evaluation paths are read, and no updated checkpoint is saved. The mathematical bound permits general optimizer histories; this numerical study tests only the first step. All four measured gradients are below the clipping threshold.

The new-row query schedules use zero, one quarter, one half or all of the source-ordered eight-anchor blocks, fixed before the new fields are measured. Bounds use the old cache, actual factor changes and only the designated new rows. Complete new fields are acquired separately as a charged reference, after the bounds are fixed. The full reference is not available for choosing the earlier query schedules or tightening their intervals.

All four updates reduce actual work, but no batch improvement is established at zero, quarter or half budgets. Full querying recovers all four improvements. The same partial-query failure holds for every constituent path; paths within a batch share the update and are not independent runs. Probability

and work inclusion pass the fixed tolerances of 2×10^{-6} and 2×10^{-8}, with a 2×10^{-5} log-odds margin. An independent differentiable-batch gradient has maximum coordinate discrepancy below 4.2×10^{-8}. Separate synthetic checks enumerate 2,358 corners across 36 probability boxes, with maximum recursion discrepancy below 8.9×10^{-16}. These are numerical checks, not effectiveness evidence.

For each model the old full cache costs 1,936 physical head rows and stores 16,128 bytes of target log probabilities, besides the matching parameter state. The nominal half-block query costs 1,024 physical head rows; it is not exactly half the active row count. Computing the actual full gradient costs 3,872 forward and recomputation rows. The unembedding norm scan reads about 1.05 GB for SmolLM3 and 0.82 GB for OLMo. Factor work, bias scans, the independent gradient reference and the final true-field calculation are charged separately in the released results. Table 15 gives the actual first-step changes beside the half-block query intervals, one row per shared-parameter update. The tested interval is valid within the declared numerical tolerances but too wide to support a cheap update decision. This negative result rules out the tested global-norm bound as that practical solution, not every possible partial-query method.

L Heterogeneous Fields and Two-Sided Query Windows

Distinct representation rows alone do not imply a length-linear query lower bound. Under a known rejection floor, heterogeneous transition rows still forget distant boundary conditions. We use classical stochastic-kernel contraction (Sethuraman & Varadhan, 2004, Section 1) to bound both occupancy and continuation credit, then include every shared-parameter gradient contribution by uniform anchor sampling. These are conditional query upper bounds, not evidence of a useful fitted estimator or a white-box computational lower bound.

L.1 Model and explicit oracle access

Fix $K \geq 1$, $n = M + 1$, and the same target path and unit-call convention as Equation (2). Anchors are $0, \ldots, M-1$, M is terminal, and $H_a = \min(K, M - a - 1)$. With $A_{a,k} = \prod_{j=1}^{k} p_{a,j}$ and $A_{a,0} = 1$, the jump law is

$$P_a(j) = A_{a,j-1}(1 - p_{a,j}) \quad (1 \leq j \leq H_a), \qquad P_a(H_a + 1) = A_{a,H_a},$$

padded with zeros through $K + 1$. Thus jumps never pass M, the final anchor has no proposal, $V(M) = 0$, and

$$V(a) = 1 + \sum_{j=1}^{K+1} P_a(j) V(a + j), \qquad F = (1 + V(0))/(M + 1).$$

Values beyond M are zero. Prefill and the fixed target length have zero derivative. Both $M = 0$ and $M = 1$ have zero gradient.

Assume known constants $0 < \eta < 1$ and $B < \infty$ such that, at the current parameters,

$$P_a(1) \geq \eta, \qquad \psi_{a,j} = \nabla_\theta \log p_{a,j}, \qquad \|\psi_{a,j}\|_2 \leq B. \tag{57}$$

Active match probabilities are differentiable and in $(0, 1)$. All trainable factor and vocabulary-bias coordinates belong to the same declared Euclidean parameter space. The floor is a promise on the entire field, not something certified by querying a small subset. No homogeneity, stationary law, repeated representation, or condition $S_p \leq 1$ is assumed.

The fixed path permits random indexed access. A probability-anchor query returns all its active match probabilities; a separately charged local vector-Jacobian product returns $\sum_j c_j \psi_{a,j}$ for supplied coefficients. Both answers are exact in this oracle. No occupancy, continuation, global probability cache, or unqueried Jacobian is supplied for free. Public parameters and representations define a different information model: evaluating their known formulas cannot be excluded by merely renaming a row query. The operation below can use local neural evaluation, but its query count does not bound all input, cache, or arithmetic costs.

L.2 A COUNTDOWN CHAIN CONTROLS BOTH BOUNDARIES

Let e_r be the unit column at state $r \in \{0, \dots, K\}$. Define

$$Q_a(0, j-1) = P_a(j), \qquad Q_a(r, r-1) = 1 \quad (1 \le r \le K). \tag{58}$$

At state zero a verifier call samples a jump; positive states count down without a call. Spatial position advances by one per matrix step. Thus

$$d(a) = e_0^\top Q_0 \cdots Q_{a-1} e_0, \qquad W_b = \sum_{t=b}^{M-1} Q_b \cdots Q_{t-1} e_0, \tag{59}$$

where empty products are identities. Starting in state r postpones the first call until $b+r$, so $W_b(r) = V(b+r)$, including zero beyond the terminal. In particular $W_b = e_0 + Q_b W_{b+1}$ and $W_M = 0$.

Write $\delta(T) = \frac{1}{2} \max_{r,s} \|T(r,:) - T(s,:)\|_1$ and set

$$\rho = 1 - \eta^K, \qquad \alpha(t) = \rho^{\lfloor t/K \rfloor}, \qquad C_\eta = K/\eta^K. \tag{60}$$

Intuition. A bounded countdown and a positive chance of rejecting the first draft repeatedly erase information about both the preceding visit distribution and the remote suffix.

Lemma 2 (Heterogeneous contraction and legal value span)**.** *Every available t-step consecutive product satisfies*

$$\delta(Q_b \cdots Q_{b+t-1}) \le \alpha(t).$$

Every legal boundary column W_b has $\operatorname{span}(W_b) \le C_\eta$. The same span bound holds when rewards are stopped at an earlier boundary with zero terminal values.

Proof. In any K steps, a countdown starting at r reaches zero after r deterministic steps. Repeatedly choosing unit jumps for the remaining steps ends at zero with probability at least $\eta^{K-r} \ge \eta^K$. Thus every row of the product has a common column-zero mass of at least η^K. Removing this common mass bounds its contraction coefficient by ρ. For any row-stochastic T, $\operatorname{span}(Tx) \le \delta(T) \operatorname{span}(x)$: the positive and negative masses of a difference of rows are equal to its total-variation norm. This also gives $\delta(TU) \le \delta(T)\delta(U)$ and $\mathrm{TV}(\mu T, \nu T) \le \delta(T)\,\mathrm{TV}(\mu, \nu)$. Partition into $\lfloor t/K \rfloor$ full blocks to obtain the first claim. The reward representation then yields

$$\operatorname{span}(W_b) \le \sum_{s=0}^{M-b-1} \alpha(s) \le K \sum_{q=0}^{\infty} \rho^q = C_\eta.$$

Earlier truncation only shortens the sum. No common invariant distribution or commutativity was used.

This controls a value contrast, not the absolute continuation value, which can grow with length. Without the global promise in Equation (57), it does not replace the length-dependent fallback of Appendix J.

L.3 LOCAL WINDOWS AND THE FULL SHARED GRADIENT

Choose integers $L, R \ge 0$. At a central anchor a, query only

$$\ell_a = \max(0, a - L), \dots, b_a - 1, \qquad b_a = \min(M, a + K + R).$$

Set $\widetilde{d}(a) = e_0^\top Q_{\ell_a} \cdots Q_{a-1} e_0$. This is exact if $\ell_a = 0$; otherwise its error is at most $\alpha(L)$ by the total-variation contraction. This cold start replaces neither the true occupancy by a free oracle nor a charged full-prefix simulation.

Set $\widetilde{V}(s) = 0$ for $s \ge b_a$ and run the queried Bellman recursion backwards to a. Write $D_s = V(s) - V(s+1)$ and $\widetilde{D}_s = \widetilde{V}(s) - \widetilde{V}(s+1)$. For $s \le b_a$,

$$D_s - \widetilde{D}_s = (e_0 - e_1)^\top Q_s \cdots Q_{b_a - 1} W_{b_a}. \tag{61}$$

Indeed the error column propagates homogeneously from boundary W_{b_a}, and its first two entries are the adjacent-start errors. For $s = b_a$ the product is empty. If $b_a = M$ the error is zero; otherwise

$$|D_s - \widetilde{D}_s| \le C_\eta \alpha(b_a - s).$$

All required $s = a + k$, $k \le H_a \le K$, have $b_a - s \ge R$ when the boundary is not terminal. Equivalently the signed advantage is

$$V(a + k + 1) - V(a + k) = W_{a+1}(k) - W_{a+1}(k - 1).$$

At that common position its error is a row contrast of $Q_{a+1} \cdots Q_{b_a - 1}$, with $K + R - 1 \ge R$ transitions. Thus the indexing provides at least R mixing steps even for the last advantage. Both true and truncated adjacent differences have magnitude at most C_η.

Define

$$C_{a,j} = -\sum_{k=j}^{H_a} A_{a,k} D_{a+k}, \qquad G_{a,j} = d(a) C_{a,j},$$
$$\widetilde{C}_{a,j} = -\sum_{k=j}^{H_a} A_{a,k} \widetilde{D}_{a+k}, \qquad \widetilde{G}_{a,j} = \widetilde{d}(a) \widetilde{C}_{a,j}. \tag{62}$$

Let $e_L(a) = 0$ if $\ell_a = 0$ and $\alpha(L)$ otherwise; let $e_R(a) = 0$ if $b_a = M$ and $\alpha(R)$ otherwise. Also put

$$T_K = K(K + 1)/2, \qquad R_g = B C_\eta T_K, \qquad h_a = \sum_j G_{a,j} \psi_{a,j}, \quad \widetilde{h}_a = \sum_j \widetilde{G}_{a,j} \psi_{a,j}.$$

Intuition. Charging a left occupancy window and a right credit window bounds the complete shared-parameter contribution of each anchor, including parameters that affect many other anchors.

Proposition 13 (Two-sided full-gradient coefficient bound)**.** *The exact normalized gradient is*

$$g = \nabla_\theta F = \frac{1}{M + 1} \sum_{a=0}^{M-1} h_a.$$

For every active (a, j)*,*

$$|G_{a,j} - \widetilde{G}_{a,j}| \le (H_a - j + 1) C_\eta [e_L(a) + e_R(a)],$$
$$\|h_a\|_2, \|\widetilde{h}_a\|_2 \le R_g,$$
$$\|h_a - \widetilde{h}_a\|_2 \le R_g [e_L(a) + e_R(a)].$$

Consequently, with $\widetilde{g} = (M + 1)^{-1} \sum_a \widetilde{h}_a$*,*

$$\|g - \widetilde{g}\|_2 \le \frac{R_g}{M + 1} \sum_a [e_L(a) + e_R(a)] \le R_g [\alpha(L) + \alpha(R)].$$

Proof. Summation by parts gives $V(a) = 1 + V(a + 1) - \sum_k A_{a,k} D_{a+k}$. Its direct row derivative with respect to $\log p_{a,j}$ is $C_{a,j}$. The acyclic chain multiplies this by the probability $d(a)$ of reaching the row. Summing every row's derivative and applying the parameter chain rule proves the gradient identity; no unqueried direct effect is frozen.

Since $A_{a,k} \le 1$, both $|C_{a,j}|$ and $|\widetilde{C}_{a,j}|$ are bounded by $(H_a - j + 1) C_\eta$, and their difference is bounded by the same quantity times $e_R(a)$. Use

$$G_{a,j} - \widetilde{G}_{a,j} = [d(a) - \widetilde{d}(a)] C_{a,j} + \widetilde{d}(a) [C_{a,j} - \widetilde{C}_{a,j}],$$

with $0 \le \widetilde{d}(a) \le 1$, to obtain the coefficient bound. The vector bounds follow from $\|\psi_{a,j}\|_2 \le B$ and $\sum_j (H_a - j + 1) \le T_K$. Summing and dividing by $M + 1$ completes the proof. These bounds permit shared-parameter cancellation.

L.4 A RANDOMIZED QUERY UPPER BOUND, NOT A PRACTICAL TRAINER

Intuition. Uniformly sampling anchors includes every shared-parameter effect in expectation once the magnitude and boundary bias of each contribution are controlled.

Theorem 9 (Fixed-accuracy randomized row queries)**.** *Suppose $M \geq 1$. Sample m anchors U_i independently and uniformly from $\{0, \dots, M-1\}$, compute only their charged windows and central local vector-Jacobian products, and return*

$$\widehat{g}_m = \frac{1}{m} \sum_{i=1}^{m} \frac{M}{M+1} \widetilde{h}_{U_i}.$$

For $0 < \delta < 1$, with probability at least $1 - \delta$,

$$\|\widehat{g}_m - g\|_2 \leq R_g[\alpha(L) + \alpha(R)] + \frac{R_g}{\sqrt{m\delta}}. \tag{63}$$

For $R_g > 0$ and $0 < \varepsilon < R_g$, the choices

$$L = R = K \left\lceil \frac{\log(4R_g/\varepsilon)}{-\log \rho} \right\rceil, \qquad m = \left\lceil \frac{4R_g^2}{\delta \varepsilon^2} \right\rceil \tag{64}$$

give error at most ε. Their bounds are independent of M when $K, \eta, B, \varepsilon, \delta$ are fixed uniformly in M.

Proof. For $X_i = M\widetilde{h}_{U_i}/(M+1)$, the target-length factor gives $\mathbb{E}X_i = \widetilde{g}$ and $\|X_i\|_2 \leq R_g$. Independence and the Euclidean variance identity imply

$$\mathbb{E}\|\widehat{g}_m - \widetilde{g}\|_2^2 = \frac{1}{m}\mathbb{E}\|X_1 - \widetilde{g}\|_2^2 \leq R_g^2/m.$$

Markov's inequality for the squared norm and the deterministic bias bound give Equation (63). The specified windows make its first term at most $\varepsilon/2$, and the sample size makes its second term at most $\varepsilon/2$. All coefficients are stop-gradient inputs to the local vector-Jacobian product. The estimator is unbiased for $\widetilde{g}$, not generally for g.

If additionally $\|g\|_2 \geq g_0 > 0$ uniformly in M, choosing $\varepsilon = \gamma g_0$, $0 < \gamma < 1$, gives $g^\top \widehat{g}_m > 0$ and $\angle(g, \widehat{g}_m) \leq \arcsin \gamma$ on the successful event. The first claim follows from $g^\top \widehat{g}_m \geq \|g\|_2(\|g\|_2 - \varepsilon)$; the distance from g to the ray through $\widehat{g}_m$ then bounds the angle's sine. This margin assumption is substantive. With one independent logit per anchor, normalized gradients can have order $M^{-1/2}$ norm even for fixed interior probabilities. The theorem does not exclude length-dependent queries for vanishing margins or exact answers.

