RIGHT DIRECTION, WRONG STEP: GEOMETRIC ANALYSIS OF FINITE-STEP FAILURE IN LOOPED TRANSFORMERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Looped Transformers reuse shared layers to provide a parameter-efficient route to test-time scaling through iterative latent reasoning. Additional loops can lower support for a reference answer, but a harmful endpoint does not reveal whether the update direction is unhelpful or a useful direction was applied too far. We study this distinction by measuring reference utility, a continuous measure of support for the reference answer, along the model's own update path under teacher forcing. We vary the fraction of the proposed displacement supplied to the readout. A finite-step failure occurs when utility initially improves along the update direction but decreases at the full step. Pathwise curvature shows how this mismatch arises, while a local quadratic model predicts full-step utility changes and useful finite scales. Integrated curvature variation bounds the approximation error of these predictions. Across three models and three tasks, finite-step failures occur in all nine model–task conditions. A fixed quarter step recovers positive reference-utility gains for 72.2–83.2% of finite-step failures across four settings, and geometry-based scale selection further improves recovery on a full test split. These results identify update direction and step scale as distinct factors in recurrent progress, revealing recoverable progress within harmful updates and suggesting that step-scale control could help looped Transformers make better use of their existing computation.

1 INTRODUCTION

Looped Transformer models repeatedly apply shared Transformer layers, making computational depth an inference-time resource (Giannou et al., 2023). Following Universal Transformers (Dehghani et al., 2019), Huginn and Ouro bring recurrent computation to language-model pretraining (Geiping et al., 2025; Zhu et al., 2026), while pretrained models acquire recurrence through parameter-sharing conversions (Bae et al., 2025) or retrofitting with recurrence curricula (McLeish et al., 2025). Expressivity analyses (Giannou et al., 2023; Xu and Sato, 2025) and studies of in-context algorithm learning (Yang et al., 2024; Gatmiry et al., 2024) explain how repeated blocks support iterative computation. Empirical work connects recurrent depth to reasoning and generalisation across problem difficulty or length (Bansal et al., 2022; Fan et al., 2025; Saunshi et al., 2025). Connecting this extra computation to task progress requires understanding its learned updates.

Recent work (Yang et al., 2026) brings the fixed-point stability perspective developed for deep equilibrium models (Bai et al., 2019; 2021) to looped language models, combining Jacobian regularisation and random loop sampling to improve test-time scaling. Input-dependent depth selection (Elbayad et al., 2020) and updates conditioned on recurrent time and step size (Jeddi et al., 2026) adapt computation to different inputs and budgets, while learned correctness signals guide latent optimisation (Du et al., 2026). Empirically, however, additional recurrent computation can make a model worse. A harmful next state leaves an unresolved ambi-

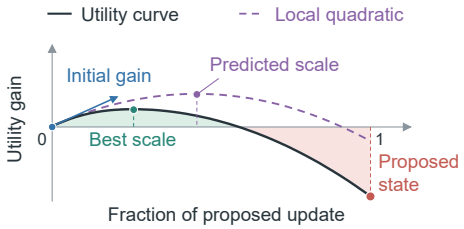

Figure 1: A shorter step can recover progress. Utility gain measures the change in support for the reference answer relative to the current state (0). The proposed state is at (1).

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.

guity: did the model follow an unhelpful direction, or did it move too far along a useful one? Building on interpolation-based analyses of reference-token loss between recurrent states (Asadulaev et al., 2026), we use the same interpolation to examine changes in reference support along the path between recurrent states.

Along the segment between the current and proposed states, support for the reference answer can have a positive initial slope yet end below its starting level, producing the *finite-step failure* illustrated in Figure 1. A sufficiently negative accumulated curvature correction can outweigh the first-order gain. A positive initial slope guarantees improvement only infinitesimally close to the current state. The substantive question is whether this useful direction persists to nontrivial finite scales. When it does, a shorter displacement can reach an intermediate state that supports the reference more strongly than both endpoints. The initial slope and curvature suggest how far to move, while variation in curvature over the segment determines how faithfully this local prediction describes the change in support along the update. This leaves a question: *how much progress remains recoverable when a learned update reaches a harmful endpoint?*

We address this question through pathwise analysis and fixed-direction interventions. Under teacher forcing, reference utility is the mean log-probability assigned to reference solution tokens, or the log-softmax score of the correct option among length-normalised option scores for multiple-choice tasks. It provides a continuous measure of support for the answer. Finite-step failures occur in all nine model–task conditions and account for 1.4–49.2% of harmful updates, including conditions with positive mean gain. Local quadratic geometry predicts full-step gain and useful update scales. A fixed quarter step recovers positive gains in reference utility for 72.2–83.2% of selected failures across four settings, and quadratic selection improves recovery and mean gain on an additional full test set. These recoveries indicate that positive reference-utility progress can remain available along the model's own update direction even when the full-step endpoint is harmful. To characterise the error of the local predictions, we use bounds based on integrated curvature variation along the same path. Interventions measure reference utility after proposal computation and before further recurrence. *Our contributions are*:

1. **Failure diagnosis.** We distinguish locally unhelpful directions from locally useful directions whose full displacement is harmful, and characterise the latter through pathwise curvature.
2. **Finite-scale prediction.** Initial slope and curvature predict full-step utility change and useful finite scales, with error controlled by curvature variation along the path.
3. **Recoverable progress.** Fixed step contractions and quadratic scale selection recover positive reference-utility gains on selected failures, identifying recoverable progress within the proposed update.

2 Related Work

Recurrent Computation and Overthinking. Shared recurrent computation (Schwarzschild et al., 2021; Bansal et al., 2022; Fan et al., 2025; Saunshi et al., 2025) supports depth and length generalisation, while adaptive computation (Graves, 2017; Banino et al., 2021; Elbayad et al., 2020) learns how much processing to allocate to each input. The benefits of additional computation, however, need not persist with increasing depth. Shallow-Deep Networks (Kaya et al., 2019) identify destructive overthinking when deeper processing turns correct predictions into errors, and Think-at-Hard (Fu et al., 2026) observes analogous reversals during latent iterations. To understand how additional recurrence loses task progress, we examine the update producing a harmful endpoint, separating its initial direction from the effect of its full displacement.

Stability and Step Control. Remaining path length and decoder margin (Viakhirev et al., 2026) yield conditions for answer stability, while differences between successive updates (Pappone et al., 2025) provide an early-exit signal. Alongside stopping decisions, residual-dependent damping (Movahedi et al., 2026) and input-dependent channel-wise gating (Park et al., 2026) regulate how individual updates are applied. To understand when a smaller step can recover task progress, we analyse how reference utility changes along a proposed update while holding its direction and readout fixed.

Task Feedback and Update Geometry. Latent thought policy optimisation (Du et al., 2026; Ye et al., 2026), post-training latent refinement (Wang et al., 2026), and manifold steering (Huang et al., 2025) use task feedback or representation geometry to guide internal-state refinement. Recent

analysis of Tiny Recursive Models (Asadulaev et al., 2026) derives reference-token loss derivatives along an interpolation between pre- and post-update distributions. Under affine logit readouts, our state interpolation induces the same distribution path. We extend this analysis to how accumulated curvature can turn local utility improvement into full-step harm, and test whether shortening the same update recovers positive utility gain.

3 METHOD

Each application of a looped Transformer's shared block determines an update direction and a displacement magnitude. For a frozen model and a fixed readout, we define reference utility along the proposed update in § 3.1. We then use pathwise curvature to explain how local improvement can become full-step harm (§ 3.2), and analyse useful step scales through a local quadratic approximation and its error bounds (§ 3.3). Finally, fixed-direction interventions test whether shorter displacements recover immediate reference utility (§ 3.4).

3.1 RECURRENT UPDATES AND TASK UTILITY

Looped language models repeatedly apply a shared transformation (Geiping et al., 2025; Zhu et al., 2026). For a frozen model and a fixed input, write the recurrence as $H_{t+1} = F(H_t)$, where $H_t \in \mathcal{H}$ is the hidden state after $t = 0, 1, \ldots$ iterations, H_0 is the initial state, and $\mathcal{H}$ is a finite-dimensional real inner product space. The map $F : \mathcal{H} \to \mathcal{H}$ is fixed. We analyse a consecutive pair $H = H_t$, $H^+ = H_{t+1}$ and its displacement

$$H^+ := F(H), \qquad D := H^+ - H. \tag{1}$$

The fixed output layers, or readout, map hidden states to output scores. A scalar task utility $U : \Omega \to \mathbb{R}$ evaluates these scores against a fixed reference, where $\Omega \subseteq \mathcal{H}$ is an open set containing the segment from H to H^+. An increase in U defines task progress as stronger support for the reference. Reference utility and discrete answer correctness are distinct outcomes. The state representation, readout, and reference remain fixed throughout each comparison, and Appendix D.2 defines the utilities.

To isolate the effect of displacement magnitude, we preserve the direction proposed by the recurrent model. For a nonzero update, every displacement in the same direction is a positive scalar multiple of that update. We therefore use a dimensionless step scale α to define the interpolated state and its utility gain:

$$\gamma(\alpha) := H + \alpha D, \qquad \phi(\alpha) := U(\gamma(\alpha)) - U(H), \qquad \alpha \in [0, 1]. \tag{2}$$

Thus $\phi(0) = 0$ and $\phi(1) = \Delta U := U(H^+) - U(H)$. An additional recurrent iteration takes the full displacement. For $D \neq 0$, intermediate scales retain its direction while varying the state supplied to the readout, after both endpoints have been computed. We relate this path to its local derivatives under the following assumption.

Assumption 3.1 (Readout Regularity)**.** The utility U is twice continuously differentiable on the open set Ω containing $\{H + \alpha D : \alpha \in [0, 1]\}$.

The evaluated readouts consist of smooth operations under fixed attention masks, so their real-valued utilities satisfy this regularity (Appendix D.1). Since the proposed displacement is held fixed, no smoothness assumption on the recurrent map F is required. Assumption 3.1 gives $\phi \in C^2(\mathcal{O})$ on an open neighbourhood $\mathcal{O}$ of $[0, 1]$. The initial slope A and directional curvature $2Q$ are

$$A := \phi'(0) = \big\langle \nabla U(H), D \big\rangle_{\mathcal{H}}, \qquad Q := \frac{1}{2}\phi''(0) = \frac{1}{2}\big\langle D, \nabla^2 U(H)[D] \big\rangle_{\mathcal{H}}. \tag{3}$$

These coefficients measure the utility response to the shared block's own displacement (derivation in Appendix A.1). The direction is supplied by learned recurrence, and the utility derivatives assess its alignment with the task.

The sign of A describes the initial effect of following the update, whereas ΔU measures the result of the full displacement. We use their joint signs to distinguish two kinds of harmful update.

Definition 3.2 (Finite-Step Failure)**.** An update is progressing if $\phi(1) > 0$ and neutral if $\phi(1) = 0$. A harmful update ($\phi(1) < 0$) is a directional failure if $A \leq 0$ and a finite-step failure if $A > 0$.

In a finite-step failure, the shared block proposes a direction that initially improves utility, but taking its full displacement lowers utility at the next recurrent state. The curvature accumulated within this update determines how the initial improvement is lost.

3.2 Direction–Step Compatibility

A scale is compatible with task progress when $\phi(\alpha) \geq 0$. The integral form of Taylor's theorem expresses this gain as the initial linear gain plus a pathwise curvature correction.

Proposition 3.3 (Pathwise Progress Decomposition)**.** *Let $\mathcal{O} \subseteq \mathbb{R}$ be open with $[0,1] \subseteq \mathcal{O}$. If $\phi \in C^2(\mathcal{O})$ and $\phi(0) = 0$, then for every $\alpha \in [0,1]$,*

$$\phi(\alpha) = \alpha A + \int_0^\alpha (\alpha - s)\phi''(s)\, ds. \tag{4}$$

In particular, a finite-step failure satisfies $\int_0^1 (1-s)\phi''(s)\, ds < -A < 0$.

The proof is given in Appendix A.2. At $\alpha = 1$, the integral corrects the local gain A for directional curvature accumulated across the full update. A finite-step failure occurs when this correction outweighs the positive linear term.

For an affine logit path, the curvature correction admits an exact interpretation in terms of the output distributions. Fix a question i with a finite nonempty set M_i of reference-token indices. Let $p_j(\alpha)$ denote the softmax distribution predicting reference token $j \in M_i$ over a finite nonempty vocabulary.

Proposition 3.4 (Exact Endpoint Decomposition)**.** *Suppose the logits for every $j \in M_i$ are affine in α and U is the mean reference-token log probability. Then*

$$\Delta U = A - \mathcal{C}, \qquad \mathcal{C} := \frac{1}{|M_i|} \sum_{j \in M_i} D_{\mathrm{KL}}(p_j(0) \| p_j(1)) \geq 0. \tag{5}$$

A finite-step failure occurs exactly when $0 < A < \mathcal{C}$. Moreover, $\mathcal{C} = -\int_0^1 (1-s)\phi''(s)\, ds$.

Here D_{KL} denotes Kullback–Leibler divergence. The proof is in Appendix C.2. The full update loses utility when the distributional correction exceeds the first-order gain. Since this correction uses only the endpoint distributions, it also provides a separate numerical check of the pathwise decomposition. For a general readout, we obtain an explicit sufficient range of improving scales by bounding negative directional curvature uniformly along the segment.

Corollary 3.5 (Beneficial Step Scales)**.** *Under Proposition 3.3, suppose $A > 0$ and $\phi''(\alpha) \geq -L_D$ on $[0,1]$ for a constant $L_D \geq 0$. Then for every $\alpha \in [0,1]$,*

$$\phi(\alpha) \geq \alpha A - \frac{L_D}{2}\alpha^2. \tag{6}$$

For $L_D > 0$, every $0 < \alpha \leq 1$ with $\alpha < 2A/L_D$ has positive utility gain. For $L_D = 0$, every $\alpha \in (0,1]$ has positive gain.

The proof also appears in Appendix A.2. The sufficient range turns the directional diagnosis into an intervention: it specifies how far the proposed update can move while improving utility. A valid L_D requires a lower bound on directional curvature along the whole segment. The experiments obtain this bound from an FP64 derivative bound for Ouro's affine readout and have no such bound for Huginn (Appendix E.5). The path defines the progress boundary, the endpoint of its largest initial interval of nonnegative gain:

$$r_1 := \sup\Big\{ r \in [0,1] \,:\, \phi(\beta) \geq 0 \quad \text{for every } \beta \in [0,r] \Big\}. \tag{7}$$

For a finite-step failure, the positive initial derivative and negative endpoint give $0 < r_1 < 1$. The argument is in Appendix A.3. This boundary describes how far progress persists. Selecting a scale further requires locating the utility maximum.

3.3 PREDICTING THE UTILITY-MAXIMISING SCALE

The quadratic model predicts $\widehat{\alpha}$ for the path's continuous optimum $\alpha^\star$:

$$q(\alpha) := A\alpha + Q\alpha^2, \qquad \alpha^\star := \min \arg\max_{\alpha\in[0,1]} \phi(\alpha), \qquad \widehat{\alpha} := \min \arg\max_{\alpha\in[0,1]} q(\alpha). \tag{8}$$

Both optimisation problems use the smallest maximiser to resolve ties. The same local model gives two predictions: $q(1) = A + Q$ for the gain of the full update and $\widehat{\alpha}$ for the scale at which the gain is largest. For $Q < 0$, the predicted scale is

$$\widehat{\alpha} = \mathrm{clip}_{[0,1]}\left(-\frac{A}{2Q}\right). \tag{9}$$

Here $\mathrm{clip}_{[0,1]}(x) := \min\{1, \max\{0, x\}\}$. For $Q \geq 0$, $\widehat{\alpha}$ is 1 if $A + Q > 0$ and 0 otherwise.

For $A > 0, Q < 0$, the ratio $-A/(2Q)$ balances the initial slope against its decrease under the quadratic model. Clipping gives the best scale within the original displacement. The polynomial's positive root, $r^{(2)} = -A/Q$, predicts where that gain returns to zero. It is twice the unclipped maximising scale and may lie beyond the search interval. The scale bounds below do not apply to this root. Under a two-sided curvature bound, Appendix A.3 places $r^{(2)}$ and r_1 in a common interval, and the experiments compare $r^{(2)}$ with the sampled path. Existence of the smallest maximisers and the derivation of Eq. (9) are given in Appendix A.4.

The accuracy of this prediction depends on how the directional slope changes along the segment. We measure the accumulated departure from the initial curvature by

$$C(a) := \int_0^a |\phi''(s) - \phi''(0)|\, ds, \qquad a \in [0, 1]. \tag{10}$$

This quantity measures variation around the curvature already represented by q. A curved path can still have small approximation error: constant curvature gives $C(1) = 0$ and makes q exact, as shown in Appendix A.5.5. Under Assumption 3.1, the full-update prediction error satisfies

$$|\Delta U - (A + Q)| \leq C(1). \tag{11}$$

In particular, $A + Q$ has the same sign as ΔU whenever $|A + Q| > C(1)$ (proof in Appendix A.5.1). For $Q < 0$, this deviation also bounds scale error and regret relative to the initial curvature.

Theorem 3.6 (Integrated Curvature Approximation)**.** *Let $\phi \in C^2(\mathcal{O})$ on an open neighbourhood of $[0, 1]$, and set $A = \phi'(0)$ and $Q = \phi''(0)/2 < 0$. Set $\kappa := -2Q > 0$ and use the maximisers in Eq. (8). Then*

$$|\widehat{\alpha} - \alpha^\star| \leq \min\{1, C(1)/\kappa\}, \qquad 0 \leq \phi(\alpha^\star) - \phi(\widehat{\alpha}) \leq \frac{C(1)^2}{2\kappa}. \tag{12}$$

The bounds hold for every global maximiser of ϕ, including endpoints, without assuming concavity.

The proof and a tighter regret bound for $C(1) > \kappa$ are given in Appendix A.5.3. The same accumulated deviation governs scale error linearly and regret quadratically, relative to the initial negative curvature κ. The bound characterises when local utility geometry remains informative across the shared block's full displacement. The predictor uses derivatives at the current recurrent state. Measurements along the resulting path quantify its approximation error. The pointwise location bound in Appendix A.5.2 additionally retains signed cancellation up to the true maximiser. Appendix A.5.5 treats zero deviation and the sensitivity of the maximising scale when $Q \geq 0$.

Finite measurements can bound the accumulated deviation when variation between samples is controlled. Proposition A.6 constructs lower and upper sums for $C(1)$ from endpoint curvatures and valid Lipschitz constants on each partition cell. Its upper sum $\overline{C}$ can replace $C(1)$ in the endpoint, scale, and regret bounds. A bound on the signed slope residual retains cancellation of curvature deviations (Appendix E.9). Cellwise third-derivative bounds supply these constants, and a global bound M recovers $C(a) \leq Ma^2/2$ under C^3 regularity (Appendix A.5.6). The global bound applies worst-case derivative control throughout the segment. The integrated quantity accumulates the curvature deviation along the actual path. This distinction determines how tightly the path measurements can bound the error of a local prediction. These bounds concern real-valued functions. Their numerical evaluation requires separate floating-point error accounting.

3.4 Prediction and Fixed-Direction Intervention

After computing the next recurrent state, we evaluate the local model on the resulting update. The endpoint comparison measures how closely $A + Q$ predicts ΔU. To evaluate the scale prediction, we additionally measure utility along the segment. These evaluations are made on a finite grid $\mathcal{A} \subseteq [0, 1]$ containing 0 and 1, with the grid optimum as the comparison target:

$$\alpha^\star_{\mathcal{A}} := \min \arg\max_{\alpha \in \mathcal{A}} \phi(\alpha). \tag{13}$$

The grid target uses sampled utilities, while $\widehat{\alpha}$ uses the derivatives at the current recurrent state H. Appendix B gives location and utility bounds relating this grid target to the continuous optimum $\alpha^\star$, under their respective path-shape conditions. We additionally refine the numerical optimum along the same segment, using the readout's concavity when available and evaluating further candidate scales for a general readout. For $A > 0, Q < 0$, we also evaluate the root prediction $r^{(2)}$ on the same grid against the interval ending at the first observed negative gain. This compares the predicted extent of progress with the sampled path. Locating the continuous boundary r_1 additionally requires controlling the behaviour between grid points.

The curve comparison evaluates a scale predicted separately for each update. We additionally measure recovery of finite-step failures under a common reduction in update magnitude. The local positivity argument in Appendix A.3 guarantees a beneficial scale for each such failure. We assess recovery at a shared scale by choosing a prespecified fraction $\alpha_c \in (0, 1)$ and evaluate the readout at $H + \alpha_c D$, holding H, D, and the readout fixed. This is the step-contraction intervention. A failure is recovered when $\phi(\alpha_c) > 0$. The recovery fraction is the proportion of selected failures satisfying this condition, and the mean gain averages $\phi(\alpha_c)$ over that same subset. When a uniform negative-curvature bound is available, we also select a scale separately for each update from the sufficient range in Corollary 3.5 and evaluate its measured gain.

Recovery is defined on finite-step failures, where the current state already dominates the harmful endpoint. To place the finite set of evaluated step scales relative to this endpoint choice, let $\mathcal{B} \subseteq [0, 1]$ contain both endpoints:

$$U_{\text{halt}} := \max\{U(H), U(H^+)\}, \qquad U_{\text{step}} := \max_{\alpha \in \mathcal{B}} U(H + \alpha D). \tag{14}$$

Since this set contains both endpoints, $U_{\text{step}} \geq U_{\text{halt}}$. For a recovered finite-step failure, any evaluated $\alpha \in \mathcal{B}$ with $\phi(\alpha) > 0$ strictly exceeds $U_{\text{halt}} = U(H)$. Thus recovery identifies cases where an intermediate action contains utility unavailable from either original endpoint. All comparisons use the same reference utility before further recurrence.

4 Experiments

We first trace utility across recurrent depth and examine the effect of one additional loop at fixed transitions (§ 4.2). Path measurements then test whether local geometry predicts useful scales within those updates (§ 4.3). Fixed-direction interventions evaluate whether changing the displacement recovers progress within harmful updates (§ 4.4).

4.1 Setup

Models, tasks, and transitions. We study Ouro-1.4B and Ouro-2.6B (Zhu et al., 2026) alongside Huginn-0125 (3.5B) (Geiping et al., 2025) to examine the mechanism across model families and scales, with different recurrent and readout designs. Following mathematical evaluations of looped language models (Yang et al., 2026), we use the complete MATH-500 (Lightman et al., 2024) test set ($N = 500$) and GSM8K (Cobbe et al., 2021) test set ($N = 1319$). We extend the comparison to commonsense completion using the full HellaSwag (Zellers et al., 2019) validation set ($N = 10042$). All three models are evaluated on each task, giving nine conditions. For this matrix, transitions are fixed across tasks: $4 \to 5$ for both Ouro models and $19 \to 20$ for Huginn. A broader depth sweep covers seven labelled splits, including additional question-answering tasks, giving 21 conditions (Appendix E.1). The four original intervention cohorts are the two Ouro models on MATH-500, the Huginn/GSM8K training cohort, and the Ouro-1.4B/HellaSwag validation cohort. We also include a separate CommonsenseQA (Talmor et al., 2019) training extension.

Table 1: Finite-step failures and endpoint prediction. All nine conditions contain finite-step failures. For predicting ΔU, adding Q improves sign accuracy and reduces MAE (bold). Counts are finite-step/harmful updates, and mean gain averages ΔU over all questions. Rank correlation is Spearman between $A + Q$ and ΔU.

Model	Task	Mean gain	Finite-step failures		Sign accuracy (%) ↑		Rank corr. ↑	MAE (10^{-4}) ↓	
		(10^{-3})	Count	Share (%)	A	$A+Q$		A	$A+Q$
Ouro 1.4B	MATH-500	−8.73	179/364	49.18	64.20	**98.20**	0.9929	139.41	**11.99**
	GSM8K	−14.47	429/1066	40.24	67.48	**98.41**	0.9985	147.51	**6.20**
	HellaSwag	−7.75	261/5771	4.52	92.49	**98.87**	0.9983	73.11	**12.45**
Ouro 2.6B	MATH-500	−4.40	95/329	28.88	81.00	**98.60**	0.9973	51.12	**4.10**
	GSM8K	−7.77	325/874	37.19	75.36	**97.19**	0.9952	114.61	**13.13**
	HellaSwag	+1.49	88/4719	1.86	94.24	**99.49**	0.9995	29.56	**3.05**
Huginn 3.5B	MATH-500	+0.75	45/203	22.17	91.00	**99.60**	0.9998	6.81	**0.23**
	GSM8K	−0.48	102/702	14.53	92.27	**99.77**	0.9998	10.39	**0.42**
	HellaSwag	+0.34	65/4695	1.38	98.58	**99.92**	1.0000	2.06	**0.17**

Evaluation protocol. We follow the fixed-direction analysis in § 3.1 with frozen models, measuring support for reference solutions or correct options under teacher forcing. We compare A and $A + Q$ with full-step utility gains, and predicted scales with measured path optima. Interventions test whether shortening the same updates recovers positive utility gains before further recurrence. Additional FP64 path evaluations assess prediction error and curvature bounds. Utility definitions, comparison metrics, numerical precision, and statistical procedures are given in Appendices D.2 and E.2, with FP64 analyses in Appendix E.5.

4.2 RQ1: Local Direction versus Finite Progress

Useful directions within harmful updates. Depth sweeps identify utility declines in 19 of 21 model–task conditions (Appendix E.1). To examine how progress is lost within an additional loop, we compare the initial utility slope with the full-step gain at the fixed transitions in Table 1. Finite-step failures as defined in Definition 3.2 occur in all nine conditions, accounting for 1.4–49.2% of harmful updates: the proposed direction initially improves reference utility, but its full displacement lowers it. These failures also occur in all three conditions with positive mean gain, so the mechanism appears within individual updates even when the population average improves. A harmful full-step outcome can therefore conceal a direction along which a shorter move would improve reference utility. The compact scatter in Figure 2 makes this separation visible across the full matrix. Within Ouro-1.4B's original four-step computation range, a token-level audit finds the same mechanism, with quadratic selection recovering 97.77–99.11% of finite-step failures (Appendix E.8). Additional effect-margin checks are reported in Appendix E.2.

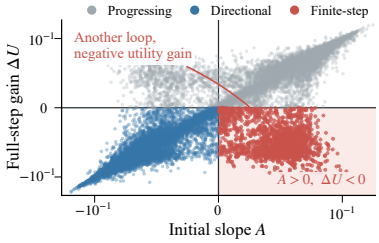

Figure 2: Finite-step failures. Red points fall in the $A > 0, \Delta U < 0$ quadrant: an additional loop lowers reference utility despite a locally improving direction.

Accumulated curvature and endpoint prediction. Proposition 3.3 decomposes full-step gain into the initial slope and an accumulated curvature correction. We test how well local curvature approximates this correction by comparing A and $A + Q$ with the observed ΔU. Adding Q improves sign accuracy in all nine matched conditions, reaching 97.2–99.9%, and reduces MAE by factors of 5.9–29.8 (Table 1). The largest sign-accuracy gains occur on Ouro-1.4B's mathematical tasks, where harmful full updates often have positive initial slopes. These results support the role of curvature in the mismatch between local improvement and full-step harm. Additional checks of the affine-readout identity in Proposition 3.4 are reported in Appendix E.2.

We also examine whether changes in hidden states and outputs track task progress. Across the four original intervention cohorts, state- and output-change diagnostics have absolute Spearman correlation at most 0.285 with ΔU, while the reference-based prediction $A + Q$ exceeds 0.99 on

Table 2: Quadratic scale prediction against grid optima. Results on Ouro/MATH-500 test and Huginn/GSM8K train256, restricted to $A > 0, Q < 0$. Correlation and MAE compare predicted scales with 21-point grid optima. Regret is evaluated at the nearest grid action.

Model / task	n	Spearman	Scale MAE	Grid regret
Ouro-1.4B / MATH-500	315	**0.9951**	0.0211	$\mathbf{1.747 \times 10^{-5}}$
Huginn / GSM8K	129	**0.9998**	0.0072	$\mathbf{3.482 \times 10^{-7}}$

the same teacher-forced states. These results indicate that generic change measures provide limited information about the ranking of utility gains, motivating an analysis of utility along the update direction. Diagnostic definitions and detailed comparisons are provided in Appendix E.3. These comparisons show why a harmful endpoint alone does not identify the source of lost progress. The next question is whether the same local geometry also indicates how far the update should move.

4.3 RQ2: Predicting the Scale of Progress

Prediction along the fixed direction. Accurate endpoint prediction motivates evaluating scales inside the update. On the primary subset $A > 0, Q < 0$, we compare the quadratic maximiser with the measured grid optimum, keeping the initial coefficients fixed. The first-order rule takes the full step throughout this subset. On Ouro-1.4B/MATH-500 and Huginn/GSM8K train256, predicted and empirical scales have correlation above 0.995, and the quadratic rule reduces mean grid regret by more than two orders of magnitude relative to the first-order rule (Table 2 and Figure 3(e)). The same panel compares the quadratic rule with all five fixed scales. This grid-regret comparison evaluates the utility at the predicted scale, complementing the agreement in scale. Within the primary subset, it tests whether adding curvature to the first-order full-step choice improves utility. The best fixed scale in hindsight differs between these evaluation sets: 0.5 for Ouro and 1 for Huginn. Even against these separately selected choices, the quadratic rule reduces mean grid regret by more than two orders of magnitude. Its advantage therefore extends beyond replacing the full step with a uniformly smaller displacement. Appendix E.4 gives the comparison protocol and results on the additional tasks.

Intermediate optima and prediction error. To test whether local quadratic geometry locates intermediate utility maxima, we examine paths whose numerical optima lie strictly inside $[0, 1]$. Mean scale error is below 0.019 across 269 Ouro-1.4B/MATH-500 cases and 0.018 across 36 Huginn/GSM8K cases (Table 9). These results support the prediction in Eq. (9): the initial slope and curvature contain information about how far to move along the proposed direction (Figure 3(a–d)). Theorem 3.6 relates scale-prediction error to accumulated curvature variation. We evaluate this bound on the primary $A > 0, Q < 0$ records from both Ouro models on the complete MATH-500 and GSM8K test sets. Across these four conditions, the integrated bound $\overline{C}/\kappa$ has median 0.058–0.114, compared with 2.84–4.61 for the global bound $M/(2\kappa)$. It covers the numerical scale-error intervals of all 2034 primary updates and is below the full search range on 99.8% of them (Table 14). These quantities instantiate the two controls in Theorem 3.6: the ratio $\overline{C}/\kappa$ controls scale-location error, while the corresponding quadratic term $\overline{C}^2/(2\kappa)$ controls utility regret. The comparison is between path-specific curvature variation and the global third-derivative envelope. Both are evaluated on the same learned recurrent transitions used for scale prediction, so the tighter integrated bound reflects the geometry of the observed model trajectories. Thus, accounting for curvature variation along the path gives informative error bounds for the same local prediction. Appendix E.9 gives the bound distributions and numerical optimum checks for all four conditions. The grid comparison therefore tests the utility of the predicted action, while the bound comparison tests how far the local approximation remains informative along the observed path.

4.4 RQ3: Recovering Progress by Step Contraction

The preceding analyses show that local geometry predicts both full-step utility change and useful intermediate scales. Here, we use fixed-direction interventions to test whether shorter updates recover the progress suggested by this geometry. Additional comparisons of the predicted zero crossing with sampled utility paths are reported in Appendix E.4.