Complete computational bill. There are at most $m(L + K + R)$ probability-anchor requests, hence at most $mK(L + K + R)$ evaluated head rows, plus at most mK central backward head rows and any mK forward rows needed for their recomputation. Duplicate caching is optional and valid only at unchanged parameters. Physical chunk padding, not just logical requests, must be charged. The sparse scalar recursions cost $O(mK(L+K+R))$; dense gradient accumulation costs $O(mD)$ and $O(D)$ storage for D parameter coordinates, in addition to the local vector-Jacobian products. Full-vocabulary normalization and head arithmetic depend on representation dimension, factor rank and vocabulary size. Creating the target path and random-access representation cache, loading parameters, and checking global promises are separate costs. Thus the dimension-free row bound is not a dimension-free or end-to-end runtime bound. Its constants can be too large to save any work at the experimental cap.

Even the displayed sample-size certificate is not operationally small at that cap. Since $\|g\|_2 \leq R_g$, consider the optimistic thought experiment $\varepsilon = 0.1\|g\|_2$ and $\delta = 0.05$, without assuming that the oracle knows $\|g\|_2$. Equation (64) then prescribes

$$m \geq \frac{4R_g^2}{0.05 \cdot 0.01\|g\|_2^2} \geq 8000.$$

Using a guaranteed $g_0 \le \|g\|_2$ instead cannot reduce this sufficient budget. Our target-length cap is 64, hence $M \le 63$; the stated certificate therefore supplies no nontrivial row-saving guarantee at these lengths, even before windows, vector-Jacobian products, and cache costs. Repeated samples may share queried rows. The number 8000 is a conservative sufficient sample budget from this proof, not a lower bound on necessary samples or distinct queries, and it does not exclude a sharper estimator or analysis.

L.5 Qualified transfer to one actual optimizer map

For alignment with stabilized norm clipping, allow any $\tau \ge 0$ and define

$$P_{C,\tau}(x) = x \min\{1, C/(\|x\|_2 + \tau)\}, \qquad P_{C,\tau}(0) = 0,$$

where $C > 0$ and the zero convention includes $\tau = 0$. The latter case is Euclidean projection onto the radius-C ball. This stabilized map is also nonexpansive. Indeed its radial magnitude $f(r) = r \min\{1, C/(r+\tau)\}$ obeys $0 \le f(r) \le r$ and is one-Lipschitz: its derivative is one in the identity region and $C\tau/(r+\tau)^2 \le 1$ in the shrinking region, with continuity at the boundary. For unit vectors u, v,

$$\|f(r)u - f(s)v\|_2^2 = [f(r) - f(s)]^2 + f(r)f(s)\|u - v\|_2^2 \le \|ru - sv\|_2^2.$$

Zero inputs follow by continuity, including $C < \tau$.

Fix identical pre-step parameters, moments $m_{\mathrm{prev}}, v_{\mathrm{prev}}$, step $t \ge 1$, $\beta_1, \beta_2 \in [0, 1)$, learning rate λ, and $\epsilon_A > 0$. Assume $v_{\mathrm{prev}} \ge 0$ coordinatewise. For $c = P_{C,\tau}(g)$ write

$$A = \frac{\beta_1 m_{\mathrm{prev}}}{1 - \beta_1^t}, \qquad b = \frac{1 - \beta_1}{1 - \beta_1^t},$$
$$q_i = \frac{\beta_2 v_{\mathrm{prev},i}}{1 - \beta_2^t}, \qquad d = \frac{1 - \beta_2}{1 - \beta_2^t}.$$

Standard bias-corrected Adam with epsilon outside the square root has the mathematical increment

$$U_i(g) = -\lambda \frac{A_i + b c_i}{\sqrt{q_i + d c_i^2} + \epsilon_A}.$$

This statement assumes zero weight decay and excludes AMSGrad and floating-point effects; it requires identical handling of every gradient coordinate, including inactive tensors.

Intuition. At one shared optimizer state, a global continuity bound can transfer gradient accuracy through clipping and Adam without asserting that the resulting step decreases work.

Corollary 2 (Same-state clipped-Adam stability)**.** *The mathematical update satisfies*

$$\|U(g) - U(g')\|_2 \le L_t \|g - g'\|_2, \qquad L_t = |\lambda| \left(\frac{b}{\epsilon_A} + \frac{\sqrt{d}\,\|A\|_\infty}{\epsilon_A^2} \right). \tag{65}$$

No positive lower bound on v_{prev} is required.

Proof. Fix a coordinate and put $s(x) = \sqrt{q_i + dx^2}$, $D(x) = s(x) + \epsilon_A$. Reverse triangle inequality for $(\sqrt{q_i}, \sqrt{d}\,x)$ gives $|s(x) - s(y)| \le \sqrt{d}|x - y|$. Since $D \ge \epsilon_A$, the reciprocal $1/D$ is $\sqrt{d}/\epsilon_A^2$-Lipschitz. Moreover $h(x) = x/D(x)$ is globally $1/\epsilon_A$-Lipschitz. For $q_i > 0$,

$$h'(x) = \frac{q_i/s(x) + \epsilon_A}{(s(x) + \epsilon_A)^2} \in [0, 1/\epsilon_A],$$

using $q_i/s(x) \le s(x)$. For $q_i = 0$, differentiate $x/(\sqrt{d}|x| + \epsilon_A)$ on either side of zero; its derivative at zero also exists and is $1/\epsilon_A$. Decomposing the update as $-\lambda[A_i/D(x) + bh(x)]$ gives its coordinate Lipschitz bound. Squaring and summing uses the maximum coordinate constant, then composition with the nonexpansive $P_{C,\tau}$ proves the result.

If $m_0 = 0$ and all previous gradients were clipped at the same radius C, their corrected prior weight gives $\|A\|_\infty \le (1-b)C$. Since $b, d \le 1$,

$$L_t \le |\lambda| \left[\frac{b}{\epsilon_A} + \frac{(1-b)C\sqrt{d}}{\epsilon_A^2} \right] \le |\lambda| \max\{\epsilon_A^{-1}, C\epsilon_A^{-2}\}.$$

Thus the event $\|\widehat{g}_m - g\|_2 \le \varepsilon$ implies a one-step parameter discrepancy at most $L_t\varepsilon$. Length-independent absolute update accuracy additionally needs a length-uniform L_t; small ϵ_A can make the bound uninformative. This is neither optimizer unbiasedness, work decrease, a useful decision at partial budgets, nor a trajectory comparison after the two states diverge. It supplies no fitted method guarantee. Kernel contraction, sample-mean bounds, and this elementary optimizer stability argument are not claimed as new general principles.

M Comparing Actual Updates from the Same Optimizer State

A decrease relative to doing nothing does not establish an advantage over local training. We instead compare two finite updates formed from identical parameters and Adam history. The comparison is conditional on a fixed target path; a second training batch tests whether the observed advantage transfers within the available development pool. Neither calculation alone establishes population generalization or an economical training method.

M.1 A finite comparison that includes the unknown suffix

Use the capped countdown kernels $Q_{X,a}$ and values $W_{X,a}$ of Appendix L, where $X = C, L$ denotes the whole-response candidate or the local candidate, respectively. Both are realized post-update states, including clipping, all Adam moments, both factor changes and vocabulary biases. Write $q_{X,a} = Q_{X,a}(0,:)$, $d_C(a) = [e_0^\top Q_{C,0} \cdots Q_{C,a-1}]_0$, and $n = M+1$. Subtracting the two Bellman recursions and telescoping gives

$$F_C - F_L = \frac{1}{n} \sum_{a=0}^{M-1} d_C(a)(q_{C,a} - q_{L,a}) W_{L,a+1}. \tag{66}$$

This is the classical performance-difference identity specialized to the capped decoder (Schulman et al., 2015), not a new policy-improvement theorem. It compares actual states without approximating the optimizer map, but forming those states is still charged.

For a queried prefix $0, \dots, b-1$, let $\mu_{X,b}$ be its outgoing countdown distribution and $R_{X,b}$ its cumulative expected calls. Starting with $\mu_{X,0} = e_0^\top$ and $R_{X,0} = 0$, compute

$$R_{X,a+1} = R_{X,a} + \mu_{X,a}(0), \qquad \mu_{X,a+1} = \mu_{X,a} Q_{X,a}.$$

The unqueried tail obeys the exact decomposition

$$n(F_C - F_L) = R_{C,b} - R_{L,b} + \mu_{C,b} W_{C,b} - \mu_{L,b} W_{L,b}. \tag{67}$$

Thus no unqueried position is assigned zero credit or omitted.

For countdown state r, define

$$s_r = (M - b - r)_+, \quad \ell_r = \lceil s_r/(K+1) \rceil, \quad u_r = s_r, \quad w_r = u_r - \ell_r.$$

Every legal tail value belongs to $[\ell_r, u_r]$: each call advances between one and $K+1$ positions. If separately obtained bounds satisfy $\mathrm{TV}(q_{C,a}, q_{L,a}) \le \tau_a \le 1$, set

$$e_r = \min \left\{ w_r, \sum_{a=\min(M, b+r)}^{M-1} \tau_a (M - a - 1) \right\}.$$

Applying Equation (66) to each suffix, using $d_C(a) \le 1$ and a remaining-value span at most $M - a - 1$, gives $|W_{C,b}(r) - W_{L,b}(r)| \le e_r$. The default $\tau_a = 1$ requires no probability access beyond the queried prefix.

Intuition. Comparing the two tails jointly can retain their common cost, but every unknown position still contributes to the uncertainty.

Proposition 14 (Complete paired-prefix enclosure)**.** *Put* $p_r = \mu_{C,b}(r)$, $q_r = \mu_{L,b}(r)$ *and* $\delta_r = p_r - q_r$. *Define*

$$a_r = \delta_r \ell_r - q_r e_r + \min(0, \delta_r)(w_r - e_r),$$
$$z_r = \delta_r \ell_r + p_r e_r + \max(0, \delta_r)(w_r - e_r).$$

Then

$$\frac{R_{C,b} - R_{L,b} + \sum_r a_r}{n} \le F_C - F_L \le \frac{R_{C,b} - R_{L,b} + \sum_r z_r}{n}.$$

The endpoints are exact over the independent rectangular tail relaxations with the stated pairwise differences, not over all realizable neural tails. At $b = M$ *both bounds equal the actual difference. The cases* $M = 0, 1$ *and zero width require no active draft probabilities and are included.*

Proof. For fixed r, optimize $p_r x - q_r y$ subject to $\ell_r \le x, y \le u_r$ and $|x - y| \le e_r$. For the upper endpoint, take $x = y + e_r$ and place the remaining common component at its lower or upper endpoint according to the sign of δ_r. This gives z_r; the symmetric construction gives a_r. The same formulas cover zero masses or zero width. Sum these coordinate extrema in Equation (67). The relaxation does not impose that different boundary coordinates share the same future neural kernels, so its sharpness does not assert neural attainability.

An alternative window calculation propagates the signed row difference itself. For queried local kernels after anchor a up to boundary b, put $h = (q_{C,a} - q_{L,a}) Q_{L,a+1} \cdots Q_{L,b-1}$. The omitted term is $hW_{L,b}$ and $h\mathbf{1} = 0$. If the tail span is at most B_b, then $|hW_{L,b}| \le B_b \|h\|_1 / 2$. Equality holds over arbitrary vectors of that span. This can be tighter than separately bounding the two candidate values, but applying overlapping windows at every anchor can still acquire the complete field.

Paid parameter envelope. For hidden row H, target item y, frozen output columns W_v, and $t = (H\Omega_C)\Gamma_{C,j} - (H\Omega_L)\Gamma_{L,j}$, let $\Delta b = b_{C,j} - b_{L,j}$. The bound

$$B_{a,j} = \|t\|_2 \left(\max_v \|W_v\|_2 + \|W_y\|_2 \right) + \max\{ |\min_v \Delta b_v - \Delta b_y|, |\max_v \Delta b_v - \Delta b_y| \}$$

controls every target-relative logit change. Consequently $|p_{C,a,j} - p_{L,a,j}| \le \tanh(B_{a,j}/4)$ and a coupling of the prefix decisions gives

$$\mathrm{TV}(q_{C,a}, q_{L,a}) \le \min\{1, \sum_j \tanh(B_{a,j}/4)\} =: \tau_a.$$

This uses complete finite factor products, not a first-order approximation. The output-column norm scan, both factor evaluations, target-column norms, bias scans and transfers are separately charged. The implementation uses FP64 algebra on stored FP32 parameters with the fixed 2×10^{-5} log-odds margin. Inclusion checks allow 2×10^{-8}; these are numerical tests, not directed-rounding certificates.

Fixed acquisition schedule and exact checks. Both the unconditional and paid-envelope comparisons acquire the first quarter, half and all of the eight-anchor blocks, with upward block rounding. A callback supplies only requested rows, retains their union, and never receives true future values or visitation weights. Dense symbolic rows, actual active head rows and complete candidate-formation costs are distinct. Both actual candidates have already required complete gradients; a cheaper comparison cannot remove that cost. Exact rational checks cover 96 field pairs, 3,392 signed windows and 26,410 checks, with zero equality residual. Floating comparisons have maximum discrepancy below 2.3×10^{-16} on these finite instances. This is not a guarantee for every floating input.

M.2 A RESERVED-TRAINING COMPARISON OF TWO ACTUAL STEPS

Execution chronology. The initial implementation incorrectly assumed 384 training paths per corpus and stopped before forming a new history or candidate. The immutable roster has 512 GSM8K and 256 MBPP paths. A correction changed this assertion and its synthetic test, not the reserved paths, optimizer, candidates or analysis. The corrected implementation completed all six runs before any paired outcome was read. Loading and the output-norm scan in the first execution were not zero-cost; its failure was not a scientific null.

Candidate evaluation on a separate batch is not a new general technique: safe policy improvement separates candidate construction from assessment (Thomas et al., 2015), and validation-gradient reweighting uses another batch to guide training (Ren et al., 2018). Our diagnostic instead asks whether continuation information adds value relative to an all-anchor local step at exactly the same Adam state.

For each model and each of three warm seeds, reserve the first 64 training paths per corpus, in original source order. Partition each corpus into four groups of 16; their first four paths contribute to an eight-path update batch A, and their other twelve contribute to a 24-path probe batch B. The 128 reserved paths are excluded from both new 32-step histories. They were used in the older warm-start training, so B is not untouched data.

Use three anchor states per model and seed: the warm head with empty fine Adam moments, and separate 32-step local and whole-response histories on the remaining 640 training paths. Both histories use the same batch schedule. At each of the 18 states, form two disposable candidates for each of the four batch pairs, giving 72 paired comparisons. Copy all parameter tensors and every Adam first moment, second moment and step counter before branching. Candidates advance from step zero to one or from step 32 to 33. Neither candidate propagates into a history. Both post-update models are fixed before reading their probe values or the separately charged probe gradient.

The learning rate is 10^{-4}, the batch size eight, and all other head, precision, normalization, clipping and Adam settings match the all-anchor comparison. The complete anchor optimizer states are saved. Independent differentiable objective calculations agree with 18 synthetic candidate cases, including inactive gradients, nonempty history and active clipping; the maximum parameter discrepancy is below 2.2×10^{-11}. Both paths use PyTorch Adam, so this is not an independent implementation of Adam itself.

For each pair record

$$D_A = F_A(\theta + u_C) - F_A(\theta + u_L), \qquad D_B = F_B(\theta + u_C) - F_B(\theta + u_L).$$

Negative values favor the whole-response candidate. We also record both candidates relative to the anchor and the whole-gradient inner product with $u_C - u_L$ on each batch. The sign tolerance is 10^{-8} calls per token, not a practical-benefit threshold. Repeated prompts, histories and seeds make these dependent development contrasts. No population interval, independence of the 72 comparisons, deployment gain or trained-method superiority follows from their directions.

Complete results. All 72 pairs and 2,304 prompt-level measurements are retained. With empty fine-training Adam moments, the whole-response candidate has lower work on A in all 24 pairs, but only one of those pairs retains that direction on B. In the 48 comparisons with nonempty moments, local training has lower work on A in every pair; it also has lower work on B in 43. Local training has lower work on B in 66 of 72 comparisons. Every model, history and corpus mean on B favors local training (Table 16). This does not establish a population ranking. The parameters as well as moments differ between history groups, so their contrast does not isolate the causal effect of Adam history.

First-order directions and finite changes. Define $L_X = \nabla F_X(\theta)^\top (u_C - u_L)$ and $R_X = D_X - L_X$ for $X = A, B$. The recorded terms satisfy the exact descriptive decomposition

$$D_B - D_A = (L_B - L_A) + (R_B - R_A).$$

Table 17 reports all six model-history means. The sign of L_X agrees with that of D_X in all 72 update-batch comparisons and 71 of 72 probe comparisons. At the empty-history states, the first-order direction itself favors whole-response training on A and local training on B in all 24 pairs. Thus a nonlinear remainder is not necessary for that sign change. This identity does not identify causal contributions or a percentage of benefit explained. Both full candidate gradients and the probe gradient were paid for; the last was measured only after both candidates were fixed. Clipping occurs in none of the 144 candidate or 384 preceding-history steps.

Partial acquisition still does not decide the comparison. Neither unconditional nor parameter-based bounds decide any pair on A or B at the quarter or half block budgets (Table 18). Full acquisition resolves all pairs. The complete numerical analysis checks 13,824 path enclosures and

Table 16: Actual whole-response minus local work, in 10^{-3} calls per token. A forms the update; B does not. Each entry averages four pairs and three seeds within the indicated corpus. These are dependent development means without a population confidence interval.

Model	Prior updates	Corpus	D_A (10^{-3})	D_B (10^{-3})
SmolLM3-3B	None	GSM8K	-8.779	+0.177
SmolLM3-3B	None	MBPP	-6.182	+0.879
SmolLM3-3B	32 local	GSM8K	+1.180	+0.025
SmolLM3-3B	32 local	MBPP	+1.139	+0.049
SmolLM3-3B	32 whole-response	GSM8K	+2.090	+0.082
SmolLM3-3B	32 whole-response	MBPP	+2.436	+0.244
OLMo-2-1B	None	GSM8K	-7.247	+0.576
OLMo-2-1B	None	MBPP	-7.559	+0.533
OLMo-2-1B	32 local	GSM8K	+1.253	+0.031
OLMo-2-1B	32 local	MBPP	+1.109	+0.070
OLMo-2-1B	32 whole-response	GSM8K	+2.412	+0.068
OLMo-2-1B	32 whole-response	MBPP	+2.363	+0.242

Table 17: Recorded first-order and remainder differences. Each row averages 12 dependent batch pairs. The last two columns sum to $D_B - D_A$ before rounding; neither is a causal attribution.

Mean differences in 10^{-3} calls/token					
Model	Prior updates	D_A	D_B	$L_B - L_A$	$R_B - R_A$
SmolLM3-3B	None	-7.480	+0.528	+8.811	-0.803
SmolLM3-3B	32 local	+1.159	+0.037	-1.136	+0.013
SmolLM3-3B	32 whole-response	+2.263	+0.163	-2.212	+0.113
OLMo-2-1B	None	-7.403	+0.554	+7.942	+0.016
OLMo-2-1B	32 local	+1.181	+0.050	-1.142	+0.011
OLMo-2-1B	32 whole-response	+2.387	+0.155	-2.335	+0.102

independently reconstructs the raw measurements, cell means and directions. Maximum measurement discrepancy is below 4.3×10^{-14}; maximum positive enclosure excess is below 1.2×10^{-15}, within the declared 2×10^{-8} tolerance. This does not establish a useful partial-query method or a lower bound against every possible comparison algorithm.

Measured cost and the remaining computational question. Table 19 separates candidate formation from the complete comparison. Local and whole-response formation each require one full gradient and have nearly identical measured time. The pair includes both candidates, reference fields, parameter envelopes, probe evaluations and the reference probe gradient; these nested times must not be added. The summed execution times are 1,240.90 seconds for SmolLM3 and 961.00 for OLMo, including respectively 0.415 and 0.166 seconds of output-norm scans. Maximum allocated memory is 1.166 and 0.953 GiB. The head graph incurs 4,553,472 and 4,527,390 vocabulary projection rows, including preceding histories and forward recomputation in reverse operations. Historical target caching and warm training are inherited costs, not included or zero. No serving latency is measured.

M.3 Varying the length of the preceding history

The comparison above uses a single nonempty history length. A preregistered extension, fixed before any of its outputs existed, repeats it after 1, 2, 4, 8, 16, 32, 64 and 128 prior updates. Nothing else changes: the same models, warm seeds, reserved batches A and B, four batch pairs and learning rate, with local and whole-response histories drawn with replacement from the remaining 640 training paths by the same seeded batch generator. Each length gives 48 comparisons after nonempty histories. The 24 empty-history comparisons do not depend on the length.

Table 18: Comparisons resolved by the fixed prefix acquisition schedules. The nominal budget counts eight-anchor blocks with upward rounding, not an equal fraction of all training work. $C < L$ favors whole-response training. All candidate-formation costs have already been paid.

Bound	Blocks	$A:\ C < L$	$A:\ L < C$	$B:\ C < L$	$B:\ L < C$
Unconditional	25%	0/72	0/72	0/72	0/72
Unconditional	50%	0/72	0/72	0/72	0/72
Unconditional	100%	24/72	48/72	6/72	66/72
Parameter-based	25%	0/72	0/72	0/72	0/72
Parameter-based	50%	0/72	0/72	0/72	0/72
Parameter-based	100%	24/72	48/72	6/72	66/72

Table 19: Mean seconds per candidate and complete comparison on Tesla V100 GPUs (one 32 GB device and five 16 GB devices). Each row contains 24 candidate steps and 12 complete comparisons. Clipped counts concern candidates only; all 384 preceding steps are also unclipped. Total execution time and peak memory are not measured by history.

Model	Prior updates	Local (s)	Whole (s)	Pair (s)	Clipped
SmolLM3-3B	None	2.333	2.318	21.629	0/24
SmolLM3-3B	32 local	2.289	2.285	21.448	0/24
SmolLM3-3B	32 whole-response	2.289	2.287	21.429	0/24
OLMo-2-1B	None	1.808	1.787	16.627	0/24
OLMo-2-1B	32 local	1.775	1.772	16.453	0/24
OLMo-2-1B	32 whole-response	1.775	1.773	16.460	0/24

Two checks passed before the primary count was read. Rerun through the extension code, the 32-update arm reproduces all 72 endpoints of the preceding subsection with maximum difference zero, and at every length the 24 recomputed empty-history comparisons match exactly.

The preregistered primary quantity counts nonempty-history comparisons in which the whole-response candidate has lower work on A. It is 29 of 48 after one prior update, 22 after two, 5 after four and zero from eight onward (Table 20). No count rises between adjacent lengths, which meets the preregistered monotone criterion. The mean difference on A changes sign between one and two prior updates, and the local advantage keeps growing up to 128 updates. Split by history type, which the preregistration did not specify, the whole-response candidate wins on A in 24, 22 and 5 of 24 comparisons after one, two and four whole-response updates, but in only 5, 0 and 0 of 24 after the same numbers of local updates. No endpoint at any length lies within the 10^{-8} sign tolerance, so strict and tolerant counts agree.

On B, which never forms a candidate, the whole-response candidate wins at most 14 of 48 comparisons at any length and 1 of 24 at the empty state, and every mean favors local training. The mean single-step advantage of the whole-response objective in these native states therefore appears only after short histories and on the batch that formed the update. This agrees with the all-anchor comparison of Section N.1, where whole-response training passes none of eight improvement tests after 128 updates.

The history length changes the parameters and the complete Adam state together. Appendix M.4 separates them.

Aggregate local-loss reversals. Theorem 1 requires coordinatewise nondecreasing match probabilities, with some strict improvements. After the extension finished, a single exploratory scan, fixed in advance but not preregistered, counted candidate-batch cases in which a candidate lowers the local loss yet raises calls per token relative to its own anchor. Of 1,632 distinct cases across all lengths, 9 qualify, all after 128 prior updates and all on B. The largest increase is 8.1×10^{-5} calls per token, against a baseline near 0.48. None of the 288 cases of the preceding subsection qualifies. These are aggregate-loss reversals, not verified instances of coordinatewise dominance. The scan

Table 20: Whole-response minus local work after preceding histories of different lengths. $C < L$ counts comparisons in which the whole-response candidate has lower work; means are in 10^{-3} calls per token. The first row is the empty fine-training Adam state shared by every length; each other row has 24 comparisons after local and 24 after whole-response histories. The 32-update row reproduces the preceding subsection. These are dependent development comparisons without a population interval.

Prior updates	$C < L$ on A	D_A (10^{-3})	$C < L$ on B	D_B (10^{-3})
None	24/24	-7.442	1/24	+0.541
1	29/48	-1.067	2/48	+0.411
2	22/48	+0.364	6/48	+0.264
4	5/48	+1.172	7/48	+0.187
8	0/48	+1.411	5/48	+0.104
16	0/48	+1.557	2/48	+0.106
32	0/48	+1.748	5/48	+0.101
64	0/48	+2.043	6/48	+0.103
128	0/48	+2.282	14/48	+0.076

did not test whether every match probability was nondecreasing and therefore does not estimate how often the theorem's stronger condition occurs.

M.4 Swapping optimizer states between histories

A second preregistered experiment separates parameters from optimizer state. At one, two and 32 prior updates, each run first recomputes the three states of the preceding subsections, which reproduce every earlier endpoint exactly. It then assembles four anchors from the two history snapshots without further training: local-history parameters with the whole-response history's complete Adam state, the reverse, and each history's parameters with an empty Adam state. Both histories share the step count, so a swap changes moment values, not the bias-correction clock. Every anchor receives the same four batch pairs (Tables 21 and 22).

After one or two prior updates, the win counts on A follow the state supplying the moments. After one prior update, the whole-response candidate wins 24 of 24 comparisons when local parameters carry whole-response moments, as on its own history, and 5 of 24 when whole-response parameters carry local moments, as after local training. After two updates the counts are 22 and 0, again equal to the states that supplied the moments. This meets the preregistered criterion at both lengths. With a reset Adam state the whole-response candidate wins all 24 comparisons on either history's parameters at every length, including after 32 updates, and the mean difference is close to the warm-state mean. At 32 updates, where every nonempty state favors local training, the larger local advantage after whole-response histories also moves with the moments: $+2.305$ against $+1.237 \times 10^{-3}$ calls per token after swapping. On B, the whole-response candidate wins at most 9 of 24 comparisons in every assembled state except one, local parameters with empty moments after 32 updates, where it wins 15 of 24 with mean -0.080×10^{-3}.

Resetting clears both moments and sets the step count to zero. Standard bias-corrected Adam then takes coordinate step $-\alpha g_i/(|g_i| + \epsilon)$, for the gradient supplied to the optimizer; it is not exact sign normalization when $\epsilon > 0$. The swaps establish that optimizer state affects the update-batch comparison at fixed parameters. They do not show that parameters are irrelevant or isolate the first moment, second moment, gradient scale or reset clock. The assembled anchors are diagnostic states, not a training method.

The strong all-anchor local implementation already obtains the complete current probability field and computes both local and continuation credit arrays. Acquiring current continuation credit is therefore not its incremental bottleneck. A useful new method must turn that information into an update that transfers beyond its construction batch, while charging any additional reverse pass or post-update evaluation. More accurate imitation of the complete gradient alone does not meet this requirement.

Table 21: Whole-response minus local work on the update batch A from assembled anchors, in 10^{-3} calls per token; $C < L$ counts comparisons in which the whole-response candidate has lower work. Each cell has 24 dependent comparisons. The first three rows reproduce earlier states exactly.

	1 prior update		2 prior updates		32 prior updates	
Anchor state	$C < L$	D_A	$C < L$	D_A	$C < L$	D_A
Warm state, empty moments	24/24	-7.442	24/24	-7.442	24/24	-7.442
Local history	5/24	+0.638	0/24	+1.852	0/24	+1.170
Whole-response history	24/24	-2.772	22/24	-1.124	0/24	+2.325
Local parameters, whole-response moments	24/24	-2.765	22/24	-1.106	0/24	+2.305
Whole-response parameters, local moments	5/24	+0.591	0/24	+1.799	0/24	+1.237
Local parameters, empty moments	24/24	-7.424	24/24	-7.375	24/24	-7.141
Whole-response parameters, empty moments	24/24	-7.454	24/24	-7.461	24/24	-7.592

Table 22: The same anchors scored on the disjoint probe batch B.

	1 prior update		2 prior updates		32 prior updates	
Anchor state	$C < L$	D_B	$C < L$	D_B	$C < L$	D_B
Warm state, empty moments	1/24	+0.541	1/24	+0.541	1/24	+0.541
Local history	0/24	+0.485	1/24	+0.302	5/24	+0.044
Whole-response history	2/24	+0.337	5/24	+0.226	0/24	+0.159
Local parameters, whole-response moments	3/24	+0.276	9/24	+0.102	5/24	+0.072
Whole-response parameters, local moments	0/24	+0.496	1/24	+0.337	0/24	+0.082
Local parameters, empty moments	2/24	+0.451	5/24	+0.352	15/24	-0.080
Whole-response parameters, empty moments	1/24	+0.515	1/24	+0.496	4/24	+0.190

M.5 Separating moments, clock and current-gradient scale

The preceding swap changes both Adam moments together. A separately preregistered intervention uses its 36 saved snapshots at one, two and 32 updates, with no history retraining. At each fixed parameter vector, let L and W name the local- and whole-history source of a moment. We assemble (m_L, v_L), (m_L, v_W), (m_W, v_L) and (m_W, v_W) at the existing clock H; additionally we clear only the first moment, only the second moment, or both moments. The two single-moment clearing operations each retain both possible sources of the other moment. Clearing both is evaluated at clock H and, separately, at clock zero. This gives ten states.