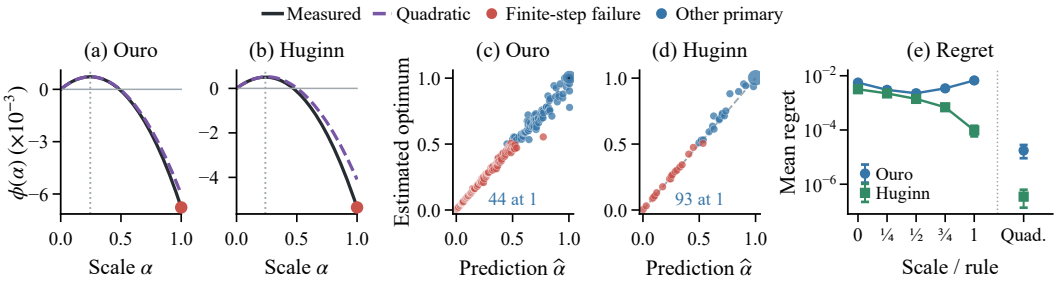

Figure 3: Scale prediction on Ouro-1.4B/MATH-500 and Huginn/GSM8K. (a,b) Example utility paths and local quadratic approximations. (c,d) Predicted scales versus numerical optima. (e) Grid regret for fixed scales and quadratic selection. Labels in (c,d) count overlapping points at $(1, 1)$.

A common step contraction across questions. Finite-step failure guarantees an improving neighbourhood along the proposed direction, and Corollary 3.5 relates its extent to directional curvature. We first test how often a common scale $\alpha_c = 0.25$ reaches this improving region across questions, using the fixed-direction intervention in § 3.4. On the four original intervention cohorts, this quarter step recovers 72.2–83.2% of finite-step failures (Figure 4). Each recovery therefore tests the endpoint relation in Eq. (14): the intermediate state has higher reference utility than both endpoints of the original harmful update. These results show that a single fraction reaches useful intermediate states across most selected failures in each evaluation set. Effect margins, task extensions, and transition selection are detailed in Appendix E.2.

Selecting a scale from local geometry. To test whether adapting the scale to each update improves recovery, we apply the quadratic rule in Eq. (8) using local derivatives of reference utility. On the same Ouro-1.4B/GSM8K test updates analysed in RQ1 at $4 \to 5$, it recovers 99.1% of the 429 finite-step failures, compared with 70.4% for the quarter step. Across all 1319 questions, the rule yields a mean gain of 2.766×10^{-3} and a paired advantage of 3.610×10^{-3} over the quarter step. Under reference-conditioned immediate readout, this corresponds to 425/429 recovered failures, compared with 302/429 for the quarter step. It therefore quantifies the immediate recovery obtained by adapting the scale along the frozen update direction. The direction and recurrent endpoints are unchanged in this comparison, so the difference measures the value of adapting the step scale to local geometry. The recovery phenomenon also appears on Ouro-2.6B/HellaSwag, where the quadratic rule recovers 88.0–100% of finite-step failures across four separately evaluated transitions from $1 \to 2$ through $4 \to 5$ (Appendix E.7). These results connect the scale prediction examined in RQ2 to improved recovery along the model's own update directions. The intervention protocol is detailed in Appendix E.6. A curvature-bound selection check is reported separately in Appendix E.5.

Additional evidence and future use. Supplementary experiments extend recovery to CommonsenseQA, with protocol details in Appendix E.2. These gains in immediate reference utility motivate using the geometry as offline supervision for learned exit and step-scale decisions (Appendix E.8).

5 Conclusion

This work characterises finite-step failure in looped Transformers: an additional recurrent update can lower reference utility despite a locally improving direction. For a frozen model and a fixed readout, accumulated curvature explains this mismatch, while local quadratic geometry predicts useful step scales and curvature variation bounds approximation error. Under teacher forcing, we observe recovery of finite-step failures in both model families when evaluating shorter displacements along the same direction before further recurrence. A fixed quarter step recovers 72.2–83.2% of selected finite-step failures across four settings. On Ouro-1.4B/GSM8K test, quadratic scale selection recovers 99.1% of finite-step failures, compared with 70.4% for the quarter step. These results show that, under reference-conditioned immediate readout, scale adaptation can recover reference-utility progress from harmful endpoints along the model's proposed direction. For looped Transformers, this distinction suggests treating step scale as a first-class control alongside recurrent depth in future model designs.

LLM USE DISCLOSURE

In this work, we used LLM-based assistants for refinement of mathematical exposition and derivations and interpretation, manuscript drafting and editing, and visualization code. The assistants were not used to generate datasets, experimental results, or reported numerical values. The authors reviewed AI-assisted text and code against the paper's definitions, source literature, and recorded analysis outputs, and take responsibility for the final content and artifacts.

REPRODUCIBILITY STATEMENT

This work studies publicly released looped Transformer checkpoints on the MATH-500, GSM8K, HellaSwag, and CommonsenseQA benchmarks. The experiments used more than 100 GPU-hours on a shared NVIDIA H200 GPU cluster. The reported utilities, derivatives, interventions, and derived results were checked against the recorded evaluation outputs and reproduced across independent analysis passes. The theoretical statements are made under explicit regularity assumptions, with proofs in Appendices A–C.2; the model instantiations, evaluation sets, intervention procedures, and numerical validation protocols are recorded in Appendices D, E.2, and E.6. Code and analysis artifacts will be released upon acceptance.

REFERENCES

Arip Asadulaev, Rayan Banerjee, Fakhri Karray, and Martin Takáč. Latent reasoning in TRMs is secretly a policy improvement operator. In *Forty-third International Conference on Machine Learning*, 2026.

Sangmin Bae, Adam Fisch, Hrayr Harutyunyan, Ziwei Ji, Seungyeon Kim, and Tal Schuster. Relaxed recursive transformers: Effective parameter sharing with layer-wise lora. In *13th ICLR 2025: Singapore*, 2025.

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. Deep equilibrium models. In *Advances in Neural Information Processing Systems*, 2019.

Shaojie Bai, Vladlen Koltun, and J. Zico Kolter. Stabilizing equilibrium models by jacobian regularization. In *38th ICML 2021: Virtual Event*, pages 554–565, 2021.

Andrea Banino, Jan Balaguer, and Charles Blundell. PonderNet: Learning to Ponder. In *ICML 2021 Workshops: AutoML*, 2021.

Arpit Bansal, Avi Schwarzschild, Eitan Borgnia, Zeyad Emam, Furong Huang, Micah Goldblum, and Tom Goldstein. End-to-end algorithm synthesis with recurrent networks: Extrapolation without overthinking. In *Advances in Neural Information Processing Systems*, 2022.

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. *arXiv preprint arXiv:2110.14168*, 2021.

Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, and Lukasz Kaiser. Universal transformers. In *7th ICLR 2019: New Orleans, LA, USA*, 2019.

Hanwen Du, Yuxin Dong, and Xia Ning. Latent thinking optimization: Your latent reasoning language model secretly encodes reward signals in its latent thoughts. In *International Conference on Learning Representations*, 2026.

Maha Elbayad, Jiatao Gu, Edouard Grave, and Michael Auli. Depth-adaptive transformer. In *International Conference on Learning Representations*, 2020.

Ying Fan, Yilun Du, Kannan Ramchandran, and Kangwook Lee. Looped transformers for length generalization. In *International Conference on Learning Representations*, 2025.

Tianyu Fu, Yichen You, Zekai Chen, Guohao Dai, Huazhong Yang, and Yu Wang. Think-at-hard: Selective latent iterations to improve reasoning language models. In *Forty-third International Conference on Machine Learning*, 2026.

Khashayar Gatmiry, Nikunj Saunshi, Sashank J. Reddi, Stefanie Jegelka, and Sanjiv Kumar. Can looped transformers learn to implement multi-step gradient descent for in-context learning? In *41st ICML 2024: Vienna, Austria*, pages 15130–15152, 2024.

Jonas Geiping, Sean McLeish, Neel Jain, John Kirchenbauer, Siddharth Singh, Brian Bartoldson, Bhavya Kailkhura, Abhinav Bhatele, and Tom Goldstein. Scaling up test-time compute with latent reasoning: A recurrent depth approach. In *Conference on Neural Information Processing Systems*, 2025.

Angeliki Giannou, Shashank Rajput, Jy-Yong Sohn, Kangwook Lee, Jason D. Lee, and Dimitris Papailiopoulos. Looped transformers as programmable computers. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 11398–11442. PMLR, 23–29 Jul 2023.

Alex Graves. Adaptive computation time for recurrent neural networks, 2017.

Yao Huang, Huanran Chen, Shouwei Ruan, Yichi Zhang, Xingxing Wei, and Yinpeng Dong. Mitigating overthinking in large reasoning models via manifold steering. In *Conference on Neural Information Processing Systems*, 2025.

Ahmadreza Jeddi, Marco Ciccone, and Babak Taati. Loopformer: Elastic-depth looped transformers for latent reasoning via shortcut modulation. In *International Conference on Learning Representations*, 2026.

Yigitcan Kaya, Sanghyun Hong, and Tudor Dumitras. Shallow-deep networks: Understanding and mitigating network overthinking. In *36th ICML 2019: Long Beach, California, USA*, pages 3301–3310, 2019.

Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. In *12th ICLR 2024: Vienna, Austria*, 2024.

Sean Michael McLeish, Ang Li, John Kirchenbauer, Dayal Singh Kalra, Brian R. Bartoldson, Bhavya Kailkhura, Avi Schwarzschild, Jonas Geiping, Micah Goldblum, and Tom Goldstein. Teaching pretrained language models to think deeper with retrofitted recurrence. In *NeurIPS 2025 Workshop on Efficient Reasoning*, 2025.

Sajad Movahedi, Vera Milovanović, Shlomo Libo Feigin, Alexander Theus, Thomas Hofmann, Valentina Boeva, T. Konstantin Rusch, and Antonio Orvieto. Fixed-point reasoners: Stable and adaptive deep looped transformers, 2026.

Francesco Pappone, Donato Crisostomi, and Emanuele Rodolà. Two-scale latent dynamics for recurrent-depth transformers. In *UniReps: 3rd Edition of the Workshop on Unifying Representations in Neural Models*, 2025.

Taekhyun Park, Yongjae Lee, Dohee Kim, and Hyerim Bae. Loopus: Recasting pretrained llms into looped latent refinement models, 2026.

Nikunj Saunshi, Nishanth Dikkala, Zhiyuan Li, Sanjiv Kumar, and Sashank J. Reddi. Reasoning with latent thoughts: On the power of looped transformers. In *13th ICLR 2025: Singapore*, 2025.

Avi Schwarzschild, Eitan Borgnia, Arjun Gupta, Furong Huang, Uzi Vishkin, Micah Goldblum, and Tom Goldstein. Can you learn an algorithm? generalizing from easy to hard problems with recurrent networks. In *Advances in Neural Information Processing Systems*, 2021.

Alon Talmor, Jonathan Herzig, Nicholas Lourie, and Jonathan Berant. CommonsenseQA: A question answering challenge targeting commonsense knowledge. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 2019.

Ivan Viakhirev, Kirill Borodin, Amirah Almutairi, Serguei Barannikov, Maxim Abramov, and Grach Mkrtchian. Think shallow, solve deep: Controlling recurrent dynamics for reliable test-time depth, 2026.

Xinyuan Wang, Dongjie Wang, Wangyang Ying, Haoyue Bai, Nanxu Gong, Sixun Dong, Kunpeng Liu, and Yanjie Fu. Efficient post-training refinement of latent reasoning in large language models. In *40th AAAI 2026: Singapore*, pages 33692–33700, 2026. doi: 10.1609/aaai.v40i40.40659.

Kevin Xu and Issei Sato. On expressive power of looped transformers: Theoretical analysis and enhancement via timestep encoding. In *International Conference on Machine Learning*, 2025.

Liu Yang, Kangwook Lee, Robert D. Nowak, and Dimitris Papailiopoulos. Looped transformers are better at learning learning algorithms. In *International Conference on Learning Representations*, 2024.

Xiao-Wen Yang, Ziyu Han, Xi-Hua Zhang, Wen-Da Wei, Jie-Jing Shao, Lan-Zhe Guo, and Yu-Feng Li. Stabilizing recurrent dynamics for test-time scalable latent reasoning in looped language models. In *Forty-third International Conference on Machine Learning*, 2026.

Wengao Ye, Yan Liang, and Lianlei Shan. Thinking on the fly: Test-time reasoning enhancement via latent thought policy optimization. In *International Conference on Learning Representations*, 2026.

Rowan Zellers, Ari Holtzman, Yonatan Bisk, Ali Farhadi, and Yejin Choi. HellaSwag: Can a machine really finish your sentence? In *Annual Meeting of the Association for Computational Linguistics*, 2019.

Rui-Jie Zhu, Zixuan Wang, Kai Hua, Tianyu Zhang, Ziniu Li, Haoran Que, Boyi Wei, Zixin Wen, Fan Yin, He Xing, Lu Li, Jiajun Shi, Kaijing Ma, Shanda Li, Taylor Kergan, Andrew Smith, Xingwei Qu, Mude Hui, Bohong Wu, Qiyang Min, Hongzhi Huang, Xun Zhou, Wei Ye, Jiaheng Liu, Jian Yang, Yunfeng Shi, Chenghua Lin, Enduo Zhao, Tianle Cai, Ge Zhang, Wenhao Huang, Yoshua Bengio, and Jason Eshraghian. Scaling latent reasoning via looped language models, 2026.

A Proofs of the Path Analysis

We use the state space, utility, displacement, and path defined in § 3.1. In particular, $D = F(H) - H$ is fixed, $\phi(\alpha) = U(H + \alpha D) - U(H)$, $A = \phi'(0)$, and $Q = \phi''(0)/2$. Smoothness is imposed on an open neighbourhood of the tested segment, and endpoint derivatives use the resulting real-valued continuation.