Each state receives three disposable candidates: the local gradient, the whole-response gradient, and the whole-response gradient multiplied by $\|g_L\|_2/\|g_W\|_2$. The multiplier is computed only on A, before the common norm-five clip; it never uses B and never rescales stored moments. If $g_W = 0$, the multiplier is one; we record whether both gradients vanish or norm matching is impossible. If only $g_L = 0$, it is zero. The same two complete gradients are reused for all ten states. All parameter tensors receive gradients, including exact zero gradients, and their clocks advance together. Adam keeps learning rate 10^{-4}, moment decay rates $(\beta_1, \beta_2) = (0.9, 0.999)$, $\epsilon = 10^{-8}$ and zero weight decay. No candidate is selected or propagated.

There are 144 fixed-parameter/batch blocks and 4,320 candidate steps. Of these, 864 reproduce historical candidates, all with zero observed work discrepancy; 3,456 are new component or scale interventions. The replay threshold was 10^{-10} calls/token. Every original model, seed, parameter origin, history, corpus and batch pair is retained. The 4,320 candidates are not independent experiments, and the reused warm-development paths are not a confirmation set.

Registered continuous effects. Let D_{ab} be whole minus local expected work per token when the first moment comes from a and the second from b, at fixed parameters and clock. The primary effects are

$$E_m = \tfrac{1}{2}[(D_{WL} - D_{LL}) + (D_{WW} - D_{LW})],$$
$$E_v = \tfrac{1}{2}[(D_{LW} - D_{LL}) + (D_{WW} - D_{WL})],$$
$$I_{mv} = D_{WW} - D_{WL} - D_{LW} + D_{LL}.$$

We report each on both A and B, using the original and norm-matched whole gradient separately (Figure 3). Four paired batches are averaged within each seed and origin/corpus group; the three fixed seed means and their range are descriptive, not confidence intervals. Clearing and clock contrasts, all origin/corpus groups and absolute candidate changes are retained in the accompanying machine-readable tables.

Real but restricted transfer. Figure 4 shows the same OLMo effects on both batches at one, two and 32 prior updates, each panel on its own scale. Figure 5 retains every combination of the two uncleared moments. In OLMo at one prior update, the original-gradient B effects are $E_v = -0.0011024$, $E_m = +0.0009706$ and $I_{mv} = +0.0013444$. The large interaction and opposing first-moment effect explain why mixing the components behaves differently from exchanging the complete state. After two updates the corresponding values are -0.0010515, $+0.0009110$ and $+0.0013471$. Neither one moment nor its direction of effect explains both models at every history length. At 32 updates, for example, the original gradient's A second-moment effect is positive in both models, whereas its norm-matched counterpart is negative.

The favorable mixed-state B comparisons in this appendix reflect actual improvement, not two damaged candidates. For one disposable candidate step, OLMo's whole and local changes from the common parameter baseline are -0.0052066 and -0.0038334 at the one-update anchor, then -0.0055419 and -0.0040998 at the two-update anchor. SmolLM3's norm-matched changes at its two-update anchor are -0.0040956 and -0.0038201. In each of these three cases, both candidates improve in all twelve origin/corpus/seed means, and the whole candidate improves more. There are explicit counterexamples across models, histories and norm controls. Across all 120 model/history/state/gradient B groups, 21 pooled means are negative, eleven have three negative seed means, and only these three have all twelve negative origin/corpus/seed means, using the 10^{-8} numerical sign tolerance. These are dependent descriptive counts, not success probabilities or a registered practical-value test.

Relation to multi-objective optimization. MTAdam maintains loss-specific first and second moments and balances layerwise gradient magnitudes (Malkiel & Wolf, 2021). Bloop uses a moving average to estimate the primary-gradient direction before projecting an auxiliary gradient (Hsieh et al., 2024). MAdam instead studies mismatches between reconciled objective directions and Adam's adaptive metric (Liu et al., 2026a). Our intervention assembles stored moments at fixed parameters, preserves clock controls, and measures absolute and relative response work on a disjoint batch. It is a controlled response-work diagnosis, not a new multi-objective optimizer or a transfer of those methods' guarantees to arbitrary mixed Adam histories.

Clock and scale controls. For clipped current gradient c, the standard Adam update is

$$\delta_i = -\alpha \frac{(0.9m_i + 0.1c_i)/(1 - 0.9^{H+1})}{\sqrt{(0.999v_i + 0.001c_i^2)/(1 - 0.999^{H+1})} + \epsilon}.$$

Thus scaling only c at fixed m, v is not an invariance. Matching gradient norms also need not match step norms. With both moments zero, changing the clock still changes bias correction. Resetting this clock strengthens the whole-relative improvement on A at early histories while worsening the comparison on B. Clearing only v can create unusually large steps and does not describe a plausible ordinary Adam history. These controls constrain mechanistic explanations; none defines a new optimizer or establishes a generalization guarantee. Parameters differ across history lengths, so those cross-length comparisons are not a pure clock intervention.

Measured cost. The intervention forms 288 complete gradients on A, none on B, and evaluates 4,608 baseline and 138,240 candidate prompts. The 18 runs sum to 13,947.62 seconds on Tesla V100 GPUs. They charge 71,099,748 physical full-vocabulary forward projection rows, including reverse recomputation, and 557,568 backward head rows. Original caching and warm/history fitting are inherited acquisition costs, not zero costs or part of this sum. Logged tensor payload transfers are not a complete inter-device traffic profile. A usable training procedure would additionally need a way to obtain and maintain its state, an independent evaluation and a complete training-cost comparison.

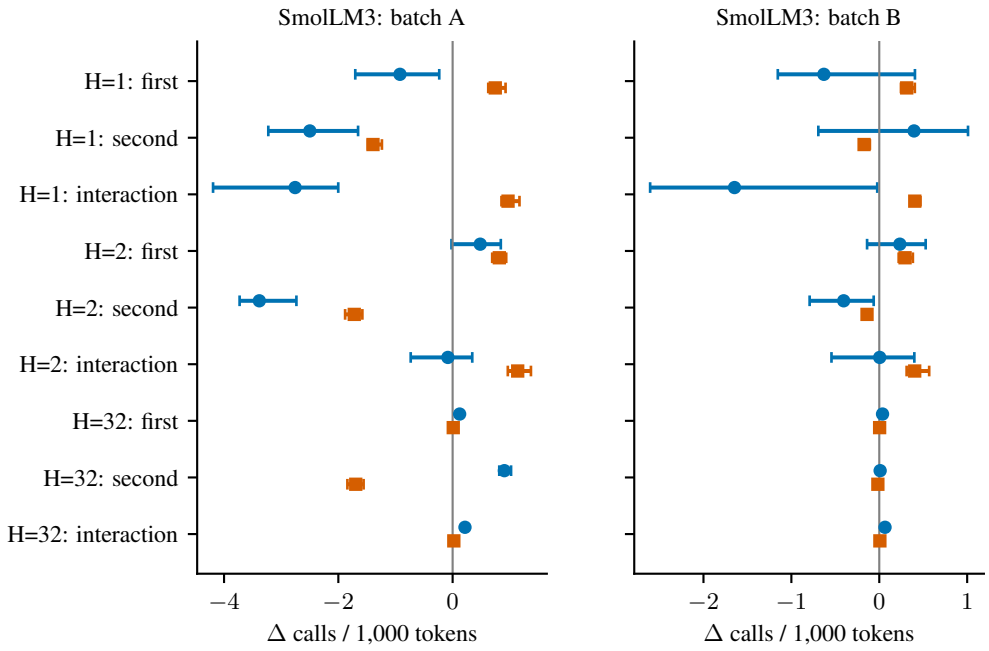


Figure 3: SmolLM3 moment effects on both batches, averaging four pairs within each seed and then the two parameter origins and corpora; bars span three fixed-seed means, not confidence intervals. Blue uses the original gradient and orange matches its batch-A norm; positive increases whole-minus-local work.

M.6 WHY PERSISTENT MIXED MOMENTS ARE NOT A GENERAL OPTIMIZER

The favorable disposable state combinations do not by themselves specify a valid persistent optimizer. Consider storing the local objective's first moment and the whole objective's second moment, while inserting the current whole gradient into a temporary first-moment numerator. This differs from ordinary Adam, whose two moments consume the same gradient sequence.

Take $W(x) = x^2/2$ and $L(x) = (x-a)^2/2$, where $0 < |a| < 5$. For $0 < \beta < 1$, $0 \le \beta_2 < 1$, $\eta > 0$ and $\epsilon > 0$, define

$$\begin{aligned} g_t^W &= x_t, \qquad g_t^L = x_t - a, \\ u_t &= \beta m_t^L + (1-\beta) g_t^W, \qquad v_{t+1}^W = \beta_2 v_t^W + (1-\beta_2)(g_t^W)^2, \\ x_{t+1} &= x_t - \eta \, \frac{u_t/(1-\beta^{t+1})}{\sqrt{v_{t+1}^W/(1-\beta_2^{t+1})} + \epsilon}, \\ m_{t+1}^L &= \beta m_t^L + (1-\beta) g_t^L. \end{aligned} \tag{68}$$

The last line is the saved first moment. Saving u_t instead defines a different algorithm.

Intuition. A local moment can move parameters away from a stationary whole objective even when the current whole gradient is zero.

Proposition 15 (Two-step stationary-point obstruction)**.** *Starting at $x_0 = m_0^L = v_0^W = 0$, the recurrence satisfies*

$$x_1 = 0, \qquad x_2 = \frac{\eta\beta a}{(1+\beta)\epsilon}, \qquad W(x_2) - W(0) = \frac{\eta^2\beta^2 a^2}{2(1+\beta)^2\epsilon^2} > 0.$$

Input-gradient clipping at norm five does not change either step. Ordinary whole-objective Adam with the same zero-state initialization stays at zero.

Proof. The first whole gradient and both initial moments are zero, so the first displacement vanishes while $m_1^L = -(1-\beta)a$. At the second step $v_2^W = 0$ and the corrected temporary numerator is $-\beta a/(1+\beta)$. Substitution gives the stated displacement and increase. At both steps the input gradients are zero and $-a$, so the declared clipping threshold is inactive. The whole-only comparator has zero gradient and moments at every step.

The obstruction vanishes at this initialization if $a = 0$ or $\beta = 0$. More generally, if both objectives have identical gradients along a whole trajectory and start with identical states, the temporary numerator equals the ordinary persistent first-moment update at every step. Induction then recovers ordinary Adam, including its second moment and clock. Arbitrary retained momentum need not preserve stationarity even in ordinary Adam; the comparison above deliberately uses the same zero-state start.

If a constant-step, unclipped trajectory and its moments converge, their limits satisfy $m^L \to x - a$, $v^W \to x^2$ and $u \to x - \beta a$. Vanishing displacement therefore requires $x = \beta a$. This is a necessary condition under the stated assumptions, not a convergence theorem. Ordinary Adam on $(1-\beta)W + \beta L$ preserves that mixture's stationary point from zero moments, but follows a different trajectory in general.

The quadratic pair is not an embedding into the MTP objective class. It refutes a general stationary-point guarantee for this precise persistence rule, not every finite-step MTP intervention. It leaves the recorded conditional gains and their counterexamples intact. MTAdam, Bloop and MAdam use different moment combinations or direction transformations (Malkiel & Wolf, 2021; Hsieh et al., 2024; Liu et al., 2026a); none supplies a guarantee for Equation (68). The calculation verifies all 144 declared parameter settings in exact arithmetic and compares them with an actual double-precision Adam assembly under the fixed numerical tolerances. Those are implementation checks, not fitted comparisons or empirical validation of an optimizer.

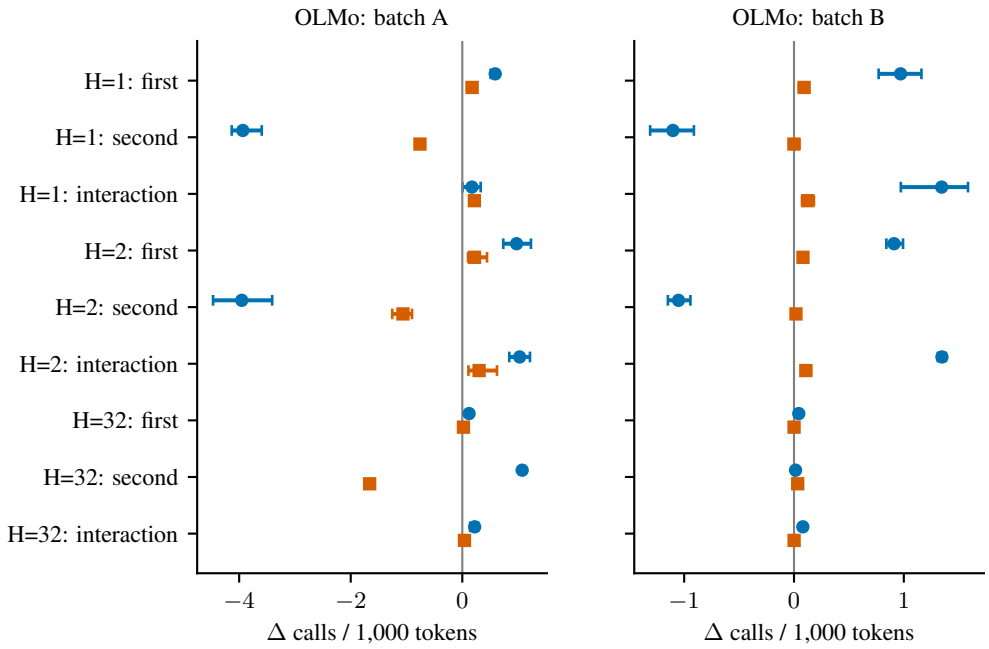


Figure 4: OLMo moment effects on both batches, at one, two and 32 prior updates. Symbols, aggregation and descriptive seed ranges follow Figure 3; each panel has its own labeled scale.

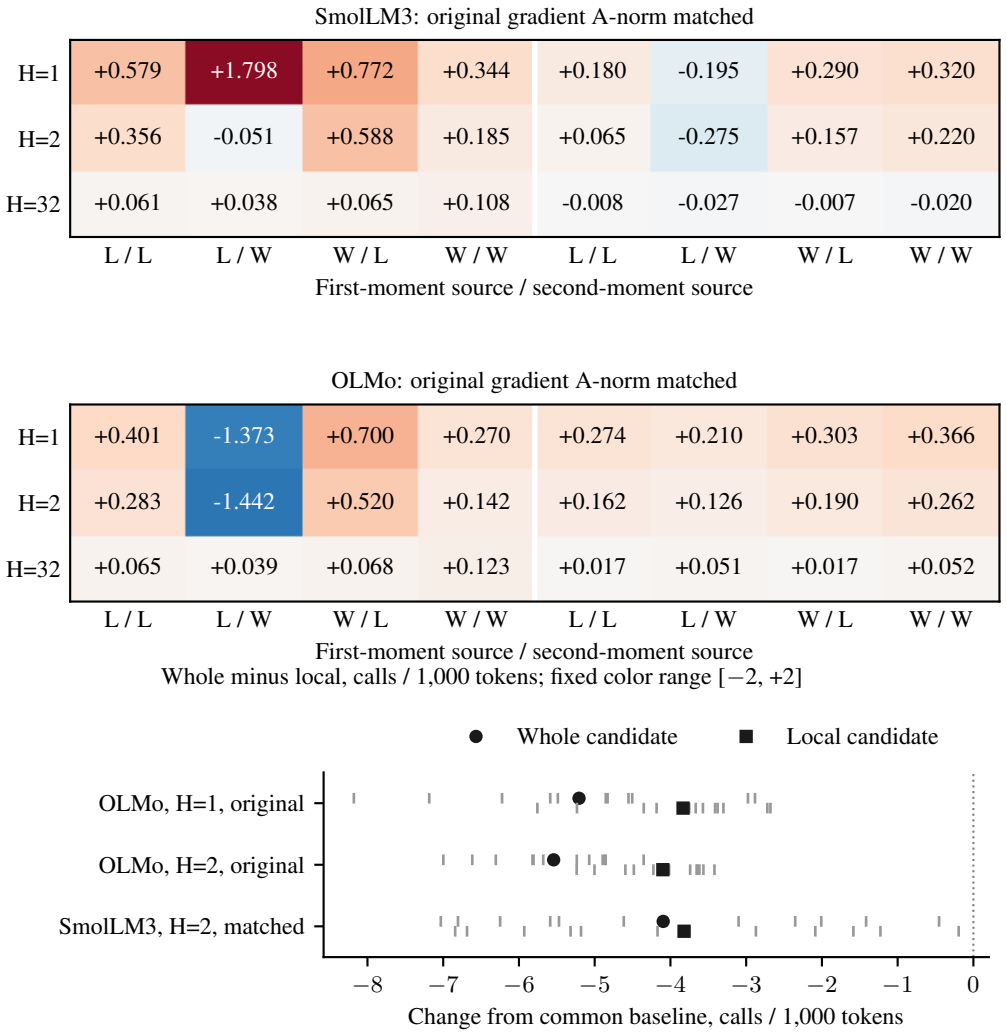


Figure 5: All 48 retained-moment batch-B means (blue favors whole-response) and absolute changes from the common baseline in the three identified conditional regions; each heatmap's left four columns use the original gradient and its right four use update-batch norm matching. L/W names each moment's source, H is the clock, and short marks show twelve origin/corpus/seed means rather than intervals; these regions are development findings, not confirmation sets.

N EARLIER TRAINING CONTROLS AND FIELD DIAGNOSTICS

N.1 SAME ANCHORS, LEARNING RATES AND COMPUTATION

The earlier development comparison uses SmolLM3-3B and OLMo-2-1B on GSM8K and MBPP (Cobbe et al., 2021; Austin et al., 2021), three warm seeds, rank-108 heads, every active anchor and 128 clipped Adam updates of eight paths. Both rates, 3×10^{-5} and 10^{-4}, are retained: 24 fits. There are 128 previously opened prompts per corpus. Seeds are averaged within prompts before 20,000 paired bootstrap draws and Bonferroni adjustment over eight contrasts, conditional on the three fitted seeds.

Seven intervals favor local training, with 0.23 to 0.68 percent higher whole-response work and the same direction in all seeds; the eighth includes zero (Figure 6, Appendix P). Physical rows are equal, update times match to 0.4 percent and neither arm clips, although local update norms are larger. Sampled replay favors local training in six contrasts. Under argmax, seven contrasts include zero and one favors whole-response training. Full levels, norms, target-generation settings and both decoding replays are in Appendix P.

All four adjusted intervals favor local training on each of Yi-1.5-6B, Granite-4.1-3B, Qwen3.8-27B and MiMo-7B-Base. Granite-4.2-8B favors local training on GSM8K; Granite-4.2-30B does so at the lower rate on both corpora. Their other intervals, and all four on Granite-4.2-3B and Qwen3.5-0.8B, include zero. Each backbone retains its own development comparison family, not pooled confirmation. Complete cohorts and all weaker controls appear in Appendix Q; no partial fits are mixed.

N.2 STRENGTHENING THE LOCAL CONTROL CHANGES THE RESULT

Whole-response training's earlier advantage depends on the comparison. Against a single sampled local anchor, it reduces work by 3.11 to 4.97 percent in six model-corpus cells across Qwen2.5-1.5B, OLMo and SmolLM3; only the two SmolLM3 cells were prospective. It uses 8.10 to 8.12 times the physical vocabulary rows and 8.23 to 8.25 times the update time. Against eight-anchor blocks on a separately registered cohort, the gain is 1.46 to 2.61 percent in four cells. These changes in anchor coverage and gradient truncation do not isolate one cause of the gain.

The economical correction instead worsens work by 3.60 to 5.26 percent in all four cells, despite meeting its physical-cost limit. Every seed is unfavorable; it meets neither superiority nor half-gap recovery in any cell. The four-arm comparison retains 48 fits, both tuned rates and 6,144 endpoints, using 20,000 prompt-bootstrap draws and one-sided 98.75% bounds within each four-cell family. Appendix O reports all objectives, tuning rules, bootstrap schedules, norms and costs.

N.3 FIXED PARAMETERS DO NOT FIX THE UPDATE RANKING

Exchanging Adam state changes which objective wins on the update batch. From identical parameters and optimizer state, we form local and whole-response candidates using eight paths, batch A, and evaluate 24 disjoint paths, batch B; neither candidate is selected or propagated. In the initial 72 comparisons, whole-response wins all 24 empty-history A comparisons and none of 48 after 32 updates, whereas local wins 66/72 on B; all prompts are development data.

Complete-state swaps at fixed parameters give A win counts of 24/24 versus 5/24 after one prior update, and 22/24 versus 0/24 after two. They exclude a parameter-only explanation, not separate effects of first moments, second moments, clocks or gradient scale. Resetting changes both moments and clock; the native 32-step count gap is zero.

A subsequent registered component intervention *does* separate these quantities. It retains 864 exact historical replays and 3,456 new candidates. Only three of 120 dependent descriptive groups improve disjoint-batch work in all twelve parameter-origin, corpus and seed means. Two OLMo gains, 1.373 and 1.442 calls per 1,000 tokens, reverse under current-gradient norm matching; SmolLM3 gains 0.275 with norm matching in another state. These assembled states do not establish a general optimization advantage. Both partial-query bounds decide 0/72 pairs at quarter and half budgets, after already paying for two complete candidate gradients (Appendix M).

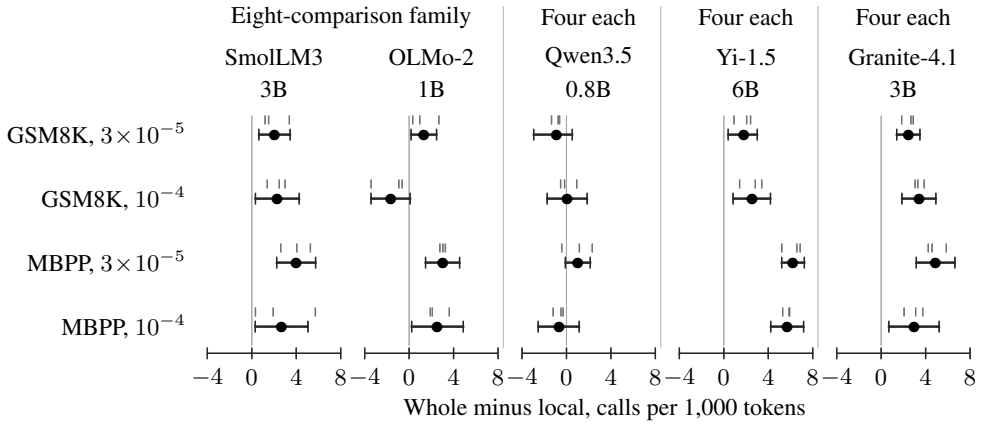


Figure 6: Whole-response minus all-anchor local work, in calls per 1,000 tokens; positive favors local training. Points and bars are prompt means and adjusted bootstrap intervals conditional on three fitted seeds, with the original eight and each additional four treated as separate families.

O COMPLETE WHOLE-RESPONSE TRAINING RESULTS

This section retains every objective in the completed three-model comparison. The four discovery-model contrasts and the two preregistered third-model contrasts have different inferential status. Repeated OLMo endpoints are reported once in the three-model summary; exact reproduction does not add an independent family, prompt or training-noise sample.

Single-anchor estimator outcomes and complete cost. Neither single-anchor estimator improves on local-prefix training in the registered two-model evaluation. The bounded estimator passes zero of four local-superiority tests, four of eight noninferiority tests, and zero of two compute tests. All four noninferiority passes concern the one-anchor exact comparator, not the complete gradient. The third-model outcomes are also unfavorable. For whole-response versus sampled local training, including diagnostics and tuning gives elapsed-time ratios of 5.62 to 5.63; these differ from the update-only ratios because the complete bill includes other work.

Bootstrap reconstruction and the preserved first report. The reconstruction uses the registered seed, 20,000 paired prompt draws and the declared corpus order. The original contrast implementation consumed its random stream in a different draw order. On SmolLM3, the reconstructed upper bounds are -2.801566% on GSM8K and -4.031055% on MBPP; the first report gave approximately -2.801% and -4.017%, respectively. The MBPP difference is about 0.014 percentage points. Both schedules satisfy the original two confirmation tests, and neither changes the observed point effects or seed directions. The first reported results remain in the artifact; they are not silently replaced or presented as an identical bootstrap execution.

Model	Corpus	Change (%)	Upper (%)	Seeds
Qwen2.5-1.5B	GSM8K	-3.443	-3.062	3/3
Qwen2.5-1.5B	MBPP	-4.455	-4.086	3/3
OLMo-2-1B	GSM8K	-4.286	-3.738	3/3
OLMo-2-1B	MBPP	-4.966	-4.521	3/3
SmolLM3-3B†	GSM8K	-3.112	-2.802	3/3
SmolLM3-3B†	MBPP	-4.455	-4.031	3/3

Table 23: Exact sequence training relative to the local-prefix objective. Negative changes favor sequence training. The endpoint is expected target calls per generated token on frozen canonical paths, not serving latency. Each cell contains 128 paired prompts and three training seeds. One-sided 97.5% upper bounds use the registered paired prompt bootstrap schedule, reconstructed directly in this displayed orientation; the historical implementation used a different draw schedule, retained separately with unchanged decisions. † marks the two prospectively registered third-model comparisons; the other four are exploratory. OLMo refitting reproduces the same endpoints and adds no independent sample.

Table 24: Complete six-objective results. Changes and 95% paired prompt intervals are relative to the local-prefix objective. These intervals are descriptive except for the two registered exact-sequence comparisons. The seed column counts favorable relative changes; all three numerical seed changes are released in the result data.

Model	Corpus	Training objective	Mean J/L	Change (%)	95% interval (%)	Seeds
Qwen2.5-1.5B	GSM8K	Token cross-entropy	0.63217	+1.596	[+1.370, +1.826]	0/3
Qwen2.5-1.5B	GSM8K	Local prefix	0.62224	+0.000	[+0.000, +0.000]	0/3
Qwen2.5-1.5B	GSM8K	Smoothed prefix	0.62917	+1.114	[+0.926, +1.303]	0/3
Qwen2.5-1.5B	GSM8K	Exact sequence	0.60082	-3.443	[-3.841, -3.062]	3/3
Qwen2.5-1.5B	GSM8K	Sampled sequence	0.63272	+1.684	[+1.464, +1.902]	0/3
Qwen2.5-1.5B	GSM8K	Adaptive sequence	0.63272	+1.684	[+1.464, +1.901]	0/3
Qwen2.5-1.5B	MBPP	Token cross-entropy	0.59784	+1.088	[+0.875, +1.302]	0/3
Qwen2.5-1.5B	MBPP	Local prefix	0.59141	+0.000	[+0.000, +0.000]	0/3
Qwen2.5-1.5B	MBPP	Smoothed prefix	0.59457	+0.535	[+0.372, +0.698]	0/3
Qwen2.5-1.5B	MBPP	Exact sequence	0.56506	-4.455	[-4.834, -4.086]	3/3
Qwen2.5-1.5B	MBPP	Sampled sequence	0.59837	+1.177	[+0.972, +1.383]	0/3
Qwen2.5-1.5B	MBPP	Adaptive sequence	0.59837	+1.177	[+0.972, +1.383]	0/3
OLMo-2-1B	GSM8K	Token cross-entropy	0.59185	+0.668	[+0.434, +0.916]	0/3
OLMo-2-1B	GSM8K	Local prefix	0.58792	+0.000	[+0.000, +0.000]	0/3
OLMo-2-1B	GSM8K	Smoothed prefix	0.58909	+0.199	[+0.026, +0.376]	0/3
OLMo-2-1B	GSM8K	Exact sequence	0.56272	-4.286	[-4.875, -3.738]	3/3
OLMo-2-1B	GSM8K	Sampled sequence	0.59227	+0.740	[+0.545, +0.948]	0/3
OLMo-2-1B	GSM8K	Adaptive sequence	0.59188	+0.673	[+0.479, +0.878]	0/3
OLMo-2-1B	MBPP	Token cross-entropy	0.62635	+0.696	[+0.417, +0.996]	0/3
OLMo-2-1B	MBPP	Local prefix	0.62202	+0.000	[+0.000, +0.000]	0/3
OLMo-2-1B	MBPP	Smoothed prefix	0.62411	+0.335	[+0.111, +0.570]	0/3
OLMo-2-1B	MBPP	Exact sequence	0.59113	-4.966	[-5.418, -4.521]	3/3
OLMo-2-1B	MBPP	Sampled sequence	0.62815	+0.986	[+0.743, +1.245]	0/3
OLMo-2-1B	MBPP	Adaptive sequence	0.62800	+0.962	[+0.726, +1.212]	0/3
SmolLM3-3B	GSM8K	Token cross-entropy	0.64962	+1.393	[+1.181, +1.609]	0/3
SmolLM3-3B	GSM8K	Local prefix	0.64069	+0.000	[+0.000, +0.000]	0/3
SmolLM3-3B	GSM8K	Smoothed prefix	0.64649	+0.904	[+0.721, +1.098]	0/3
SmolLM3-3B	GSM8K	Exact sequence	0.62075	-3.112	[-3.435, -2.802]	3/3
SmolLM3-3B	GSM8K	Sampled sequence	0.65161	+1.705	[+1.510, +1.908]	0/3
SmolLM3-3B	GSM8K	Adaptive sequence	0.65182	+1.737	[+1.541, +1.942]	0/3

Continued on the next page.