A.1 Directional Derivatives

Derivation of Eq. (3). The affine path has derivative $\gamma'(\alpha) = D$. Differentiating $\phi(\alpha) = U(\gamma(\alpha)) - U(H)$ by the chain rule gives $\phi'(\alpha) = dU_{\gamma(\alpha)}[D]$. In the inner product space, the differential is represented by the gradient. Differentiating its inner product with the fixed vector D then gives

$$\phi'(\alpha) = \big\langle \nabla U(\gamma(\alpha)), D \big\rangle_{\mathcal{H}}, \qquad \phi''(\alpha) = \big\langle D, \nabla^2 U(\gamma(\alpha))[D] \big\rangle_{\mathcal{H}}. \tag{15}$$

Evaluation at zero gives $\phi'(0) = A$ and $\phi''(0) = 2Q$. The identities involve the fixed vector D and the derivatives of U at points of the segment. □

A.2 Finite Progress and Beneficial Scales

Lemma A.1 (Increment Comparison)**.** *Let f, g be continuously differentiable on an open neighbourhood of $[0, 1]$. If $f'(s) \le g'(s)$ for every $s \in [0, 1]$, then $f(\alpha) - f(0) \le g(\alpha) - g(0)$ for every $\alpha \in [0, 1]$.*

Proof. Continuity makes both derivatives integrable on $[0, \alpha]$. The fundamental theorem of calculus and monotonicity of the integral give

$$f(\alpha) - f(0) = \int_0^\alpha f'(s)\, ds \le \int_0^\alpha g'(s)\, ds = g(\alpha) - g(0). \tag{16}$$

□

Proof of Proposition 3.3. Fix $\alpha \in [0, 1]$ and define $g_\alpha(s) := \phi(s) + (\alpha - s)\phi'(s)$. The product rule gives $g'_\alpha(s) = (\alpha - s)\phi''(s)$. This derivative is continuous on $[0, \alpha]$, so the fundamental theorem of calculus yields

$$\int_0^\alpha (\alpha - s)\phi''(s)\, ds = g_\alpha(\alpha) - g_\alpha(0) = \phi(\alpha) - \alpha A. \tag{17}$$

The last equality uses $\phi(0) = 0$ and $\phi'(0) = A$. At a finite-step failure, $\phi(1) < 0$ and $A > 0$, so $\int_0^1 (1 - s)\phi''(s)\, ds = \phi(1) - A < -A < 0$. □

Proof of Corollary 3.5. First compare the functions $s \mapsto -L_D s$ and $s \mapsto \phi'(s)$. Their derivatives satisfy $-L_D \le \phi''(s)$, so Lemma A.1 gives $\phi'(s) \ge A - L_D s$ on $[0, 1]$. Next define $b(s) := As - L_D s^2/2$, whose derivative is $A - L_D s$. Applying the same comparison to b and ϕ, with $b(0) = \phi(0) = 0$, gives

$$\phi(\alpha) \ge \alpha A - \frac{L_D}{2}\alpha^2 = \alpha \left(A - \frac{L_D}{2}\alpha \right). \tag{18}$$

For $L_D > 0$, this expression is positive whenever $0 < \alpha < 2A/L_D$. For $L_D = 0$, it equals $\alpha A > 0$ at every positive scale in the interval. □

A.3 The Continuous Progress Boundary

Justification of $0 < r_1 < 1$ for a finite-step failure. Differentiability at zero, $\phi(0) = 0$, and $A > 0$ imply $\phi(\alpha)/\alpha \to A$ as $\alpha \downarrow 0$. Hence $\phi(\alpha) > 0$ on some interval $(0, \varepsilon)$. Set $r_0 = \min\{\varepsilon/2, 1/2\}$. Then r_0 belongs to the prefix set in Eq. (7). This set is nonempty and bounded above by one, so its supremum satisfies $r_1 \ge r_0 > 0$. Since $\phi(1) < 0$, continuity at one gives $\eta > 0$ such that $|\alpha - 1| < \eta$ implies $\phi(\alpha) < 0$. Put $b = 1 - \min\{\eta/2, 1/2\}$. Then $0 \le b < 1$ and $\phi(b) < 0$. Every member of the prefix set is strictly less than b, making b an upper bound. Hence $r_1 \le b < 1$. □

The local positive interval requires only the derivative at zero. The curvature bound in Corollary 3.5 additionally gives an explicit sufficient range of scales. A two-sided curvature bound also places the boundary and the quadratic root prediction in a common interval.

Proposition A.2 (Root Prediction under Two-Sided Curvature Bounds)**.** *Under Proposition 3.3, suppose $A > 0$ and $-L_D \le \phi''(s) \le -\mu$ on $[0,1]$ for constants $0 < \mu \le L_D$. Then $\kappa = -\phi''(0) \in [\mu, L_D]$, so $r^{(2)} = -A/Q = 2A/\kappa \in [2A/L_D,\ 2A/\mu]$, and $r_1 \ge \min\{1, 2A/L_D\}$. If moreover $2A/\mu < 1$, then $r_1 \le 2A/\mu$, and consequently*

$$|r^{(2)} - r_1| \le 2A\left(\frac{1}{\mu} - \frac{1}{L_D}\right). \tag{19}$$

Proof. The membership of κ and hence of $r^{(2)}$ follows from the curvature bounds at $s = 0$. For the lower bound on r_1, Corollary 3.5 gives $\phi(\alpha) \ge \alpha(A - L_D\alpha/2) > 0$ for $0 < \alpha < 2A/L_D$. With $\phi(0) = 0$, every $r < \min\{1, 2A/L_D\}$ belongs to the prefix set in Eq. (7), so its supremum is at least $\min\{1, 2A/L_D\}$. For the upper bound, apply Lemma A.1 to ϕ' and $s \mapsto A - \mu s$, whose derivatives satisfy $\phi'' \le -\mu$. This gives $\phi'(s) \le A - \mu s$ on $[0,1]$. Applying the same comparison to ϕ and $b(s) = As - \mu s^2/2$ gives $\phi(\alpha) \le \alpha(A - \mu\alpha/2)$, which is negative for $\alpha > 2A/\mu$. If $2A/\mu < 1$, then $\phi < 0$ on $(2A/\mu, 1]$, so no $r > 2A/\mu$ belongs to the prefix set and $r_1 \le 2A/\mu$. Both r_1 and $r^{(2)}$ then lie in $[2A/L_D, 2A/\mu]$, whose length is the right-hand side of Eq. (19). □

In the affine-logit case, ϕ'' is the negative mean variance in Eq. (53), so μ and L_D bound the minimum and maximum of that variance along the segment. The experiments evaluate only the upper envelope L_D and compare $r^{(2)}$ with the sampled crossings (Appendix E.4). No bound of the form Eq. (19) is computed.

A.4 Constrained Quadratic Maximisation

Existence and tie convention in Eq. (8). Let f be either ϕ or q. Compactness of $[0,1]$ and continuity of f give a maximiser b. The set $S = [0,1] \cap f^{-1}(\{f(b)\})$ is a closed subset of a compact interval and contains b. Minimising the identity function on S gives a point $a \in S$ no larger than any other member. Since $f(a) = f(b)$, this point maximises f on $[0,1]$ and is its smallest maximiser. If a and $\widetilde{a}$ both have this property, maximality gives $f(a) = f(\widetilde{a})$, and their respective minimality gives $a \le \widetilde{a}$ and $\widetilde{a} \le a$. Thus they coincide. □

Derivation of Eq. (9). Suppose $Q < 0$. For any candidate $c \in [0,1]$,

$$q(x) - q(c) = (A + 2Qc)(x - c) + Q(x - c)^2. \tag{20}$$

If $(A + 2Qc)(x - c) \le 0$ for every $x \in [0,1]$, this identity makes c a maximiser, and the strict inequality $Q(x-c)^2 < 0$ for $x \ne c$ makes it unique. Write $u = -A/(2Q)$, so $A + 2Qu = 0$. If $u \le 0$, then $A \le 0$ and $c = 0$ satisfies the required inequality. If $u \ge 1$, then $A + 2Q \ge 0$ and $c = 1$ satisfies it. In the remaining case $0 < u < 1$, take $c = u$, where the derivative vanishes. These branches give the clipped formula.

For $Q \ge 0$, the inequality $x^2 \le x$ on $[0,1]$ gives $q(x) \le x(A + Q)$. If $A + Q > 0$, this is at most $A + Q = q(1)$, with strict inequality when $x < 1$. If $A + Q \le 0$, it is at most zero, and $c = 0$ is the smallest maximiser. This includes $Q = 0$ and the constant-zero quadratic. For $A > 0, Q < 0$, the factorisation $q(\alpha) = \alpha(A + Q\alpha)$ also gives the positive root $r^{(2)} = -A/Q = 2(-A/(2Q))$. □

A.5 Quadratic Approximation Error

Lemma A.3 (Power Increment Bound)**.** *Let f be continuously differentiable on an open neighbourhood of $[0,1]$, and suppose $|f'(s)| \le Bs^n$ on $[0,1]$ for $B \in \mathbb{R}$ and an integer $n \ge 0$. Then*

$$|f(\alpha) - f(0)| \le \frac{B}{n+1}\alpha^{n+1}, \qquad \alpha \in [0,1]. \tag{21}$$

Proof. The derivative bounds are $-Bs^n \le f'(s) \le Bs^n$. The two bounding polynomials $g_\pm(s) = \pm Bs^{n+1}/(n+1)$ have derivatives $\pm Bs^n$ and vanish at zero. Applying Lemma A.1 first to f, g_+ and then to g_-, f gives the two sides of the absolute-value inequality. □

A.5.1 Endpoint Prediction Error

Derivation of Eq. (11). Under Assumption 3.1, the fundamental theorem of calculus gives $\phi'(a) - q'(a) = \int_0^a [\phi''(s) - \phi''(0)]\, ds$, whose absolute value is at most $C(a) \le C(1)$ on $[0,1]$. Applying the mean value inequality to $\phi - q$ between zero and one, with $\phi(0) = q(0) = 0$, yields $|\Delta U - (A+Q)| \le C(1)$. If $A + Q > C(1)$, then $\Delta U \ge A + Q - C(1) > 0$, and if $A + Q < -C(1)$, then $\Delta U \le A + Q + C(1) < 0$. This proves the sign statement without a sign restriction on A or Q. □

A.5.2 Integrated Curvature Deviation

Proposition A.4 (Scale Error from Integrated Curvature Deviation)**.** *Let $\phi \in C^2(\mathcal{O})$ on an open neighbourhood $\mathcal{O}$ of $[0,1]$, and suppose $\phi''(0) < 0$. Set $\kappa := -\phi''(0) = -2Q > 0$, and let $\alpha^\star$ and $\widehat{\alpha}$ be as in Eq. (8). Then*

$$|\widehat{\alpha} - \alpha^\star| \le \frac{\left|\int_0^{\alpha^\star} [\phi''(s) - \phi''(0)]\, ds\right|}{\kappa} \le \frac{C(\alpha^\star)}{\kappa} \le \frac{C(1)}{\kappa}. \tag{22}$$

The result includes endpoint maximisers and does not require concavity of ϕ.

Proof of Proposition A.4. Continuity on the compact interval ensures existence of both maximisers. The argument applies to any global maximiser $\alpha^\star$ of ϕ, and the smallest-maximiser convention merely fixes its name. Since $\kappa = -\phi''(0) > 0$, the quadratic $q(\alpha) = \phi'(0)\alpha - \kappa\alpha^2/2$ is strictly concave, so its constrained maximiser $\widehat{\alpha}$ is unique.

First, let f be differentiable at a constrained maximiser $a \in [0,1]$. For any $b \in [0,1]$, the curve $t \mapsto a + t(b-a)$ remains feasible for $t \in [0,1]$. Optimality implies $f(a + t(b-a)) - f(a) \le 0$ for $t > 0$. Dividing by t and taking $t \downarrow 0$ gives

$$f'(a)(b - a) \le 0. \tag{23}$$

This reasoning uses only feasible one-sided displacements, so it also applies when $a = 0$ or $a = 1$. It does not assert that $f'(a) = 0$ there.

Put $\delta := \widehat{\alpha} - \alpha^\star$. Applying Eq. (23) first to ϕ and then to q yields

$$\phi'(\alpha^\star)\delta \le 0, \qquad q'(\widehat{\alpha})\delta \ge 0. \tag{24}$$

Since $q'(\alpha) = \phi'(0) - \kappa\alpha$, we have $q'(\alpha^\star) - q'(\widehat{\alpha}) = \kappa\delta$. Therefore

$$\begin{aligned} \kappa\delta^2 &= \big(q'(\alpha^\star) - q'(\widehat{\alpha})\big)\delta \\ &\le q'(\alpha^\star)\delta \\ &\le \big(q'(\alpha^\star) - \phi'(\alpha^\star)\big)\delta \\ &\le |q'(\alpha^\star) - \phi'(\alpha^\star)|\, |\delta|. \end{aligned} \tag{25}$$

For $\delta \neq 0$, division by $\kappa|\delta| > 0$ gives

$$|\delta| \le \frac{|\phi'(\alpha^\star) - q'(\alpha^\star)|}{\kappa}. \tag{26}$$

For $\delta = 0$, the same inequality holds because its right-hand side is nonnegative.

Next, $q'(a) = \phi'(0) + \phi''(0)a$ and the fundamental theorem of calculus give, for every $a \in [0,1]$,

$$\phi'(a) - q'(a) = \phi'(a) - \phi'(0) - a\phi''(0) = \int_0^a [\phi''(s) - \phi''(0)]\, ds. \tag{27}$$

All integrands are continuous on the compact segment, hence integrable. Taking absolute values gives

$$\left|\int_0^a [\phi''(s) - \phi''(0)]\, ds\right| \le \int_0^a |\phi''(s) - \phi''(0)|\, ds = C(a). \tag{28}$$

For $0 \le a \le b \le 1$, $C(b) - C(a) = \int_a^b |\phi''(s) - \phi''(0)|\, ds \ge 0$. Thus $C(\alpha^\star) \le C(1)$. Substitution into Eq. (26) proves the complete chain in Eq. (22). Neither this proof nor the distance conclusion uses $\phi(0) = 0$, $\phi'(0) > 0$, or concavity of the actual path. □

A.5.3 Integrated Location and Regret Bounds

Proof of Theorem 3.6. Continuity on $[0,1]$ gives a global maximiser, and strict concavity of q gives its unique maximiser $\widehat{\alpha}$. Fix any global maximiser $\alpha^\star$ of ϕ. Define the derivative residual and value residual by $e(a) := \phi'(a) - q'(a)$ and $\mathcal{R}(a) := \phi(a) - q(a)$. Eq. (27) and monotonicity of C give $|e(a)| \le C(a) \le C(1)$ on $[0,1]$. Put $\varepsilon := C(1) \ge 0$.

Proposition A.4 bounds the location error by ε/κ. Both maximisers belong to $[0,1]$, so their distance is also at most one. This proves the first conclusion of Eq. (12).

For the value bound, $\mathcal{R}' = e$ and the derivative bound on the convex interval imply, by the mean value inequality,

$$|\mathcal{R}(v) - \mathcal{R}(u)| \le \varepsilon |v - u|, \qquad u, v \in [0,1]. \tag{29}$$

This step uses differentiability and the uniform derivative bound. It applies in either order of u, v. The exact quadratic expansion at its constrained maximiser gives

$$q(a) - q(\widehat{\alpha}) = q'(\widehat{\alpha})(a - \widehat{\alpha}) - \frac{\kappa}{2}(a - \widehat{\alpha})^2 \le -\frac{\kappa}{2}(a - \widehat{\alpha})^2. \tag{30}$$

The inequality uses the feasible first-order condition in Eq. (23). With $d := |\alpha^\star - \widehat{\alpha}|$, decompose the utility difference into the residual increment and the quadratic gap:

$$\begin{aligned} 0 \le \phi(\alpha^\star) - \phi(\widehat{\alpha}) &\le \varepsilon d - \frac{\kappa}{2} d^2 \\ &= \frac{\varepsilon^2}{2\kappa} - \frac{\kappa}{2}(d - \varepsilon/\kappa)^2 \le \frac{\varepsilon^2}{2\kappa}. \end{aligned} \tag{31}$$

The lower bound follows from maximality of $\alpha^\star$. Substitution of $\varepsilon = C(1)$ proves the result. No value of $\phi(0)$ enters the residual increment, so adding a constant to the utility leaves the argument unchanged. □

Corollary A.5 (Refinement by the Scale Domain)**.** *Under Theorem 3.6,*

$$\phi(\alpha^\star) - \phi(\widehat{\alpha}) \le \begin{cases} C(1)^2/(2\kappa), & C(1) \le \kappa, \\ C(1) - \kappa/2, & C(1) > \kappa. \end{cases} \tag{32}$$

The same conclusion holds with any upper bound $\overline{C} \ge C(1)$ in place of $C(1)$.