Table 24: Complete six-objective results (continued).

Model	Corpus	Training objective	Mean J/L	Change (%)	95% interval (%)	Seeds
SmolLM3-3B	MBPP	Token cross-entropy	0.59299	+1.014	[+0.628, +1.410]	0/3
SmolLM3-3B	MBPP	Local prefix	0.58704	+0.000	[+0.000, +0.000]	0/3
SmolLM3-3B	MBPP	Smoothed prefix	0.58958	+0.434	[+0.094, +0.785]	0/3
SmolLM3-3B	MBPP	Exact sequence	0.56088	-4.455	[-4.902, -4.031]	3/3
SmolLM3-3B	MBPP	Sampled sequence	0.59749	+1.781	[+1.472, +2.109]	0/3
SmolLM3-3B	MBPP	Adaptive sequence	0.59748	+1.779	[+1.471, +2.106]	0/3

Model	Objective	Head rows	Row ratio	Update ratio	Cell ratio
Qwen2.5-1.5B	Token cross-entropy	366,340	1.00	1.00	1.00
Qwen2.5-1.5B	Local prefix	366,340	1.00	1.00	1.00
Qwen2.5-1.5B	Smoothed prefix	366,340	1.00	1.00	1.00
Qwen2.5-1.5B	Exact sequence	2,973,696	8.12	8.24	5.63
Qwen2.5-1.5B	Sampled sequence	1,272,231	3.47	2.70	2.08
Qwen2.5-1.5B	Adaptive sequence	1,051,723	2.87	2.19	1.76
OLMo-2-1B	Token cross-entropy	363,996	1.00	1.00	1.00
OLMo-2-1B	Local prefix	363,996	1.00	1.00	1.00
OLMo-2-1B	Smoothed prefix	363,996	1.00	1.00	1.00
OLMo-2-1B	Exact sequence	2,947,048	8.10	8.23	5.63
OLMo-2-1B	Sampled sequence	1,230,261	3.38	2.65	2.06
OLMo-2-1B	Adaptive sequence	1,045,199	2.87	2.21	1.78
SmolLM3-3B	Token cross-entropy	366,340	1.00	1.00	1.00
SmolLM3-3B	Local prefix	366,340	1.00	1.00	1.00
SmolLM3-3B	Smoothed prefix	366,340	1.00	1.00	1.00
SmolLM3-3B	Exact sequence	2,973,696	8.12	8.25	5.62
SmolLM3-3B	Sampled sequence	1,258,816	3.44	2.69	2.08
SmolLM3-3B	Adaptive sequence	1,053,092	2.87	2.22	1.77

Table 25: Training cost relative to the local-prefix objective, charging both learning rates and all three seeds. Physical head rows include forward recomputation for the vector-Jacobian product. Update ratios use update-only wall time; cell ratios also include diagnostics and tuning. Shared cache and warm-up costs are outside these per-objective ratios. Timings are recorded process times with possible node interference, not controlled serving benchmarks.

The complete six-objective table above gives every contrast against the local-prefix objective, and Table 25 gives the training cost that buys it, charging both learning rates and all three seeds.

Counting independent units. The same 1,152 selected prompts are used by each model. Their generated paths differ, and counting those model paths as independent prompts would inflate the apparent sample size. The three seeds index warm checkpoints and the paired fitting random numbers. Re-running the deterministic OLMo pipeline with those same seeds reproduces the endpoint values rather than sampling a second set of training noise. The paired prompt intervals are therefore conditional on the fitted seeds and models.

O.1 The new-prompt four-arm comparison

All four arms use the same head class, warm checkpoints, optimizer steps, training prompts and two-rate tuning budget. The source-ordered confirmation cohort excludes the earlier training, tuning and evaluation prompt identities. It remains within the two previously used corpora and model families. The fitted seeds are averaged within a prompt before paired resampling; they are not treated as additional independent prompts.

Model	Corpus	Contrast	Change	Lower	Upper	Pass
OLMo-2-1B	GSM8K	$W - L$	-0.01476	-0.01730	-0.01226	Yes
OLMo-2-1B	GSM8K	$M - L$	+0.02970	+0.02595	+0.03345	No
OLMo-2-1B	GSM8K	$M - (L + W)/2$	+0.03708	+0.03261	+0.04158	No
OLMo-2-1B	MBPP	$W - L$	-0.01046	-0.01232	-0.00855	Yes
OLMo-2-1B	MBPP	$M - L$	+0.02124	+0.01890	+0.02365	No
OLMo-2-1B	MBPP	$M - (L + W)/2$	+0.02647	+0.02421	+0.02874	No
SmolLM3-3B	GSM8K	$W - L$	-0.00917	-0.01114	-0.00723	Yes
SmolLM3-3B	GSM8K	$M - L$	+0.02467	+0.02262	+0.02672	No
SmolLM3-3B	GSM8K	$M - (L + W)/2$	+0.02926	+0.02707	+0.03141	No
SmolLM3-3B	MBPP	$W - L$	-0.00994	-0.01222	-0.00765	Yes
SmolLM3-3B	MBPP	$M - L$	+0.02266	+0.02038	+0.02494	No
SmolLM3-3B	MBPP	$M - (L + W)/2$	+0.02763	+0.02554	+0.02973	No

Table 26: All twelve primary contrasts. L is block-local training, W whole-response training, and M sparse correction. The recovery contrast is $M - (L + W)/2$, with no division by an estimated gap. Units are absolute expected calls per token. Lower and upper columns copy the 1.25th and 98.75th percentiles of the registered paired prompt bootstrap; the upper endpoint supplies the one-sided decision. Pass copies the registered decision, not a newly selected threshold. Bounds are conditional on the three fitted seeds.

Model	Corpus	Training method	Mean	Seed 0	Seed 1	Seed 2
OLMo-2-1B	GSM8K	Single-anchor local	0.57928	0.58252	0.57713	0.57819
OLMo-2-1B	GSM8K	Block-local	0.56448	0.56794	0.56327	0.56223
OLMo-2-1B	GSM8K	Whole-response	0.54972	0.55237	0.54830	0.54849
OLMo-2-1B	GSM8K	Sparse correction	0.59418	0.59564	0.59993	0.58697
OLMo-2-1B	MBPP	Single-anchor local	0.60715	0.60715	0.60878	0.60552
OLMo-2-1B	MBPP	Block-local	0.58994	0.58957	0.58998	0.59028
OLMo-2-1B	MBPP	Whole-response	0.57948	0.57744	0.57998	0.58101
OLMo-2-1B	MBPP	Sparse correction	0.61118	0.60577	0.61563	0.61214
SmolLM3-3B	GSM8K	Single-anchor local	0.64048	0.64243	0.64273	0.63627
SmolLM3-3B	GSM8K	Block-local	0.62703	0.62812	0.62772	0.62526
SmolLM3-3B	GSM8K	Whole-response	0.61786	0.61769	0.61696	0.61893
SmolLM3-3B	GSM8K	Sparse correction	0.65170	0.64896	0.66083	0.64531
SmolLM3-3B	MBPP	Single-anchor local	0.60227	0.60344	0.60455	0.59881
SmolLM3-3B	MBPP	Block-local	0.58713	0.58985	0.58735	0.58419
SmolLM3-3B	MBPP	Whole-response	0.57719	0.57905	0.57996	0.57255
SmolLM3-3B	MBPP	Sparse correction	0.60979	0.60674	0.61686	0.60578

Table 27: All sixteen method means and all 48 seed-specific means on the new prompt cohort. The endpoint is expected target calls per token, lower is better. Each seed value averages the same 128 prompts within the model/corpus cell. The mean averages the three fitted seeds. One learning rate per model and method was selected on tuning data before this evaluation.

Model	Training method	Head rows	Rows/L	Time/L	Clipped	Corrections
OLMo-2-1B	Single-anchor local	363,996	0.990	0.968	754/768	0
OLMo-2-1B	Block-local	367,600	1.000	1.000	42/768	0
OLMo-2-1B	Whole-response	2,947,048	8.017	7.914	0/768	0
OLMo-2-1B	Sparse correction	724,696	1.971	1.954	360/768	854
SmolLM3-3B	Single-anchor local	366,340	0.990	0.970	751/768	0
SmolLM3-3B	Block-local	370,200	1.000	1.000	81/768	0
SmolLM3-3B	Whole-response	2,973,696	8.033	7.955	0/768	0
SmolLM3-3B	Sparse correction	731,792	1.977	1.964	444/768	854

Table 28: Complete training cost, charging both learning rates, three seeds and 128 updates per fit. Head rows include forward recomputation; ratios use block-local training (L) on the same model. Time is update-only elapsed time, not serving latency. Clipped counts actual minibatch steps out of 768. Corrections counts full-field branches per prompt, not minibatch steps. Shared caches, warm starts, tuning and diagnostic time are outside the displayed update-time ratio and are retained separately in the complete result.

Model	Training method	Before clipping		After clipping		Clipped
		Median	Maximum	Median	Maximum	steps
OLMo-2-1B	Single-anchor local	10.163	22.558	5.000	5.000	754/768
OLMo-2-1B	Block-local	3.944	7.474	3.944	5.000	42/768
OLMo-2-1B	Whole-response	0.408	0.905	0.408	0.905	0/768
OLMo-2-1B	Sparse correction	4.880	22.132	4.880	5.000	360/768
SmolLM3-3B	Single-anchor local	11.029	25.891	5.000	5.000	751/768
SmolLM3-3B	Block-local	4.168	6.632	4.168	5.000	81/768
SmolLM3-3B	Whole-response	0.476	0.716	0.476	0.716	0/768
SmolLM3-3B	Sparse correction	5.237	12.694	5.000	5.000	444/768

Table 29: Stored full-parameter minibatch gradient norms before and after norm-5 clipping. Each row describes all $2 \times 3 \times 128 = 768$ actual batch updates from both learning rates and all three seeds, with eight prompts per update. These are medians and maxima of recorded norms, not norms of mean gradients or an estimate of clipping bias. The recorded clipping indicator supplies the count. Differences across these trajectories do not identify the cause of the fitted outcome.

Table 26 gives the twelve primary contrasts, Table 27 all sixteen method means and all 48 seed-specific means, Table 28 the training cost, and Table 29 the gradient norms before and after clipping.

The recorded clipping counts and norms are descriptive properties of the actual eight-prompt updates, pooling both learning rates and all three seeds. They do not estimate the bias of those random minibatch updates or isolate clipping from momentum, preconditioning and sampling variance. The separate eight-path diagnostic evaluates gradients at predetermined warm checkpoints; its prompt-wise post-clipping discrepancy is not a batch-level causal test. The complete correction is unbiased only before these nonlinear operations.

P Complete Same-Anchor Development Results

The comparison in Section N.1 was specified after the preceding evaluation cohort had been opened. It compares two complete training operations on development data, not an additional confirmation. Each model has three warm checkpoints; both objectives and both learning rates start from the corresponding same checkpoint and use the same random batch schedule. All 24 fits finish 128 updates. Each objective uses all active anchors, the same four rank-108 residual heads, all vocabulary biases, FP32 matrix products, FP64 log softmax, norm-5 clipping and Adam with betas $(0.9, 0.999)$, epsilon 10^{-8} and zero weight decay. An inactive batch supplies explicit zero gradients to advance every optimizer clock. No learning rate or checkpoint is selected for the primary comparison.

The primary statistic is the within-prompt three-seed mean of whole-response minus local expected calls per token, averaged over each corpus's 128 prompts. The bootstrap retains their fixed order

Model	Corpus	Rate	Difference	Simultaneous interval	Seeds
SmolLM3-3B	GSM8K	3×10^{-5}	+0.00203	[+0.00064, +0.00345]	0/3
SmolLM3-3B	GSM8K	10^{-4}	+0.00228	[+0.00033, +0.00427]	0/3
SmolLM3-3B	MBPP	3×10^{-5}	+0.00398	[+0.00224, +0.00574]	0/3
SmolLM3-3B	MBPP	10^{-4}	+0.00266	[+0.00031, +0.00505]	0/3
OLMo-2-1B	GSM8K	3×10^{-5}	+0.00132	[+0.00018, +0.00247]	0/3
OLMo-2-1B	GSM8K	10^{-4}	-0.00166	[-0.00343, +0.00010]	3/3
OLMo-2-1B	MBPP	3×10^{-5}	+0.00302	[+0.00148, +0.00454]	0/3
OLMo-2-1B	MBPP	10^{-4}	+0.00254	[+0.00022, +0.00488]	0/3

Table 30: Same-anchor, same-rate development comparison. Differences are whole-response minus all-anchor local training in expected verification calls per token; negative values favor whole-response training. Intervals are two-sided nominal 95% Bonferroni-adjusted paired prompt-bootstrap intervals over all eight contrasts, conditional on three fixed training seeds. Both learning rates are retained. The last column counts favorable seed differences. Each corpus has 128 already opened prompts. These are complete training-operation comparisons, not isolated effects of clipping or serving-speed measurements.

Model	Rate	Loss	Train rows (10^6)	Train (s)	Probe rows (10^6)	Total (s)	Clipped
SmolLM3-3B	3×10^{-5}	Local	1.487	859.1	0.070	957.2	0/384
SmolLM3-3B	3×10^{-5}	Whole	1.487	855.7	0.070	952.6	0/384
SmolLM3-3B	10^{-4}	Local	1.487	855.5	0.070	950.5	0/384
SmolLM3-3B	10^{-4}	Whole	1.487	854.3	0.070	949.5	0/384
OLMo-2-1B	3×10^{-5}	Local	1.474	665.9	0.068	741.0	0/384
OLMo-2-1B	3×10^{-5}	Whole	1.474	664.7	0.068	743.2	0/384
OLMo-2-1B	10^{-4}	Local	1.474	665.7	0.068	739.5	0/384
OLMo-2-1B	10^{-4}	Whole	1.474	665.5	0.068	739.2	0/384

Table 31: Measured fitting costs, summed over three seeds at each retained rate. Local means all-anchor local and Whole means whole-response. A physical head row is a full-vocabulary head calculation; the training column includes vector-Jacobian recomputation. Extra probes are separately charged. Total seconds include within-fit probes, diagnostics and terminal tuning, but exclude separately reported cache creation, warm-start fits and final evaluation. Clipping counts are descriptive, not a causal explanation.

and all paired values. One PCG64 generator with seed 2026090801 supplies 20,000 draw matrices in GSM8K, then MBPP order; the same corpus draws are used for every model and rate. The two-sided nominal 95% Bonferroni-adjusted intervals use linear quantiles 0.003125 and 0.996875 over the eight comparisons. They condition on these three fitted seeds. They do not quantify uncertainty over new training seeds, model families or unseen corpora.