Proof. The distance d in Eq. (31) belongs to $[0,1]$. If $\varepsilon \le \kappa$, the quadratic in d has its maximum at ε/κ. If $\varepsilon > \kappa$, its maximum on $[0,1]$ is $\varepsilon - \kappa/2$ at one. For an upper bound $\overline{C}$, the same proof starts with $|e(a)| \le \overline{C}$, and nonnegativity follows from $0 \le C(1) \le \overline{C}$. □

A.5.4 Finite-Partition Enclosures

To connect this integral to finite measurements, let $g(s) := \phi''(s) - \phi''(0)$ and partition the interval as $0 = t_0 < \cdots < t_N = 1$, with $N \ge 1$. For each cell $\ell = 0, \ldots, N-1$, suppose $|g(u) - g(v)| \le L_\ell |u - v|$ for all $u, v \in [t_\ell, t_{\ell+1}]$, where $L_\ell \ge 0$. Define

$$h_\ell := t_{\ell+1} - t_\ell, \quad T_\ell := \frac{h_\ell}{2}(|g(t_\ell)| + |g(t_{\ell+1})|), \quad P_\ell := \frac{L_\ell h_\ell^2}{4}. \tag{33}$$

Proposition A.6 (Finite-Partition Curvature Bounds)**.** *Under the assumptions of Theorem 3.6 and these cellwise conditions,*

$$\underline{C} := \sum_{\ell=0}^{N-1} \max\{0, T_\ell - P_\ell\} \le C(1) \le \overline{C} := \sum_{\ell=0}^{N-1} (T_\ell + P_\ell). \tag{34}$$

Both bounds in Theorem 3.6 remain valid when $C(1)$ is replaced by $\overline{C}$.

Lemma A.7 (Nearest-Endpoint Integral Bound)**.** *Let $l < r$ and let g be L-Lipschitz on $[l, r]$, with $L \ge 0$. For $h = r - l$, $T = h(|g(l)| + |g(r)|)/2$, and $P = Lh^2/4$,*

$$\left| \int_l^r |g(s)|\, ds - T \right| \le P, \qquad \max\{0, T - P\} \le \int_l^r |g(s)|\, ds \le T + P. \tag{35}$$

Proof. Set $m = (l + r)/2$. Lipschitz continuity and the reverse triangle inequality give $\big||g(s)| - |g(l)|\big| \le L(s-l)$ on $[l, m]$ and $\big||g(s)| - |g(r)|\big| \le L(r-s)$ on $[m, r]$. Integrating each half-interval and using the triangle inequality gives

$$\left| \int_l^r |g(s)|\, ds - T \right| \le \frac{L}{2}(m-l)^2 + \frac{L}{2}(r-m)^2 = \frac{Lh^2}{4}. \tag{36}$$

The constant endpoint contributions sum to T. The absolute-value bound and nonnegativity of the integral give the two-sided enclosure. The proof requires no derivative of $|g|$. □

Proof of Proposition A.6. Apply Lemma A.7 on each cell, then sum the inequalities. Strictly increasing endpoints and additivity of adjacent integrals give

$$C(1) = \sum_{\ell=0}^{N-1} \int_{t_\ell}^{t_{\ell+1}} |g(s)|\, ds. \tag{37}$$

This proves Eq. (34). The residual bound $|e(a)| \le C(1) \le \overline{C}$ feeds directly into the location and regret arguments of Theorem 3.6.

For the third-derivative bound, suppose ϕ is C^3 on an open neighbourhood of $[0, 1]$ and $|\phi'''(s)| \le L_\ell$ at every point of each cell. Since $g' = \phi'''$, the mean value inequality makes g L_ℓ-Lipschitz there. Every cell is contained in $[0, 1]$, so these bounds supply exactly the hypotheses just used. □

The finite sum specifies an upper bound when its cellwise constants are valid throughout their intervals. A maximum over finitely many sampled third derivatives does not supply this condition. These bounds hold in exact arithmetic. Floating-point derivative evaluation and outward rounding require separate numerical error control.

A.5.5 Zero Deviation and Nonnegative Initial Curvature

Corollary A.8 (Zero Deviation and Multiple Maximisers)**.** *Under the C^2 setup, $C(1) = 0$ implies $\phi(a) = q(a) + \phi(0)$ on $[0, 1]$. If additionally $Q < 0$, every true maximiser then equals $\widehat{\alpha}$. For any value of $C(1)$, under $Q < 0$ and $\kappa = -2Q > 0$, any two global maximisers $a, b \in [0, 1]$ satisfy $|a - b| \le 2C(1)/\kappa$.*

Proof. If $C(1) = 0$, the residual derivative satisfies $|e(a)| \le 0$. The mean value inequality for $\phi - q$, with derivative zero, makes the value residual constant, proving the path identity. The zero-distance conclusion follows from Theorem 3.6. Applying its distance bound to each maximiser and using the triangle inequality through $\widehat{\alpha}$ gives the diameter bound. □

Proposition A.9 (Location Instability for $Q \ge 0$)**.** *For every $Q \ge 0$ and $\eta > 0$, the path $\phi(a) = Qa(a-1) + \eta a^3$ has Taylor quadratic $q(a) = Qa(a-1)$. Its smallest quadratic maximiser is zero and its unique true maximiser is one, while $|\phi'(a) - q'(a)| \le 3\eta$ on $[0, 1]$.*

Proof. Differentiation gives $\phi'(0) = -Q$ and $\phi''(0) = 2Q$, identifying the Taylor model. Since $a^2 \le a$ on $[0, 1]$, $q(a) \le 0 = q(0) = q(1)$, so its smallest maximiser is zero. Moreover $\phi(a) \le \eta a^3 \le \eta a \le \eta = \phi(1)$, and equality at a maximum forces $a = 1$ because $\eta > 0$. Finally $\phi'(a) - q'(a) = 3\eta a^2 \in [0, 3\eta]$. □

Thus an arbitrarily small derivative residual does not imply a small location error when the initial quadratic curvature is nonnegative.

Proposition A.10 (Value Bound for an Arbitrary Quadratic)**.** *For any A, Q, suppose ϕ is differentiable on a neighbourhood of $[0, 1]$ and $|\phi'(a) - (A + 2Qa)| \le \varepsilon$ for every $a \in [0, 1]$, with $\varepsilon \ge 0$. If $\alpha^\star$ maximises ϕ and $\widehat{\alpha}$ maximises $q(a) = Aa + Qa^2$, both on $[0, 1]$, then $0 \le \phi(\alpha^\star) - \phi(\widehat{\alpha}) \le \varepsilon$.*

Proof. The residual increment is at most $\varepsilon|\alpha^\star - \widehat{\alpha}|$ by the mean value inequality, and the quadratic gap is nonpositive by maximality. Their sum is bounded by ε because both scales lie in $[0, 1]$. Maximality of $\alpha^\star$ gives nonnegativity. □

Corollary A.11 (Combined Integrated and Taylor Regret Bounds)**.** *Under Theorem 3.6 and Assumption A.13,*

$$\begin{aligned}\phi(\alpha^\star) - \phi(\widehat{\alpha}) &\le \min\left\{\frac{M}{6}\big((\alpha^\star)^3 + \widehat{\alpha}^3\big), \frac{C(1)^2}{2\kappa}\right\} \\ &\le \min\left\{\frac{M}{3}, \frac{M^2}{8\kappa}\right\}.\end{aligned} \tag{38}$$

Proof. Apply the Taylor result in Theorem A.14 to the centred path $\phi - \phi(0)$. It has the same derivatives and maximisers, and centring cancels from its regret. Combining that bound with Theorem 3.6 gives the first minimum. Corollary A.12 gives $C(1) \le M/2$, and the two scales are in $[0, 1]$, giving the second minimum. □

Corollary A.12 (Third-Derivative Specialisation)**.** *Under Proposition A.4, suppose additionally that $\phi \in C^3(\mathcal{O})$ and $|\phi'''(s)| \le M$ on $[0, 1]$ for $M \ge 0$. Then, for every $a \in [0, 1]$,*

$$C(a) \le \frac{Ma^2}{2}, \qquad |\widehat{\alpha} - \alpha^\star| \le \frac{M(\alpha^\star)^2}{2\kappa} \le \frac{M}{2\kappa}. \tag{39}$$

Proof. A further application of the fundamental theorem of calculus gives

$$|\phi''(s) - \phi''(0)| = \left|\int_0^s \phi'''(u)\, du\right| \le \int_0^s |\phi'''(u)|\, du \le Ms. \tag{40}$$

Integrating this inequality from zero to a yields $C(a) \le \int_0^a Ms\, ds = Ma^2/2$. Proposition A.4, $\kappa > 0$, and $0 \le \alpha^\star \le 1$ give the asserted distance bound. □

A.5.6 Utility Regret and the Third-Derivative Bound

Assumption A.13 (Directional Curvature Variation)**.** Let $\mathcal{O} \subseteq \mathbb{R}$ be open with $[0, 1] \subseteq \mathcal{O}$. The utility gain along the path satisfies $\phi \in C^3(\mathcal{O})$, and there is a constant $M \ge 0$ such that

$$|\phi'''(\alpha)| \le M, \qquad \alpha \in [0, 1]. \tag{41}$$

Theorem A.14 (Quadratic Step-Scale Approximation)**.** *Under Assumption A.13, suppose $\phi(0) = 0$. With the quantities in Eqs. (3) and (8),*

$$0 \le \phi(\alpha^\star) - \phi(\widehat{\alpha}) \le \frac{M}{6}\big((\alpha^\star)^3 + \widehat{\alpha}^3\big) \le \frac{M}{3}. \tag{42}$$

If additionally $Q < 0$, let $\kappa := -2Q > 0$. Then

$$|\widehat{\alpha} - \alpha^\star| \le \frac{M}{2\kappa}(\alpha^\star)^2 \le \frac{M}{2\kappa}. \tag{43}$$

Proof of Theorem A.14. Under Assumption A.13, apply Lemma A.3 with $n = 0$ to ϕ'', whose derivative is bounded by M. This gives $|\phi''(\alpha) - \phi''(0)| \le M\alpha$. The function $f_1(\alpha) = \phi'(\alpha) - \phi''(0)\alpha$ has derivative $\phi''(\alpha) - \phi''(0)$. Applying the same lemma with $n = 1$ yields

$$|\phi'(\alpha) - (A + \phi''(0)\alpha)| \le \frac{M}{2}\alpha^2. \tag{44}$$

Next take $f_2(\alpha) = \phi(\alpha) - q(\alpha)$. Its derivative is the preceding error, and $f_2(0) = 0$. The lemma with $n = 2$ and coefficient $M/2$ gives the function-value bound. Using $\phi''(0) = 2Q$, the two bounds are

$$|\phi(\alpha) - q(\alpha)| \le \frac{M}{6}\alpha^3, \qquad |\phi'(\alpha) - q'(\alpha)| \le \frac{M}{2}\alpha^2. \tag{45}$$

Optimality of $\alpha^\star$ gives the nonnegative lower bound in Eq. (42). Optimality of $\widehat{\alpha}$ for q gives

$$\begin{aligned}\phi(\alpha^\star) - \phi(\widehat{\alpha}) &= \big[\phi(\alpha^\star) - q(\alpha^\star)\big] + \big[q(\alpha^\star) - q(\widehat{\alpha})\big] + \big[q(\widehat{\alpha}) - \phi(\widehat{\alpha})\big] \\ &\le \frac{M}{6}\big((\alpha^\star)^3 + \widehat{\alpha}^3\big) \le \frac{M}{3}.\end{aligned} \tag{46}$$

For the location bound, $Q < 0$ gives $\kappa = -2Q = -\phi''(0) > 0$. Proposition A.4 and Corollary A.12 give directly

$$|\widehat{\alpha} - \alpha^\star| \le \frac{C(\alpha^\star)}{\kappa} \le \frac{M(\alpha^\star)^2}{2\kappa} \le \frac{M}{2\kappa}. \tag{47}$$

This includes boundary maxima and requires no concavity of ϕ. □

B FINITE ACTION SETS

Let $\mathcal{A} = \{a_0, \dots, a_m\}$, with $m \geq 1$, be ordered with $0 = a_0 < a_1 < \cdots < a_m = 1$, and let $h_{\mathcal{A}} := \max_{\ell=0,\dots,m-1}(a_{\ell+1} - a_\ell)$. The sampled optimum $\alpha^\star_{\mathcal{A}}$ is defined in Eq. (13).

Proposition B.1 (Grid Resolution)**.** *If ϕ is continuous and concave on $[0,1]$ with a unique maximiser $\alpha^\star$, then*

$$|\alpha^\star_{\mathcal{A}} - \alpha^\star| \leq h_{\mathcal{A}}. \tag{48}$$

Separately, if $\phi \in C^2(\mathcal{O})$ for an open neighbourhood $\mathcal{O}$ of $[0,1]$, $\alpha^\star \in (0,1)$ is a maximiser, and $|\phi''(\alpha)| \leq B_\phi$ on $[0,1]$ with $B_\phi \geq 0$, then

$$0 \leq \phi(\alpha^\star) - \phi(\alpha^\star_{\mathcal{A}}) \leq \frac{B_\phi}{8} h^2_{\mathcal{A}}. \tag{49}$$

Proof. For the location bound, take grid points $a_{\mathrm{L}}, a_{\mathrm{R}}$ bracketing $\alpha^\star$, and write $g = \alpha^\star_{\mathcal{A}}$. Uniqueness gives $\phi(x) < \phi(\alpha^\star)$ whenever $x \neq \alpha^\star$. Suppose $g < a_{\mathrm{L}}$. If $a_{\mathrm{L}} = \alpha^\star$, this contradicts grid optimality. If $a_{\mathrm{L}} < \alpha^\star$, concavity and the strict value gap imply $\phi(g) < \phi(a_{\mathrm{L}})$, again contradicting grid optimality. The symmetric argument excludes $g > a_{\mathrm{R}}$. Thus $a_{\mathrm{L}} \leq g \leq a_{\mathrm{R}}$, and $|g - \alpha^\star| \leq a_{\mathrm{R}} - a_{\mathrm{L}} \leq h_{\mathcal{A}}$. When $\alpha^\star$ is a grid point, take $a_{\mathrm{L}} = a_{\mathrm{R}} = \alpha^\star$.

For the value bound, interior constrained optimality is local optimality on the real line, so $\phi'(\alpha^\star) = 0$. Choose a nearest grid point b with $|b - \alpha^\star| \leq h_{\mathcal{A}}/2$ and reparametrise the segment between these points:

$$\psi(u) = \phi(\alpha^\star + u(b - \alpha^\star)) - \phi(\alpha^\star), \qquad 0 \leq u \leq 1. \tag{50}$$

Its derivatives are $\psi'(u) = \phi'(\alpha^\star + u(b-\alpha^\star))(b-\alpha^\star)$ and $\psi''(u) = \phi''(\alpha^\star + u(b-\alpha^\star))(b-\alpha^\star)^2$. The segment lies in $[0,1]$. Hence $\psi(0) = \psi'(0) = 0$ and $\psi''(u) \geq -B_\phi(b-\alpha^\star)^2$. The two increment comparisons used above give the lower-bound formula also for zero initial derivative. Applying it to ψ at one yields $\phi(b) - \phi(\alpha^\star) \geq -B_\phi(b-\alpha^\star)^2/2$. Since $\phi(g) \geq \phi(b)$, the upper bound follows from the nearest-point distance. Continuous optimality gives the lower bound. □

If a global maximiser lies at either endpoint, it belongs to $\mathcal{A}$, so the grid attains the same optimal value and the value regret is zero. Thus the value bound also holds for endpoint maximisers.

Sampled crossings. Suppose consecutive grid points $\alpha_+ < \alpha_-$ satisfy $\phi(\alpha_+) \geq 0$ and $\phi(\alpha_-) < 0$. Continuity guarantees at least one zero in $[\alpha_+, \alpha_-)$. Identifying that zero as the first positive zero requires additional control of the path between samples. Nonnegative values at every grid point establish nonnegativity on the sampled set.

Proof. The intermediate value theorem gives a zero in $[\alpha_+, \alpha_-]$. The strict negative value at α_- excludes that endpoint. □

C AFFINE-LOGIT UTILITIES

This section specialises the path analysis to an average of log-softmax utilities. Fix an example i, a finite nonempty reference-token index set M_i, and a finite nonempty vocabulary $\mathcal{V}$. For each $j \in M_i$, let $z_{0,j}$ and v_j be real vectors indexed by $\mathcal{V}$, representing the initial logits and their displacement, respectively. Let $z_j(\alpha) = z_{0,j} + \alpha v_j$ be the resulting affine logit path, $p_j(\alpha) := \mathrm{softmax}(z_j(\alpha))$ its distribution, and $y_{i,j} \in \mathcal{V}$ its reference label. The component $p_{j,a}(\alpha)$ is the probability of vocabulary entry a. Suppose the utility along the segment satisfies

$$U(\gamma(\alpha)) = \frac{1}{|M_i|} \sum_{j \in M_i} \log p_{j,y_{i,j}}(\alpha). \tag{51}$$

All softmax probabilities are strictly positive. For two such distributions, use the finite KL divergence $D_{\mathrm{KL}}(p\|\widetilde{p}) := \sum_{a \in \mathcal{V}} p_a(\log p_a - \log \widetilde{p}_a)$.

C.1 Directional Curvature

Write the probability-weighted variance of a logit displacement as

$$V(p, v) := \sum_{a \in \mathcal{V}} p_a v_a^2 - \left(\sum_{a \in \mathcal{V}} p_a v_a \right)^2 . \tag{52}$$

For the basis vector $e_{y_{i,j}}$ of the reference label, the derivatives of each token utility are

$$\begin{aligned} \frac{d}{d\alpha} \log p_{j,y_{i,j}}(\alpha) &= (e_{y_{i,j}} - p_j(\alpha))^\top v_j, \\ \frac{d^2}{d\alpha^2} \log p_{j,y_{i,j}}(\alpha) &= -V(p_j(\alpha), v_j). \end{aligned} \tag{53}$$

The variance is nonnegative and is strictly positive when the displacement is nonconstant and all probabilities are positive.

Proof. The softmax derivative satisfies $p'_{j,a}(\alpha) = p_{j,a}(\alpha)(v_{j,a} - p_j(\alpha)^\top v_j)$. Positivity allows differentiation of its logarithm by division by $p_{j,a}(\alpha)$. At the reference label, this gives the first formula in Eq. (53). Differentiate the resulting directional term:

$$\begin{aligned} \frac{d}{d\alpha} \left(v_{j,y_{i,j}} - \sum_a p_{j,a}(\alpha) v_{j,a} \right) &= - \sum_a p'_{j,a}(\alpha) v_{j,a} \\ &= - \sum_a p_{j,a}(\alpha) v_{j,a}^2 + \left(\sum_a p_{j,a}(\alpha) v_{j,a} \right)^2 . \end{aligned} \tag{54}$$

This is the negative variance in the second formula. For fixed p, v, put $\mu = \sum_a p_a v_a$. Expansion of the centred sum, using $\sum_a p_a = 1$, gives $V(p, v) = \sum_a p_a (v_a - \mu)^2 \geq 0$. If the variance were zero, every nonnegative summand would vanish. With all $p_a > 0$, this forces $v_a = \mu$ for every a. A nonconstant displacement therefore has strictly positive variance. □

C.2 Endpoint Decomposition and Shape

Proof of Proposition 3.4. For a logit vector z, write $\Lambda(z) = \log \sum_{a \in \mathcal{V}} \exp z_a$. The log-softmax identity is $\log p_a = z_a - \Lambda(z)$. For arbitrary endpoint logits z, w and their distributions $p = \mathrm{softmax}(z)$ and $\widetilde{p} = \mathrm{softmax}(w)$, substitution into the finite KL sum and use of $\sum_a p_a = 1$ give

$$\begin{aligned} D_{\mathrm{KL}}(p \| \widetilde{p}) &= \Lambda(w) - \Lambda(z) - \sum_a p_a (w_a - z_a), \\ \log \widetilde{p}_a - \log p_a &= (e_a - p)^\top (w - z) - D_{\mathrm{KL}}(p \| \widetilde{p}). \end{aligned} \tag{55}$$

To establish the sign, apply $\log x \leq x - 1$ to $x = \widetilde{p}_a / p_a > 0$, multiply by p_a, and rearrange. This gives $p_a - \widetilde{p}_a \leq p_a (\log p_a - \log \widetilde{p}_a)$. Summing yields

$$0 = \sum_a (p_a - \widetilde{p}_a) \leq D_{\mathrm{KL}}(p \| \widetilde{p}). \tag{56}$$

Apply the endpoint identity to $z = z_{0,j}$, $w = z_{0,j} + v_j$, and reference label $y_{i,j}$, then average over M_i. Finite-sum differentiation in Eq. (53) identifies the average directional term with $A = \phi'(0)$. The result is $\Delta U = A - \mathcal{C}$, with $\mathcal{C} \geq 0$. The condition $A > 0, \Delta U < 0$ is consequently equivalent to $0 < A < \mathcal{C}$. Finally, the same finite-sum derivative chain has second derivative $-|M_i|^{-1} \sum_j V(p_j(\alpha), v_j)$. Applying Eq. (4) to this chain at one and comparing with the endpoint identity gives the integral expression for $\mathcal{C}$. □

Corollary C.1 (Shape of a Finite-Step Failure)**.** *Under Eq. (51), suppose at least one v_j is nonconstant across the vocabulary. Then ϕ is strictly concave on $[0, 1]$. If additionally $A > 0$ and $\phi(1) < 0$, it has a unique maximiser on $[0, 1]$, lying in $(0, 1)$, and a unique positive zero in $(0, 1)$.*

Proof. Equation (53) expresses ϕ'' as the negative average of nonnegative variances. The nonconstant displacement makes at least one variance strictly positive at every scale. Thus $\phi'' < 0$ and ϕ is strictly concave. Since $A > 0$, the local positivity argument gives $\phi(\alpha) > 0$ for $0 < \alpha < \varepsilon$. Choose $\alpha_0 = \min\{\varepsilon/2, 1/2\}$. The smallest maximiser exists by Appendix A.4, and its value is at least $\phi(\alpha_0) > 0$. It is therefore different from both endpoints, whose values are zero and negative. Strict concavity gives uniqueness of the maximiser.

Continuity on $[\alpha_0, 1]$ gives a zero r in that interval. Since $\alpha_0 > 0$ and $\phi(1) < 0$, it lies in $(0, 1)$. If $0 < r < s \le 1$ were two zeros, then $r = (1 - r/s)0 + (r/s)s$ lies strictly between zero and s. Strict concavity would give $\phi(r) > 0$, contradicting its being a zero. Interchanging r, s excludes the reverse ordering, proving uniqueness. □

D Looped Language Model Instantiation

D.1 States and Readout Boundaries

For a model with frozen parameters θ and example i, fix the input representation E_i, tokenisation, initialisation, and execution policy π, including batch composition and padding. An observed recurrent transition is

$$H_{i,t+1} = \widehat{F}_{\theta,\pi}(H_{i,t}; E_i), \qquad D_{i,t} = H_{i,t+1} - H_{i,t}. \tag{57}$$

Here $\widehat{F}_{\theta,\pi}$ is the implemented recurrent transition under execution policy π, and t counts recurrent iterations. Fixing i, t, θ, π, E_i specifies the update $H^+ = F(H)$ in § 3.1. Let n_i be the sequence length and d the hidden dimension. The single-sequence state space is $\mathbb{R}^{n_i \times d}$ with its Frobenius inner product. A joint option state belongs to the product of the option-specific sequence spaces, with inner product equal to the sum of the component Frobenius inner products.

For Ouro, the state boundary lies after the shared stack and its output normalisation, immediately before the linear language-model head. For Huginn it lies after the recurrent core and before the fixed nonlinear coda and output head. For fixed attention masks, these readouts combine affine maps, softmax attention, SiLU activations, and RMSNorm with a positive stabilising constant. These operations and the log-probability utilities below are smooth in real-valued arithmetic, giving the regularity required by Assumption 3.1. The recurrent endpoints are produced in BF16. The fixed readout is evaluated in FP32, and the final log-softmax over joint option scores is evaluated in FP64. Derivatives hold the endpoints and their displacement fixed and pass only through the readout. The mathematical statements apply to the real-valued readout continuation under their stated smoothness conditions. The numerical calculations evaluate this analysis at the computed recurrent states.

D.2 Task Utilities

Reference solutions. Let $\mathcal{V}$ be the model's finite nonempty vocabulary. For a mathematical problem x_i and its nonempty reference solution y_i, let $M_i \subseteq \{1, \dots, |y_i|\}$ be the nonempty set of evaluated solution-token indices. Define

$$U_i(H) := \frac{1}{|M_i|} \sum_{j \in M_i} \log p_\theta(y_{i,j} \mid x_i, y_{i,<j}; H). \tag{58}$$

Here p_θ is the model's conditional token distribution, $y_{i,j}$ is reference token j, and $y_{i,<j}$ is its preceding reference prefix. The length $|y_i|$ counts reference tokens. Each token is read at its preceding prediction position, and the complete reference solution is retained. With i fixed, this is the utility U used in the path analysis. Ouro's linear head supplies the affine logits in Appendix C.

Joint options. For a problem x with $K \ge 1$ nonempty option sequences $y^{(k)}$, write their teacher-forced state tuple as $\mathbf{H} = (H^{(1)}, \dots, H^{(K)})$ and let $k^\star \in \{1, \dots, K\}$ be the correct option. Suppressing the example index, define

$$\begin{aligned} s_k(H^{(k)}) &:= \frac{1}{|y^{(k)}|} \sum_{j=1}^{|y^{(k)}|} \log p_\theta(y_j^{(k)} \mid x, y_{<j}^{(k)}; H^{(k)}), \\ U(\mathbf{H}) &:= s_{k^\star}(H^{(k^\star)}) - \log \sum_{k=1}^{K} \exp s_k(H^{(k)}). \end{aligned} \tag{59}$$

Table 3: Depth-sweep coverage and observed utility declines. Peak is the evaluated depth with highest mean reference utility. Decline is the earliest running-best decrease with a pointwise 95% paired interval below zero; a dash indicates none on the evaluated grid. Non-adjacent depth pairs represent cumulative changes. Val. denotes validation.

Task	Split	N	Ouro-1.4B		Ouro-2.6B		Huginn	
			Peak	Decline	Peak	Decline	Peak	Decline
MATH-500	Test	500	4	$4 \to 5$	4	$4 \to 5$	24	$24 \to 48$
GSM8K	Test	1319	4	$4 \to 5$	4	$4 \to 5$	16	$16 \to 19$
HellaSwag	Val.	10042	2	$2 \to 3$	6	$6 \to 24$	32	$32 \to 34$
ARC-C	Test	1172	4	$4 \to 5$	4	$4 \to 5$	19	—
PIQA	Val.	1838	4	$4 \to 12$	7	$7 \to 24$	64	—
BoolQ	Val.	3270	1	$1 \to 3$	4	$4 \to 5$	24	$24 \to 32$
CommonsenseQA	Val.	1221	4	$4 \to 5$	1	$1 \to 2$	20	$20 \to 24$

The score s_k is the mean log probability of option k, and $y_{<j}^{(k)}$ is that option's prefix preceding token j. These scores are evaluated under teacher forcing: the reference prefix is provided when the next token is scored. HellaSwag uses $K = 4$. Differentiation and intervention act on the whole tuple, including the cross-option dependence introduced by log-softmax. The affine-logit special case concerns the reference-solution utility. Huginn's nonlinear readout and the joint option utility use the general pathwise derivatives.

Both utilities measure support for the reference under teacher forcing. Generation accuracy and utility are separate outcomes. Comparisons hold the checkpoint, conditioning, state boundary, and utility fixed, and report magnitudes within each task setting.

D.3 Scales and Comparisons

The step-scale prediction is evaluated on $\mathcal{A} = \{0, 0.05, \ldots, 1\}$, using the original A, Q without refitting them to the measured curve. The primary subset has $A > 0, Q < 0$. The error to the grid optimum is distinct from the continuous error in Theorem 3.6. Appendix B states the conditions under which grid resolution controls that difference. The measured prediction accuracy does not estimate $C(1)$ or the third-derivative bound M. Those bounds are evaluated separately (Appendix E.9).

The intervention set is $\mathcal{B} = \{0, 0.25, 0.5, 0.75, 1\}$, with prespecified step contraction $\alpha_c = 0.25$. Every option of a joint state receives the same scale. Recovery frequency and mean gain are evaluated on the finite-step failures selected by Definition 3.2. All interventions use reference utility after the original displacement has been computed, with the effect measured before further recurrence.

E Additional Experimental Details

E.1 Coverage and Depth Selection

The complete utility sweep evaluates the seven labelled splits in Table 3 for each of the three models. There are 19362 questions per model, giving 58086 model–question observations. Questions are shared across models. Both Ouro models are evaluated at depths one through eight, then 12, 16, 24, and 32. Huginn is evaluated at depths 4, 8, 16, 19, 20, 24, 32, 33, 34, 48, 64, 96, and 128. Every specified depth is retained, including depths after a measured loss. At each depth, the running-best comparison uses the earlier evaluated depth with the largest mean utility. The reported interval is the paired interval for the resulting per-question difference. Depths and datasets have been inspected during the study, and these pointwise intervals have no adjustment for selection or multiple comparisons.

The magnitude and interpretation of a decline depend on the task. Huginn's first confirmed running-best loss on HellaSwag is a cumulative $32 \to 34$ change of approximately -8.45×10^{-5}. For Ouro-1.4B on BoolQ, the model predicts Yes on 3268 of 3270 questions at depths three and four, and on all questions at depth twelve. This response imbalance accompanies the utility curve and limits its interpretation as a deterioration of reasoning. The original Ouro-1.4B/HellaSwag validation set supplies a separate comparison between utility and discrete correctness. Across its 1000 questions,

the $4 \to 5$ mean utility change is -0.008006, with interval $[-0.010838, -0.004991]$, while option accuracy rises from 68.4% to 69.2%. The smooth reference utility and the selected answer can therefore move differently under the same recurrent update. The main mechanism comparison instead holds the transition fixed across tasks within each model and includes all nine conditions, with 35583 model–question observations in total. It reuses the states and execution settings of the complete utility sweep.

E.2 Evaluation Sets and Numerical Protocol

Transition and sample provenance. The original Ouro/MATH-500 study used an independent 128-question training pilot to identify the depth range around the $4 \to 5$ transition, then evaluated the full 500-question test set. The original Huginn/GSM8K evaluation set consists of 256 training questions: all 256 source identifiers match the training split and none match the test split. The $19 \to 20$ transition was the earliest consecutive decrease with a negative pointwise mean-gain interval within the examined 16–24 range on those same questions. The original Ouro-1.4B/HellaSwag study fixed $4 \to 5$ before measurement and evaluated 1000 validation questions. The CommonsenseQA extension evaluates 1024 training questions per model. Ouro retains $4 \to 5$, and Huginn uses $33 \to 34$, selected by the earliest negative consecutive-change interval in a dense sweep of the same evaluation set. Earlier nested subsets of these questions do not provide independent replications.

The complete mechanism matrix applies $4 \to 5$ to both Ouro models and $19 \to 20$ to Huginn on each full split. A separate Huginn/GSM8K test evaluation also held $19 \to 20$ fixed and found 107 finite-step failures among 1319 questions. Its logged execution and initialisation settings differ from those of the complete matrix, which gives 102 such failures. Each result belongs to its own observed transitions, and the matrix uses only the latter run. The split distinction concerns evaluation provenance and does not establish absence from model pretraining.

Execution and differentiation. The complete matrix reconstructs each original batch with its recorded tokenisation, masks, candidate grouping, and natural padding width. For Huginn, the complete utility sweep initialises each question with a CPU BF16 truncated-normal state using seed 20260904 plus its source index, and candidate sequences share the question's seed and initialisation rule. The original 256-question GSM8K study instead continues cached depth-eight states with batch size 32. Those settings are preserved when analysing its step contractions and scale predictions.