All seven local-favorable intervals in Table 30 also have local-favorable differences in each of the three seeds. The only whole-favorable point estimate, OLMo on GSM8K at 10^{-4}, has three favorable seed directions but an interval crossing zero. Thus no cell satisfies the frozen whole-response improvement rule. The exception is not removed, and the higher learning rate is not substituted for the lower one.

The local-training level of each cell, needed to convert a difference into a percentage, runs from 0.551 to 0.626 expected calls per token: SmolLM3-3B gives 0.626399 and 0.615576 on GSM8K and 0.586050 and 0.574527 on MBPP at the two rates, and OLMo-2-1B gives 0.565317 and 0.551380 on GSM8K and 0.591740 and 0.576940 on MBPP. Dividing each difference by its own level reproduces the relative changes 0.324, 0.371, 0.679, 0.463, 0.233, -0.300, 0.511 and 0.440 percent in table order. Figure 6 displays these same eight intervals on a common scale, alongside the three fitted-seed means. Rescaling calls per token to calls per 1,000 tokens changes no comparison or interval.

Table 31 gives the measured fitting costs summed over three seeds at each retained rate. The physical forward and recomputation counts are exactly equal within every model-rate pair. Timing is

Backbone	Corpus	Rate	Sampled realized	Argmax realized
OLMo-2-1B	GSM8K	3×10^{-5}	$+0.00124$ $[-0.00001, +0.00252]$	$+0.00065$ $[-0.00200, +0.00338]$
OLMo-2-1B	GSM8K	10^{-4}	-0.00169 $[-0.00362, +0.00023]$	-0.00408 $[-0.00757, -0.00070]$
OLMo-2-1B	MBPP	3×10^{-5}	$+0.00315$ $[+0.00155, +0.00473]$	$+0.00073$ $[-0.00273, +0.00413]$
OLMo-2-1B	MBPP	10^{-4}	$+0.00297$ $[+0.00058, +0.00534]$	$+0.00019$ $[-0.00393, +0.00449]$
SmolLM3-3B	GSM8K	3×10^{-5}	$+0.00186$ $[+0.00036, +0.00338]$	$+0.00077$ $[-0.00220, +0.00366]$
SmolLM3-3B	GSM8K	10^{-4}	$+0.00240$ $[+0.00037, +0.00447]$	$+0.00057$ $[-0.00285, +0.00391]$
SmolLM3-3B	MBPP	3×10^{-5}	$+0.00403$ $[+0.00222, +0.00587]$	$+0.00106$ $[-0.00248, +0.00472]$
SmolLM3-3B	MBPP	10^{-4}	$+0.00268$ $[+0.00023, +0.00513]$	-0.00146 $[-0.00533, +0.00256]$

Table 32: Whole-response minus all-anchor local calls per token on the 256 evaluation paths per backbone, as realized counts under the two preregistered decodings; analytic effects are also reported in Table 30. Every column uses the same procedure: seeds averaged within prompt, 20,000 paired prompt-bootstrap draws from one fixed generator, endpoints at the 0.003125 and 0.996875 quantiles over a family of eight. The sampled column averages 32 decodes per path. A positive point favours local training.

measured, not inferred from the row counts; the small differences do not establish a meaningful hardware advantage. The table sums three fits per model, rate and objective. Training on two corpora is shared, so the training bill is not duplicated by corpus. Common historical target-cache and warm-start costs, and the new endpoint evaluation costs, are reported separately in the result files. The per-fit total is not the total cost of independently reconstructing all external assets. Neither objective clips any of its 1,536 measured training updates.

Realized calls under the decoding rule, on the saved heads. Every endpoint above is an analytic expectation of verifier calls. A preregistered replay runs the decoding rule itself on the same 24 fitted heads and the same 256 cached target paths per backbone, with no training and no new prompt. Because acceptance means equality with the cached greedy target path, the committed prefix always coincides with that path and the cached per-position hidden states are exactly the states the head reads, so the replay is exact. Two numerical checks were completed before any realized count was read: the analytic recursion applied to each published row's saved target log-probabilities reproduces its published call count within 10^{-10} on all $6{,}144$ rows, and the head logits recomputed with the fitting code's eight-anchor chunking differ from the saved log-probabilities by at most 7.0×10^{-5}. Two realizations were preregistered. In the sampled realization each proposal is an independent draw from its offset's softmax, 32 decodes per path; the analytic endpoint is the exact expectation of this realization, and in all 24 cells the realized mean lies within two Monte Carlo standard errors of it, the largest ratio being 1.95. In the argmax realization each proposal is the offset's most probable token, the deterministic policy the theory treats as a separate construction; it accepts more often, and its local level is lower than the analytic one in every cell, by 0.025 to 0.039 calls per token, 4.3 to 6.6 percent and 5.3 percent on average. Table 32 recomputes the eight contrasts on realized counts with the published inference procedure unchanged. Under the sampled realization six of eight favour local training and the seventh's lower endpoint is -0.00001; under the argmax realization no contrast favours local training and one, OLMo GSM8K at 10^{-4}, favours whole-response training. The seven-of-eight statement in the body is therefore a statement about the sampled-proposal rule that the objective and the theory define, and not about deterministic argmax proposals. Under those, seven of the eight contrasts include zero and the eighth, OLMo-2-1B on GSM8K at 10^{-4}, excludes it in favour of whole-response training at -0.00408 $[-0.00757, -0.00070]$ calls per token, after the same adjustment over a family of eight. The registered adjustment is Bonferroni. Under a Holm step-down adjustment of the same bootstrap p-values all eight analytic contrasts separate, and the eighth, OLMo-2-1B on GSM8K at 10^{-4}, then favours whole-response training, adjusted $p = 0.0093$; the seven-of-eight statement holds under the registered adjustment only.

Measured update scales and the limits of the learning-rate control. Matching a nominal learning rate does not by itself match how far an arm moves, so we report both quantities from the 96 recorded updates, four positions in each of the 24 fits. The norm of the exact whole-response gradient, evaluated along each arm's own path, has medians 0.4205 and 0.4165 for OLMo at 10^{-4}, 0.4304 and 0.4387 at 3×10^{-5}, 0.4716 and 0.4649 for SmolLM3 at 10^{-4}, and 0.4978 and 0.5130 at

Model	Rate	Loss	$g^T\Delta < 0$	$F_{\text{after}} < F_{\text{before}}$	Disagree	$\lvert R\rvert$
SmolLM3-3B	3×10^{-5}	Local	12/12	12/12	0	7.2e-06
SmolLM3-3B	3×10^{-5}	Whole	12/12	12/12	0	6.8e-06
SmolLM3-3B	10^{-4}	Local	12/12	12/12	0	3.8e-05
SmolLM3-3B	10^{-4}	Whole	12/12	12/12	0	4.1e-05
OLMo-2-1B	3×10^{-5}	Local	12/12	12/12	0	5.1e-06
OLMo-2-1B	3×10^{-5}	Whole	12/12	12/12	0	4.2e-06
OLMo-2-1B	10^{-4}	Local	12/12	12/12	0	3.2e-05
OLMo-2-1B	10^{-4}	Whole	12/12	12/12	0	3.5e-05

Table 33: Descriptive checks at updates 1, 32, 64 and 128 for every seed and retained rate. g is the exact whole-response gradient before the actual clipped Adam update Δ; F is the batch mean of normalized work, measured in expected calls per token, on that same training minibatch. Disagree counts whether the strict decrease decisions from the directional derivative and actual change differ; $|R|$ is the median absolute difference between actual change and $g^T\Delta$. These dependent training measurements are not independent trials, prospective predictions, causal clipping tests, or certified curvature bounds.

3×10^{-5}, local first, agreeing within three percent. Each arm's own training gradient is not matched: its pre-clip norm is 3.4 to 3.8 times larger under local training. Median actual update norms are larger for local training by 14.6, 19.8, 12.3 and 20.4 percent in the same four cells. Neither registered learning rate yields a whole-response-favorable adjusted interval for the analytic endpoint. However, this two-point grid does not match realized update magnitudes or optimizer histories, and therefore does not isolate or rule out an effective-step-size explanation. This differs from the earlier block-correction study, where the arms' recorded norms differ by an order of magnitude (Table 29); that study is not the comparison reported here.

Table 33 gives the descriptive checks at updates 1, 32, 64 and 128 for every seed and retained rate. There are 96 recorded updates: four positions in each of 24 fits. At each position, the exact whole-response gradient is measured before the update; its inner product uses the actual parameter displacement after Adam, not a fixed-preconditioner approximation. The measurement preserves the training gradient, optimizer and random state; its additional computation is separately charged. All 96 directional derivatives are negative, and all 96 same-batch work changes are negative. The linear and finite-step decrease decisions disagree zero times.

These observations concern training batches along dependent trajectories. They are not 96 independent experiments and do not predict the final evaluation ranking: both objectives improve their measured batches while the local objective is better on seven evaluation comparisons. Inactive clipping excludes active clipping as the explanation for this particular comparison, but does not identify the role of gradient scale, adaptive preconditioning or the different parameter trajectories. The older sparse correction's frequent clipping remains a separate observation.

An independent reconstruction uses only the 6,144 raw endpoint rows, without importing the original inference module. It reconstructs eight comparisons, the frozen random-draw identities and 6,272 numerical quantities, with maximum absolute discrepancy below 1.3×10^{-15}. This checks published endpoint statistics, not the fitted optimizer trajectories or complete training rebuild.

Available reconstruction material. The anonymous archive adds all 6,144 endpoints, all 96 update probes and the complete probability fields for the four separate actual-update batches to the other saved scientific results. Its ordinary-CPU entry points reconstruct the paired intervals and the saved-radius work bounds without weights, external data or a scheduler. The archive also discloses source-anonymized training implementations and explicit external-asset requirements. It does not reconstruct training, derive the saved radii from weights, remeasure hardware costs or fix the full software and driver environment.

P.1 Reassociating the fixed readout computation

Implementation cost is not a computational lower bound. In real arithmetic,

$$[H + (H\Omega)\Gamma_j]W + b_j = HW + (H\Omega)(\Gamma_j W) + b_j.$$

Corpus	Rate	Mean	Adjusted interval	Seed 0	Seed 1	Seed 2
GSM8K	3×10^{-5}	-0.907	[-2.950, +0.522]	-1.347	-0.753	-0.620
GSM8K	10^{-4}	+0.074	[-1.735, +1.857]	-0.170	-0.528	+0.922
MBPP	3×10^{-5}	+1.020	[-0.096, +2.139]	+2.321	+1.163	-0.424
MBPP	10^{-4}	-0.665	[-2.544, +1.141]	-0.305	-0.491	-1.199

Table 34: Qwen3.5-0.8B: whole-response minus all-anchor local work, in calls per 1,000 output tokens. Negative favors whole-response training. Approximate paired prompt-bootstrap intervals use 128 prompts per corpus and the preregistered four-comparison adjustment, conditional on three fitted seeds. Seed columns are individual means, not intervals.

On each model's first fixed eight-path batch, two warmups and five alternating paired measurements compare both associations from the same checkpoint and empty Adam state, with no retained candidate. On one V100-16GB per model, median local/whole update times change from 2.2181/2.2269 to 0.1211/0.1317 seconds for SmolLM3, and from 1.7306/1.7397 to 0.0955/0.1072 for OLMo. Packing, transfers, normalization, reverse computation, clipping and Adam are included, but common loading and post-step evaluation are separate.

Both SmolLM3 objectives meet the fixed numerical conditions. OLMo's first-pair relative gradient discrepancies, 5.2751×10^{-5} and 2.3581×10^{-5}, exceed the fixed 2×10^{-5} limit on $\|g_{\text{reuse}} - g_{\text{ref}}\|_2 / \max(\|g_{\text{ref}}\|_2, 1)$. Peak allocated bytes rise from 1.107 to 4.876 billion for SmolLM3 and 0.875 to 3.825 billion for OLMo. Relative Adam-increment discrepancies are approximately 0.00052/0.00043 and 0.00182/0.00200, respectively. There is no free cache across updates, and objective agreement does not establish trajectory identity. This is a mixed numerical pilot, not a validated replacement, new trainer or explanation of the all-anchor ranking.

Q Additional Backbone Experiments

We extend the response-work training comparison to Qwen3.5-0.8B (Qwen Team, 2026b). Every registered cell is accounted for. The checkpoint was fixed before fitting, but the corpora and prompt identities had already informed earlier development. This is a backbone extension, not confirmation on a new prompt population. We use text-only inputs and do not evaluate the model's vision branch or its published native MTP system. We additionally report complete response-training families on Yi-1.5-6B (01.AI, 2024), Granite-4.1-3B (IBM Granite Team, 2026a), MiMo-7B-Base (Xiaomi LLM Core Team, 2025), Qwen3.8-27B (Qwen Team, 2026a), and Granite-4.2 at 3B, 8B and 30B (IBM Granite Team, 2026c;b), together with all registered missing families.

Q.1 Response work under three local controls

All three training studies retain the frozen full-vocabulary head, three warm seeds, two learning rates, and the original objective and optimization rules. The complete matrix contains three warm fits and 72 subsequent fits: 36 for single-anchor controls, 24 for block controls, and 12 for all-anchor controls. All 10,752 evaluation endpoints are included. Target paths use native FP32 full-prefix calls, no key-value reuse, plain inputs capped at 256 tokens and responses capped at 64 tokens. The native stopping rule is retained; no prompt is removed because of its length or acceptance. These counts are not measured serving latency or answer quality.

All four adjusted intervals cross zero (Table 34). The two favorable whole-response means agree across all three seeds, but neither yields a favorable adjusted interval. This establishes neither superiority nor equivalence. The intervals average seeds within each of 128 prompts per corpus, then use 20,000 paired prompt-bootstrap draws with percentile endpoints 0.00625 and 0.99375. Their nominal family error is 5%; this is an approximate bootstrap procedure, not an exact finite-sample guarantee. This new four-comparison family is separate from the earlier eight-comparison family and from other backbones.

Against the single-anchor local control, whole-response work decreases by 0.022629 calls/token on GSM8K and 0.021366 on MBPP. Against the eight-anchor control, the decreases are 0.009317 and 0.008917. Each direction holds in all three seeds after the original tuning rule selects the rate. By

contrast, occasional full correction increases work relative to the block-local control by 0.030590 and 0.021877; its half-gap contrasts are also unfavorable. The adaptive single-anchor estimator increases work by 0.007810 and 0.005483 relative to single-anchor local training. These are descriptive secondary comparisons, not additional adjusted hypothesis tests. All 84 method, seed and corpus means, including the cross-entropy and position-weighted controls, are retained in the accompanying results.