All nine mechanism conditions compute A and Q by automatic differentiation and a Hessian–vector product through the actual readout utility, without constructing a full Hessian. TF32 is disabled. For HellaSwag, the product includes the cross-candidate terms of the joint option utility. Replayed endpoint utilities differ from the source values by at most 8.89×10^{-16}. On the first 16 questions of each condition, an additional scalar second-derivative calculation along the normalised full valid-token chord agrees with the Hessian–vector calculation to absolute error at most 10^{-6} or relative error at most 10^{-4}, and all 144 checks pass this disjunctive criterion. These checks compare two derivative calculations through the same numerical readout.

Intervals and effect margins. Question-level comparisons use 95% confidence intervals from 2000 paired question-bootstrap resamples within each analysis population. For the complete matrix, bootstrap resampling uses whole question records with seed 20260904. Each replicate recomputes the number of finite-step failures divided by the number of harmful updates, so the denominator varies with the resample. Table 4 reports these intervals together with the count satisfying both $A > 10^{-4}$ and $-\Delta U > 10^{-4}$. All nine conditions have zero neutral updates. Contraction intervals resample the selected finite-step failures, and primary scale-comparison intervals resample the $A > 0, Q < 0$ subset. These conditional intervals describe the selected population and do not adjust for the earlier choice of transition.

Task extensions and effect margins. Recovery at a shared scale varies across the examined paths. Nine of Huginn/GSM8K's 13 original recoveries gain at most 10^{-4}, while 19 of 20 HellaSwag recoveries exceed that margin. On the CommonsenseQA training subsets, the quarter step recovers 13 of 22 finite-step failures for Ouro-1.4B, 18 of 28 for Ouro-2.6B, and three of four for Huginn. These observations extend the recovery phenomenon to the additional model–task settings. Huginn's $33 \to 34$ transition was selected on its training subset; evaluation on the complete validation split leaves the sign of the mean gain unresolved. Figure 4 displays the original quarter-step estimates and their uncertainty, including all three CommonsenseQA evaluation sets.

Table 4: Finite-step failure shares and margin counts. Shares are relative to harmful updates. Margin counts require both $A > 10^{-4}$ and $-\Delta U > 10^{-4}$. Settings follow Table 1.

Model	Task	Failure share (%) with 95% CI	Failures beyond the margin
Ouro-1.4B	MATH-500	49.18 [44.29, 54.10]	177
	GSM8K	40.24 [37.35, 43.14]	421
	HellaSwag	4.52 [3.97, 5.04]	253
Ouro-2.6B	MATH-500	28.88 [24.10, 33.64]	90
	GSM8K	37.19 [33.79, 40.32]	321
	HellaSwag	1.86 [1.49, 2.27]	78
Huginn	MATH-500	22.17 [16.35, 27.89]	32
	GSM8K	14.53 [12.02, 17.09]	84
	HellaSwag	1.38 [1.05, 1.73]	24

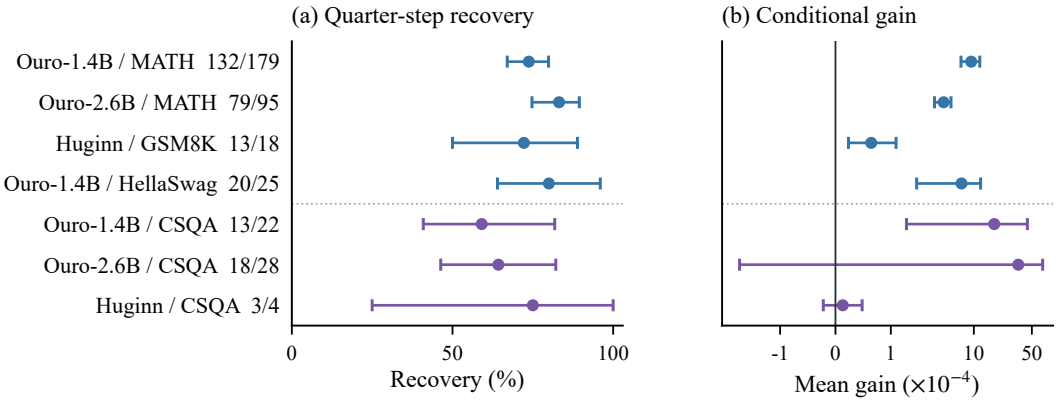

Figure 4: Quarter-step recovery. Each update uses one quarter of its proposed displacement along the same direction. (a) Fraction of finite-step failures with positive gains in reference utility, with recovered/total counts. (b) Mean gain over the same failure set, including unrecovered cases, on a symmetric-log scale. Purple rows denote CommonsenseQA.

Exact KL correction for the affine readout. For the original Ouro/MATH-500 evaluation sets, we evaluate the mean KL term $\mathcal{C}$ in Eq. (5) using the endpoint logits and a log-sum-exp calculation. This computes the correction independently of the measured difference $A - \Delta U$. Table 5 reports the maximum absolute residual of $\Delta U - (A - \mathcal{C})$ and the number of finite-step failures satisfying the exact decomposition's inequality numerically. The single exception on Ouro-2.6B has $\Delta U = -3.576 \times 10^{-7}$ and signed residual -6.044×10^{-7}. Its full-step classification is retained, but the strict inequality is unresolved at this numerical precision. This calculation applies to the affine reference-token readout. The nine-condition comparison uses automatic differentiation of each actual utility, including the nonlinear coda and joint option readout.

E.3 State and Output Diagnostics

The diagnostic comparison uses all valid token positions in the original four evaluation sets. State norms use the Frobenius norm, with all candidate states concatenated for multiple-choice questions. The signals are $\|D\|$, relative change $\|D\|/(\|H\| + 10^{-12})$, and cosine change $1 - \langle H, H^+\rangle/(\|H\|\|H^+\|)$. The norms in the cosine denominator are positive in these records. Output KL divergence uses the current token distribution as its first argument and the next-state distribution as its second, over the full vocabulary at prediction positions. Entropy change is next-state entropy minus current-state entropy at those positions. For HellaSwag, output diagnostics first average over prediction tokens within each candidate, then equally over the four candidates.

Table 6 reports the raw signal's Spearman correlation with ΔU and the harmful-update AUROC after a fixed risk orientation. Larger risk corresponds to the negative of each signal except entropy change,

Table 5: Numerical check of the exact endpoint correction. Both rows use the original MATH-500 test evaluation set ($N = 500$, $4 \to 5$). The residual is $|\Delta U - (A - \mathcal{C})|$, with $\mathcal{C}$ computed independently from endpoint logits. The last column counts finite-step failures satisfying $0 < A < \mathcal{C}$ numerically.

Model	Maximum residual	Inequality satisfied
Ouro-1.4B	2.661×10^{-6}	179/179
Ouro-2.6B	3.348×10^{-6}	94/95

Table 6: State, output, and task-geometry diagnostics. Columns use the original MATH-500 test evaluation sets ($N = 500$ each), GSM8K training subset ($N = 256$), and HellaSwag validation set ($N = 1000$). Raw correlations retain signal orientation, and AUROC uses the fixed risk signs of Appendix E.3.

Signal	Ouro-1.4B MATH-500	Ouro-2.6B MATH-500	Huginn GSM8K	Ouro-1.4B HellaSwag
Raw Spearman correlation with ΔU				
Update norm	-0.2843	0.0471	0.1006	-0.0090
Relative change	-0.2725	-0.0620	0.0010	-0.0457
Cosine change	-0.2724	-0.0640	0.0010	-0.0457
Output KL	-0.2357	-0.0985	-0.1490	-0.0057
Entropy change	-0.1573	-0.0344	-0.0110	0.0350
A	0.8276	0.9640	0.9951	0.9622
$A + Q$	**0.9929**	**0.9973**	**0.9999**	**0.9982**
Harmful-update AUROC				
Update norm	0.2893	0.4852	0.5122	0.4983
Relative change	0.3438	0.5241	0.4701	0.4960
Cosine change	0.3440	0.5229	0.4701	0.4961
Output KL	0.4390	0.5290	0.4279	0.5119
Entropy change	0.5826	0.5357	0.5105	0.4848
A	0.9411	0.9764	0.9974	0.9811
$A + Q$	**0.9974**	**0.9990**	**0.9998**	**0.9988**

whose orientation is positive. Orientations are retained when AUROC is below 0.5, and no affine mapping is fitted to the evaluation outcomes. The derivative scores use the reference utility, and the state and output diagnostics omit that task-specific derivative information while sharing the same teacher-forced states.

E.4 Scale Prediction and Curvature Measurements

The CommonsenseQA extensions have scale correlations of 0.873 and 0.837 for the two Ouro models and 0.928 for Huginn, with 268 of 275 Huginn grid optima at one. These estimates use exploratory training subsets, as specified in Appendix E.2.

The quadratic rule maximises q on $[0, 1]$ as specified in Eq. (9). It compares the endpoints and, when applicable, the interior concave maximum. Ties choose the smaller scale. Continuous scale errors are computed before mapping to the 21-point grid, and utility comparisons use the nearest grid point, with equal distances resolved toward the smaller point. Fixed-scale baselines use $0, 0.25, 0.5, 0.75, 1$. Figure 3(e) compares all five choices with the quadratic rule on the original primary evaluation sets. The paired quadratic-minus-first-order utility gain is 0.006697 with interval $[0.005622, 0.007794]$ for Ouro/MATH-500, and 9.646×10^{-5} with interval $[5.763 \times 10^{-5}, 1.427 \times 10^{-4}]$ for Huginn/GSM8K. Selecting the best fixed action on these same evaluation subsets gives a descriptive hindsight comparator.

The all-question scale correlations and MAEs are 0.9932 and 0.0133 on Ouro/MATH-500, and 0.9983 and 0.0036 on Huginn/GSM8K. On the original Huginn primary subset, 93 of 129 questions have both predicted and empirical grid optima at one. Restricting descriptively to its 35 interior grid optima gives correlation 0.9924 and MAE 0.0262. Twelve Ouro questions have $A > 0$ but a grid optimum at zero: a grid with spacing 0.05 need not resolve the beneficial interval near the origin.

Table 7: Local quadratic scale prediction. Primary subset $A > 0, Q < 0$. Correlation and scale MAE compare the continuous prediction with the 21-point grid optimum, and utility regret uses the nearest grid action. CommonsenseQA rows use exploratory training subsets.

Model / task	n	Spearman	Scale MAE	First-order regret	Quadratic regret
Ouro-1.4B / MATH-500 test, $N = 500$, $4 \to 5$	315	0.9951	0.0211	6.714×10^{-3}	$\mathbf{1.747 \times 10^{-5}}$
Huginn / GSM8K train, $N = 256$, $19 \to 20$	129	0.9998	0.0072	9.681×10^{-5}	$\mathbf{3.482 \times 10^{-7}}$
Ouro-1.4B / CommonsenseQA train, $N = 1024$, $4 \to 5$	302	0.8726	0.0146	2.044×10^{-3}	$\mathbf{1.919 \times 10^{-4}}$
Ouro-2.6B / CommonsenseQA train, $N = 1024$, $4 \to 5$	380	0.8369	0.0155	5.231×10^{-3}	$\mathbf{7.314 \times 10^{-4}}$
Huginn / CommonsenseQA train, $N = 1024$, $33 \to 34$	275	0.9275	0.0011	5.742×10^{-6}	$\mathbf{1.292 \times 10^{-7}}$

Table 8: Predicting the observed end of progress. Within $A > 0, Q < 0$, the quadratic root against the first observed nonnegative-to-negative crossing interval on the 21-point grid. Hits require the root to lie in that width-0.05 interval. The last column counts primary questions with no observed negative gain through scale one.

Model / task	Primary n	Crossings	Hits	Hit rate (%)	No negative gain observed
Ouro-1.4B / MATH-500 test, $N = 500$, $4 \to 5$	315	179	121	67.6	136
Huginn / GSM8K train, $N = 256$, $19 \to 20$	129	18	13	72.2	111

Predicting the progress boundary. Within each primary subset, we locate the first grid point with negative gain. Its immediately preceding grid point and this first negative point define the observed crossing interval. The numerical hit rule includes the lower endpoint and excludes the upper endpoint. Of 179 Ouro and 18 Huginn questions with an observed crossing, the predicted root $r^{(2)}$ lies in this interval for 121 and 13, respectively (Table 8). The remaining 136 Ouro and 111 Huginn primary questions have no observed negative gain through scale one, and their roots receive no numerical error by assigning the endpoint one as a target.

Continuity gives a zero in an observed crossing interval. In the affine reference-token case, concavity connects this crossing to the end of the nonnegative prefix. For a general nonlinear readout, negative excursions between earlier grid points are not excluded, so the interval need not locate the first continuous boundary r_1. Within $A > 0, Q < 0$, the condition $r^{(2)} < 1$ is algebraically equivalent to $A + Q < 0$. Its classification accuracy therefore reuses the endpoint prediction, while interval hits assess the additional prediction of a location along the segment.

Directional curvature along the path. The five-point derivative measurements retain the original unnormalised D at every scale. The interquartile ranges of $\phi''(0.5)/\phi''(0)$ are $[1.003, 1.133]$ for Ouro and $[0.954, 1.039]$ for Huginn on the primary subsets. The median maximum adjacent curvature difference, divided by $|\phi''(0)| + \varepsilon$, is 0.0951 and 0.0294, respectively. Here ε is 10^{-6} times the median of the original $|2Q|$ over the corresponding complete evaluation set, fixed before restricting to the primary subset. The adjacent difference is not divided by the scale spacing. Curvature ratios omit origins with $|\phi''(0)| \leq \varepsilon$. Negative curvature is observed at every sampled point up to the empirical optimum in all 315 Ouro and 129 Huginn primary cases. Of these cases, 84 and seven, respectively, include only the current state in that check. This sampled-prefix observation does not establish concavity throughout the segment.

E.5 Path Analysis and Curvature-Based Scale Selection

Fixed states and numerical readouts. The refined analysis uses all 500 original Ouro-1.4B/MATH-500 questions and all 256 original Huginn/GSM8K training questions at their respective $4 \to 5$ and $19 \to 20$ transitions. The recurrent computation and chosen endpoints are retained, and FP64

Table 9: Quadratic prediction against refined FP64 optimum estimates. Original evaluation sets of 500 Ouro-1.4B/MATH-500 test and 256 Huginn/GSM8K training questions, with primary-subset membership retained from the original analysis. Ouro entries are numerical intervals propagated from derivative-root brackets, and Huginn entries use the best evaluated candidate. These intervals are numerical, not confidence intervals. Interior and failure rows are descriptive subsets.

Model	Population	n	Mean scale error	Mean numerical regret
Ouro-1.4B	All	500	$[0.010221, 0.010254]$	$[1.150383, 1.150385] \times 10^{-5}$
Ouro-1.4B	Primary	315	$[0.016224, 0.016276]$	$[1.826005, 1.826008] \times 10^{-5}$
Ouro-1.4B	Interior	269	$[0.018928, 0.018989]$	$[2.134553, 2.134556] \times 10^{-5}$
Huginn	All	256	0.002447	1.0837×10^{-7}
Huginn	Primary	129	0.004855	2.1506×10^{-7}
Huginn	Interior	36	0.017399	7.7063×10^{-7}
Huginn	Finite-step failures	18	0.006354	1.5575×10^{-7}

readouts recompute A, Q, and utility along the numerical paths. Huginn uses its original padding and causal readout protocol. Replay checks remain below 6×10^{-6}, with primary membership and harmful-update labels unchanged.