The cache, warm fits, all fitted rates, selectors and evaluations report 16,550.6 seconds of summed process time. This is neither elapsed calendar time nor a complete resource bill: it excludes preceding model download, interface tests, process startup outside the timers and serving deployment. The result supports distinguishing sampled from fully averaged local controls; it does not assign the earlier gain to a unique causal mechanism. Separately, the resource accounting records 5.916 GPU-hours and 23.886 CPU-hours for this extension, including two registered comparison families not reported in this submission, and their explicitly listed model-specific preparation. Failed work is included; shared development is listed separately. These allocation hours are not hardware-normalized and exclude unlisted project history and deployment. A separate CPU replay package includes every saved probability field for all 10,752 endpoints. It reconstructs the 84 means and four adjusted intervals, but does not regenerate teacher paths, retrain heads or repeat rate selection.

Q.2 Completed additional response-training families

Yi, Granite-4.1, MiMo, Qwen3.8 and all three Granite-4.2 models each have all three response-training studies, with 10,752 endpoints per model. The initially preempted MiMo all-anchor and Qwen3.8 block and all-anchor families are retained as failed attempts. Separately registered replacement cohorts reuse the same completed target caches and warm starts, rerun each entire affected family, and never mix old partial fits with new ones. The counts combine earlier completed families with the registered additions: 6,144 new endpoints on Qwen3.8, 3,072 on MiMo and 10,752 on each Granite-4.2 model. The earlier Qwen3.8 single-anchor and MiMo single-anchor and block families are retained unchanged, not rerun or reanalysed. Granite-4.2-30B completes the same full matrix after one preregistered device-only repair to its numerical interface check. The earlier comparison exception and its costs remain reported; no scientific endpoint was available when that repair was specified.

Model	Single-anchor	Block-local	All-anchor
Yi-1.5-6B	4,608	3,072	3,072
Granite-4.1-3B	4,608	3,072	3,072
MiMo-7B-Base	4,608	3,072	3,072
Qwen3.8-27B	4,608	3,072	3,072
Granite-4.2-8B	4,608	3,072	3,072
Granite-4.2-3B	4,608	3,072	3,072
Granite-4.2-30B	4,608	3,072	3,072

Table 35: Complete evaluation endpoint records by registered comparison family. All seven models have all three families. Old preempted attempts remain in the resource and failure records; no partial fits enter the replacement families.

Yi, both Granite generations and MiMo use native FP32 target calls and FP32 readouts. Granite-4.1 preserves its native output operation $\mathrm{linear}(h, W)/10$ in that order; the qualified Granite-4.2 models have unit output scaling. Qwen3.8 uses the separately qualified native BF16 interface and FP32 readouts; it is not an FP32-native replication. Although the MiMo checkpoint includes native multi-token tensors, these experiments fit our registered heads, not its native multi-token system. All use the same two short-task corpora and already-opened prompt population.

All four adjusted intervals on each of Yi and Granite-4.1 favor local training, with favorable local directions in every seed. Each interval averages the three fitted seeds within each of 128 prompts per corpus before 20,000 paired bootstrap draws, with percentile endpoints 0.00625 and 0.99375. The original eight-comparison family and each new four-comparison family remain separate.

Qwen3.8 and MiMo also favor all-anchor local training in all four adjusted comparisons. Granite-4.2-8B has two positive GSM8K intervals and two MBPP intervals containing zero. Granite-4.2-30B

Model	Corpus, rate	Mean	Adjusted interval	Seed 0	Seed 1	Seed 2
Yi, 6B	GSM8K, 3×10^{-5}	+1.790	$[+0.395, +3.011]$	+2.404	+0.912	+2.055
Yi, 6B	GSM8K, 10^{-4}	+2.554	$[+0.833, +4.197]$	+3.408	+2.836	+1.417
Yi, 6B	MBPP, 3×10^{-5}	+6.218	$[+5.205, +7.249]$	+6.565	+5.218	+6.870
Yi, 6B	MBPP, 10^{-4}	+5.694	$[+4.223, +7.178]$	+5.917	+5.295	+5.870
Granite, 3B	GSM8K, 3×10^{-5}	+2.470	$[+1.397, +3.487]$	+2.680	+1.844	+2.886
Granite, 3B	GSM8K, 10^{-4}	+3.407	$[+1.869, +4.922]$	+3.874	+3.298	+3.050
Granite, 3B	MBPP, 3×10^{-5}	+4.895	$[+3.162, +6.636]$	+5.864	+4.582	+4.238
Granite, 3B	MBPP, 10^{-4}	+2.967	$[+0.689, +5.215]$	+3.737	+2.064	+3.101

Table 36: Additional all-anchor comparisons, in calls per 1,000 output tokens. Positive favors local training. Each model has its own four-comparison approximate prompt-bootstrap adjustment, conditional on the three fixed training seeds. No joint inference across models is claimed.

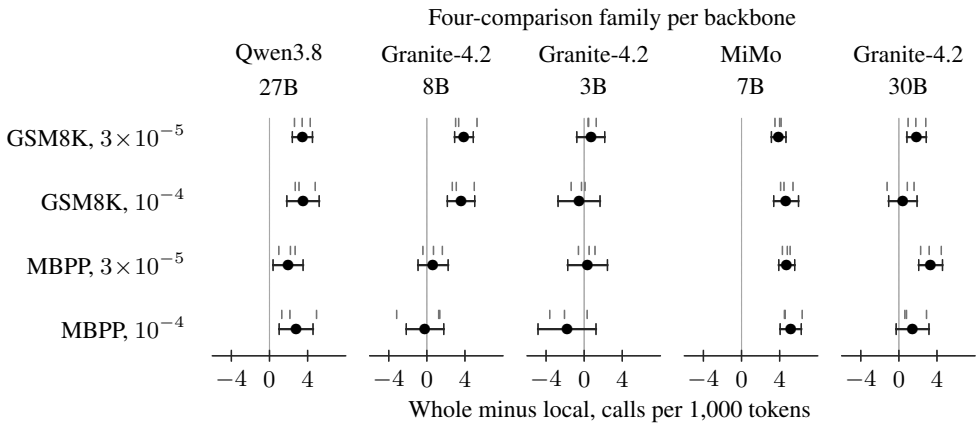


Figure 7: Complete additional all-anchor comparisons on a common scale. Positive differences favor local training. Each backbone has its own four-comparison family, with approximate prompt-bootstrap intervals conditional on three fitted seeds; short vertical marks show the seed means. All twenty comparisons, including the unresolved Granite intervals, are retained.

favors local training at the lower rate on both corpora, with both higher-rate intervals containing zero; all four Granite-4.2-3B intervals contain zero (Table 37). Thus expanding the backbone set retains both favorable local results and unresolved comparisons. It does not establish equivalence on the unresolved cells or a universal ranking of the training objectives. Each interval conditions on three fitted seeds; the new backbones do not expand the original training-seed sample.

The independent implementation reconstructs every new contrast and interval within 2.23×10^{-16} calls/token. Its forward-occupancy replay of the saved probability fields differs from the backward recursion by at most 9.95×10^{-14} calls, within the registered 10^{-10} tolerance. These are nonzero floating-point residuals. They check reconstruction of the saved experiments, not independent model retraining.

Every reported whole-response mean and seed favors whole-response training against these weaker controls. Every correction mean and seed is unfavorable. The adaptive single-anchor means are unfavorable, although one Yi/GSM8K seed is favorable. All controls, fitted rates, method/corpus/seed means, half-gap contrasts and failure records accompany the tables. These comparisons show that conclusions change with the local control; they do not separately identify sampling, clipping, gradient scale or optimizer history as its cause.

The five-model addition uses 49.046 allocated GPU-hours, including failed and cancelled attempts; Granite-4.2-30B accounts for 24.258 hours. Reused earlier Qwen3.8 and MiMo caches and warm fits are not billed again. These are allocation costs for the stated increment, not lifetime research cost or serving latency. The accompanying offline CPU package reconstructs all 41,472 new evaluation records, 33,792 tuning records, 264 fitted-cell records and twenty primary intervals. The three earlier

Backbone	Corpus	Rate	Δ	Adjusted interval
Qwen3.8-27B	GSM8K	3×10^{-5}	$+3.449$	$[+2.403, +4.512]$
Qwen3.8-27B	GSM8K	10^{-4}	$+3.539$	$[+1.820, +5.218]$
Qwen3.8-27B	MBPP	3×10^{-5}	$+1.958$	$[+0.375, +3.529]$
Qwen3.8-27B	MBPP	10^{-4}	$+2.789$	$[+1.018, +4.571]$
Granite-4.2-8B	GSM8K	3×10^{-5}	$+3.875$	$[+2.916, +4.881]$
Granite-4.2-8B	GSM8K	10^{-4}	$+3.589$	$[+2.158, +5.024]$
Granite-4.2-8B	MBPP	3×10^{-5}	$+0.639$	$[-0.923, +2.239]$
Granite-4.2-8B	MBPP	10^{-4}	-0.194	$[-2.157, +1.776]$
Granite-4.2-3B	GSM8K	3×10^{-5}	$+0.718$	$[-0.776, +2.155]$
Granite-4.2-3B	GSM8K	10^{-4}	-0.518	$[-2.741, +1.666]$
Granite-4.2-3B	MBPP	3×10^{-5}	$+0.353$	$[-1.729, +2.444]$
Granite-4.2-3B	MBPP	10^{-4}	-1.790	$[-4.845, +1.242]$
MiMo-7B-Base	GSM8K	3×10^{-5}	$+3.888$	$[+3.140, +4.665]$
MiMo-7B-Base	GSM8K	10^{-4}	$+4.655$	$[+3.379, +5.983]$
MiMo-7B-Base	MBPP	3×10^{-5}	$+4.739$	$[+3.909, +5.581]$
MiMo-7B-Base	MBPP	10^{-4}	$+5.160$	$[+4.055, +6.276]$
Granite-4.2-30B	GSM8K	3×10^{-5}	$+1.856$	$[+0.844, +2.877]$
Granite-4.2-30B	GSM8K	10^{-4}	$+0.420$	$[-1.074, +1.912]$
Granite-4.2-30B	MBPP	3×10^{-5}	$+3.311$	$[+2.081, +4.581]$
Granite-4.2-30B	MBPP	10^{-4}	$+1.443$	$[-0.281, +3.161]$

Table 37: Whole-response minus all-anchor local training, in expected verifier calls per 1,000 generated tokens. Positive values favour local training. Each backbone has its own four-comparison family: 20,000 paired prompt-bootstrap draws with percentile endpoints 0.00625 and 0.99375, conditional on three fitted seeds. Qwen3.8-27B uses a BF16 frozen trunk; the other completed backbones use FP32. The intervals are not fleet-wide or training-seed-population inference.

		Single-anchor control		Block-local control	
Model	Corpus	Whole	Adaptive	Whole	Correction
Yi-1.5-6B	GSM8K	-22.304	+5.547	-9.573	+24.789
Yi-1.5-6B	MBPP	-26.444	+10.569	-6.957	+27.821
Granite-4.1-3B	GSM8K	-18.907	+10.599	-6.283	+21.013
Granite-4.1-3B	MBPP	-22.065	+8.355	-10.485	+24.245
MiMo-7B-Base	GSM8K	-21.860	+9.732	-6.502	+30.977
MiMo-7B-Base	MBPP	-19.133	+9.172	-6.308	+24.746
Qwen3.8-27B	GSM8K	-21.065	+7.563	-7.158	+27.277
Qwen3.8-27B	MBPP	-23.565	+3.906	-8.458	+23.184
Granite-4.2-8B	GSM8K	-21.417	+7.881	-7.914	+29.472
Granite-4.2-8B	MBPP	-23.802	+7.669	-10.432	+23.515
Granite-4.2-3B	GSM8K	-26.143	+8.083	-11.738	+22.205
Granite-4.2-3B	MBPP	-35.494	+5.830	-17.722	+22.688
Granite-4.2-30B	GSM8K	-27.781	+3.864	-12.312	+26.390
Granite-4.2-30B	MBPP	-29.971	+6.151	-12.383	+26.343

Table 38: Complete weaker-control contrasts on seven further backbones, in calls per 1,000 tokens. Each entry subtracts the corresponding local control from the named training operation. These are descriptive means over the three frozen seeds after the original rate-selection rule; no new adjusted intervals are asserted. The replacement Qwen3.8 block family is complete; its earlier failed attempts remain in the resource record.

Qwen3.8 and MiMo families are supplied as unchanged historical summaries. Together these cover all three studies on five models, but the package does not regenerate target paths or retrain heads.

All seed effects, failed attempts and costs remain available; these are not native-generation measurements or validation of a new trainer. The tables of this appendix carry every record behind these statements: Table 35 the endpoint records by registered family, Tables 36 and 37 the additional all-anchor comparisons, and Table 38 the weaker-control contrasts.

R SIGNED CONTINUATION ACCOUNTING FOR FITTED HEADS

We evaluate the performance-difference identity (Equation 19) on all 3,072 paired Qwen3.5-0.8B endpoints, not only the 879 scalar reversals. For each pair, let p be the locally trained field, q the whole-response-trained field, $\Delta A_{a,k} = A_{q,a,k} - A_{p,a,k}$, $D_{a+k} = V_p(a+k) - V_p(a+k+1)$ and $w_a = d_q(a)/n$. With $x_+ = \max\{x, 0\}$ and sums over active (a, k), define

$$H_+ = \sum w_a[\Delta A_{a,k}]_+[-D_{a+k}]_+, \qquad H_- = \sum w_a[-\Delta A_{a,k}]_+[D_{a+k}]_+,$$
$$B_+ = \sum w_a[\Delta A_{a,k}]_+[D_{a+k}]_+, \qquad B_- = \sum w_a[-\Delta A_{a,k}]_+[-D_{a+k}]_+.$$

H_+ counts greater prefix survival into costlier continuations, and H_- counts lost beneficial skips; the two B terms are their beneficial counterparts. For every pair, $(J(q) - J(p))/n = H_+ + H_- - B_+ - B_-$.

Table 39: Signed accounting over all 3,072 paired fitted endpoints. Each value is a mean in verifier calls per generated token; the comparison direction is whole-response minus local training.

Term	Mean calls per generated token
H_+: costly new destinations	0.000594322
H_-: lost beneficial skips	0.011078020
B_+: beneficial new acceptance	0.018741776
B_-: beneficial avoidance	0.000768770
Whole-response minus local work	−0.007838204

The signed terms reconstruct each direct work difference. Independent recomputation from the same saved binary64 fields differs by at most 1.344×10^{-15} calls per token. These descriptive means condition on fixed head pairs and split-specific prompts from two disjoint cohorts. Single-anchor uses one prompt cohort; block and all-anchor share another, disjoint cohort. Prompts recur within, not across, the cohorts. No pair satisfies coordinatewise dominance; the four terms are neither separate causal interventions nor a training-step mechanism.