Refined numerical optima. Ouro uses concavity to select an endpoint or bracket a zero of ϕ'. The 269 interior brackets have width at most 10^{-4}, with positive left derivative and nonpositive right derivative numerically. Propagating these brackets gives the scale-error and regret intervals in Table 9. The remaining optima are 185 left endpoints and 46 right endpoints. Huginn uses 257 adaptive derivative locations per question, with additional local root refinement and evaluation of $\widehat{\alpha}$. Its numerical optimum estimate is the smallest maximiser among the evaluated candidates: 126 estimates lie at zero, 94 at one, and 36 in the interior. The primary subset contains 93 right endpoints and all 36 interior estimates. Including $\widehat{\alpha}$ among the candidates ensures nonnegative numerical regret by construction. That regret is the best evaluated utility minus $\phi(\widehat{\alpha})$, and a missed maximum could increase the true continuous regret. Equal numerical maxima select the smaller scale, so flat paths can make the reported location sensitive to finite-precision ties. Both models evaluate the continuous prediction without grid snapping. Ouro's intervals propagate numerical root brackets, whereas Huginn's entries compare with the searched candidate. Neither incorporates outward-rounded arithmetic or provides a floating-point certificate.

Figure construction. Figure 5 displays every record in the matched mechanism matrix, with identical symmetric-log transforms on the two axes, linear within $[-10^{-3}, 10^{-3}]$. Figure 3(a–d) uses the separately recomputed FP64 paths. For each mathematical evaluation set, finite-step failures are ordered by their numerical optimum estimate, with record ID breaking ties, and the lower-median record supplies the example curve in panels (a,b). This selection does not use the quadratic approximation error. Lines connect the stored path evaluations. The quadratic is evaluated from the unchanged FP64 coefficients. Every primary prediction enters panels (c,d). Identical coordinates are combined with marker area increasing as the square root of their count. No jitter is applied. Ouro's numerical root intervals are drawn vertically, and their midpoints provide display coordinates. Panel (e) uses the original 21-point-grid regrets and conditional 95% bootstrap intervals on the original primary subsets. All displayed intervals are strictly positive, allowing a logarithmic axis; the zero-regret grid oracle is omitted. The quadratic action is displayed separately from the fixed-scale sequence. Its plotted regret is evaluated at the nearest grid action, matching Table 2. Figure 4 retains the original reported bootstrap intervals. The paired FP64 quarter-step and bound-selected gains are given in Table 10. The Python generator records the selected IDs and hashes of all input files.

Derivative envelopes. Ouro evaluates directional derivatives through softmax moments. The numerical envelope construction combines global moment bounds with a fourth-derivative bound on each of 256 contiguous intervals covering $[0, 1]$. It returns upper bounds for $|\phi'''|$ and $-\phi''$ by propagating endpoint derivative values across each interval and taking the maximum of the interval bounds. The resulting constants are evaluated using ordinary FP64 arithmetic. On the primary subset, the median ratio of the computed L_D to the largest sampled $-\phi''$ is 1.0146. This ratio compares the envelope with sampled extrema, not with the exact continuous supremum. The same cellwise construction supplies the constants L_ℓ of Proposition A.6. The resulting integrated and signed-residual scale bounds are evaluated on the complete mathematical benchmarks in Appendix E.9.

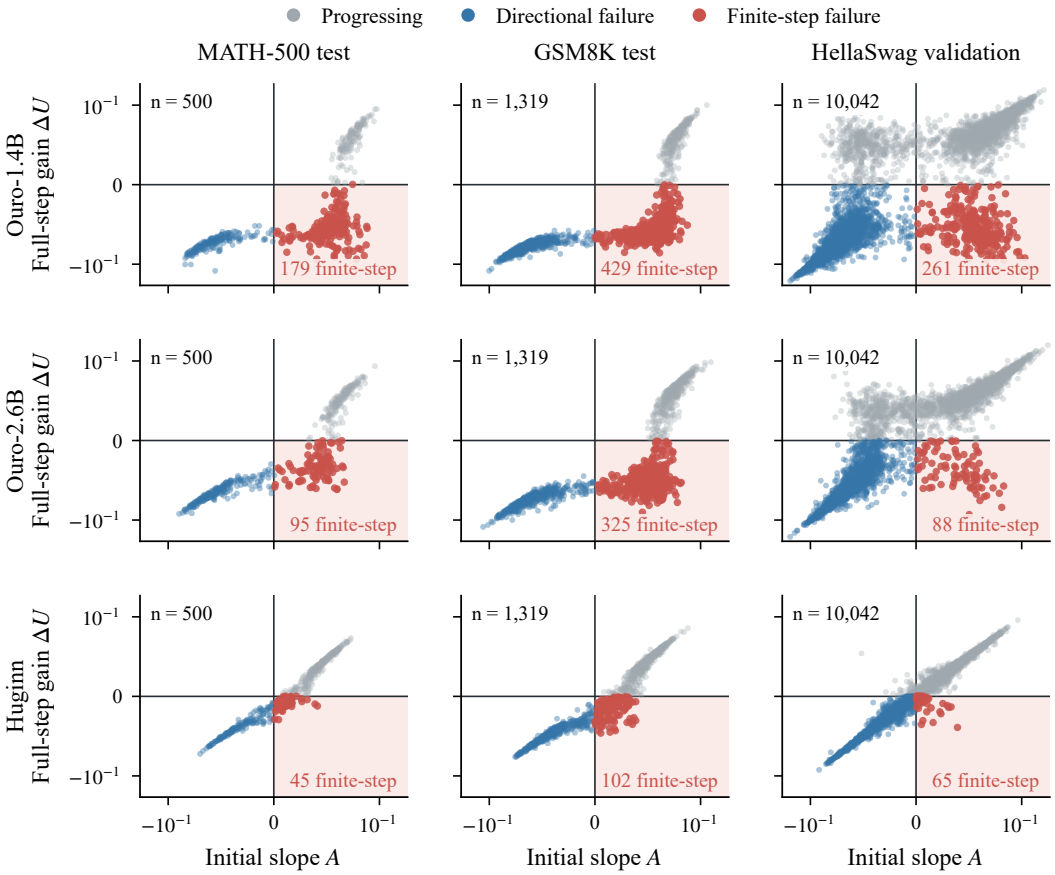

Figure 5: Local improvement and full-step harm. Each point compares the initial slope A with the full-step gain ΔU for a model–question pair in Table 1. Red points mark updates whose directions locally increase reference utility but whose full steps decrease it. Blue points mark harmful updates with nonpositive initial slopes, and grey points mark full updates that increase reference utility.

Huginn computes third derivatives by automatic differentiation. The median relative discrepancies with central differences of ϕ'' are 7.137×10^{-6}, 1.784×10^{-6}, and 4.460×10^{-7} at spacings 0.01, 0.005, and 0.0025. At the smallest spacing, the maximum absolute discrepancy is 1.069×10^{-8}. Increasing the adaptive point count from 129 to 257 changes the best sampled utility by at most 3.718×10^{-10} and its location by at most 0.000489. The sampled maxima of $|\phi'''|$ and $[-\phi'']_+$ change by at most 5.319×10^{-8} and 2.331×10^{-8}, respectively. All measured curvatures are negative on the 129 primary questions, and this holds for 254 of 256 questions overall. The primary median sampled derivative extrema are 1.789×10^{-4} for $|\phi'''|$ and 0.001847 for $[-\phi'']_+$. No continuous derivative upper bounds, global optimality certificate, or bound-selected scale are assigned to Huginn.

Scale selected from the sufficient range. For Ouro questions with $A > 0$, we use the envelope value as L_D in Corollary 3.5 and fix $\alpha_{\text{safe}} = \min\{1, 0.9(2A/L_D)\}$ before evaluating the intervention. The factor 0.9 is a fixed interior margin, unchanged across questions, rather than a fitted parameter. The reported gains and the quarter-step comparator both use the same FP64 readout. All 315 primary gains exceed the computed sufficient lower bound $\alpha_{\text{safe}}A - L_D\alpha_{\text{safe}}^2/2$, with minimum numerical margin 4.680×10^{-8}. Table 10 reports the primary and finite-step-failure populations separately, with 2000 paired question-bootstrap resamples and seed 20260904. On the 179 failures, the median selected scale is 0.32759, with interquartile range $[0.15404, 0.50776]$ and range $[0.005902, 0.821919]$. Of the 179 positive gains, 142 exceed 10^{-4} and 37 do not. This construction uses reference-dependent derivatives along the complete path and is evaluated before further recurrence.

Table 10: Selecting a scale from the negative-curvature bound. Ouro-1.4B/MATH-500 with $\alpha_{\text{safe}} = \min\{1, 0.9(2A/L_D)\}$. Both actions are evaluated in FP64 on the same records, and counts require positive utility gain. Gain columns are in units of 10^{-3} with paired 95% question-bootstrap intervals.

Population	n	Quarter-step positive gains	Bound-selected positive gains	Mean gain [95% CI]	Gain over quarter [95% CI]
Primary	315	268	**315**	5.1985 $[4.2777, 6.1934]$	**2.6278** $[1.9928, 3.3228]$
Finite-step failures	179	132	**179**	1.0738 $[0.8555, 1.3354]$	**0.1442** $[0.0610, 0.2317]$

Table 11: Quadratic selection recovers more finite-step failures than a quarter step. Ouro-1.4B/GSM8K test, $4 \to 5$, using the original updates in Table 1. Mean gains relative to H_4 use all 1319 questions, in units of 10^{-3}. Recovery counts use the 429 finite-step failures. The final row gives paired differences.

Action	Mean gain [95% CI]	Recovered / 429	Recovery (%)
Quarter step	-0.844 $[-1.093, -0.587]$	302 / 429	70.40
Quadratic rule	$\mathbf{+2.766}$ $[+2.436, +3.145]$	**425** / 429	**99.07**
Quadratic − quarter	$\mathbf{+3.610}$ $[+3.367, +3.885]$	$\mathbf{+123}$	$+28.67$ pp

E.6 Immediate Quadratic Scale Selection on the Complete GSM8K Test Set

This analysis evaluates the quadratic rule of § 3.3 on the same 1319 Ouro-1.4B/GSM8K test updates at $4 \to 5$ as Table 1. Both actions retain the original H_4, H_5 pair. The quarter step uses scale 0.25, while the quadratic rule selects its scale from the original A, Q before evaluating the intervention utility. Interpolation and immediate readout use FP32.

Table 11 reports gains relative to H_4 over all 1319 questions and recovery within the same 429 finite-step failures. Quadratic selection recovers 123 more failures than the quarter step and increases mean immediate reference utility across the full test set. In a separate precision check, casting interpolated states back to BF16 before readout gives recovery of 403/429 (93.94%) for quadratic selection and 305/429 (71.10%) for the quarter step.

E.7 Question-Level Interventions Across HellaSwag Transitions

This evaluation applies the quadratic rule to Ouro-2.6B on all 10042 HellaSwag validation questions at $1 \to 2$, $2 \to 3$, $3 \to 4$, and $4 \to 5$. Each question uses the joint option utility in Eq. (59), with one common scale across all four candidates. The derivatives include their coupling through the option-score log-softmax. Each transition uses its own pair of unmodified recurrent states and is evaluated immediately after scaling, before further recurrence. The first three transitions lie within the four-step inference configuration, and the last extends it by one step.

The recurrent endpoints use BF16. Candidate scores use the FP32 readout, followed by FP64 normalisation across options. The quadratic scale is computed from the original A, Q, maximising q on $[0, 1]$ with ties choosing the smaller scale. All questions enter every transition, giving 40168 question–transition records. The joint scalar second derivative and Hessian–vector calculation agree in all 64 numerical checks under the stated tolerance. The original depth-selection readout is reproduced bitwise in all 5021 batches, and the four-step prefix is unchanged when computing the fifth state.

Quadratic selection recovers more finite-step failures than the quarter step at each transition. It also gives higher observed mean utility gain over all questions than either the full or quarter step (Table 12). At $4 \to 5$, the rule recovers all 88 recorded finite-step failures, including 62 with gain above 10^{-4}. The first three transitions likewise contain recoveries above this margin, with 278, 170, and 211 cases,

Table 12: Quadratic scale selection across HellaSwag transitions. Ouro-2.6B, all 10042 validation questions, with one scale shared by the four candidates. (a) Recovery counts use the original finite-step failures. Mean gains use all questions, in units of 10^{-3}. (b) Full-step sign accuracy and paired quadratic-minus-full gain with 95% question-bootstrap intervals. Each transition is an independent intervention before further recurrence.

(a) Recovery and mean immediate gain

Transition	Failures	Recovered quarter → quadratic	Full	Quarter	Quadratic
$1 \to 2$	322	$231 \to \mathbf{297}$	59.317	17.370	**99.763**
$2 \to 3$	209	$162 \to \mathbf{184}$	2.337	-5.243	**37.436**
$3 \to 4$	248	$197 \to \mathbf{242}$	4.799	0.731	**18.956**
$4 \to 5$	88	$67 \to \mathbf{88}$	1.492	0.167	**9.296**

(b) Endpoint prediction and paired gain

Transition	Sign A (%)	Sign $A+Q$ (%)	Quadratic − full [95% CI]
$1 \to 2$	**85.17**	82.44	$\mathbf{40.446}$ $[38.531, 42.579]$
$2 \to 3$	84.43	**91.85**	$\mathbf{35.099}$ $[33.948, 36.268]$
$3 \to 4$	87.04	**95.94**	$\mathbf{14.157}$ $[13.655, 14.702]$
$4 \to 5$	94.24	**99.49**	$\mathbf{7.804}$ $[7.508, 8.113]$

respectively. These observations extend step-contraction recovery to a joint multiple-choice utility at several recurrent depths.

Endpoint prediction and intervention measure different outcomes. At $1 \to 2$, $A + Q$ has lower sign accuracy than A, although quadratic selection still recovers 297 of 322 finite-step failures. At $2 \to 3$, quadratic selection recovers more failures, while the mean gain over the original failure subset is 0.009511, compared with 0.009937 for the quarter step. Recovery counts and mean gains therefore retain their separate populations and interpretations. Reported paired intervals use 2000 question-bootstrap resamples with seed 20260904.

E.8 Token-Level Audit of Early Recurrent Updates

This audit evaluates Ouro-1.4B on all 1319 GSM8K test questions at $1 \to 2$, $2 \to 3$, and $3 \to 4$, using the evaluated checkpoint's original four-step depth-selection configuration. The CDF threshold is 1.0, with no fixed-exit override. Across these questions, the audit contains 167,606 reference-token prediction positions per transition. All of these positions continue through the three transitions and exit at depth 4. The gate's training-stage provenance is unresolved, and the results are attributed to this checkpoint and configuration.

The original policy selects readout depth from a dense teacher-forced trajectory. Reconstructed selected states match the original readout inputs bitwise in all 660 batches. Gate probabilities and the CDF use the original BF16 arithmetic; an FP64 reconstruction gives the same exit decisions. This audit therefore measures the geometry of continuing updates under the stored policy, without a population of early exits against which to assess gate discrimination.

At each reference prediction position, utility is the log probability of the reference next token. The coefficients A, Q and gain ΔU are computed for this scalar utility along that position's fixed displacement, using the FP32 affine readout. The quarter step uses scale 0.25, and the quadratic rule maximises its local model on $[0, 1]$. Each transition is evaluated independently from the original trajectory; the scaled states are read out immediately, with no further recurrence or cumulative intervention. Prompt positions contribute to policy coverage counts but not to the population used to evaluate reference utility.

Mean gains average over continuing prediction tokens. Failure rates use the same population, and recovery rates condition on the original finite-step failures in each transition. Confidence intervals use 2000 question-cluster bootstrap resamples with seed 20260904: all tokens of a resampled question are retained together, and each replicate recomputes the ratio of summed quantities to summed token counts. These token-weighted estimands differ from the question-weighted comparisons in the original evaluation sets.

Table 13: Finite-step failures in early recurrent updates. Ouro-1.4B/GSM8K test, 1319 questions and 167,606 continuing reference-token prediction positions per transition. (a) Failure rates use all continuing positions; recovery rates use the original failures. (b) Mean immediate gains and paired advantages are token-weighted reference log-probability increments, in units of 10^{-3}. Paired 95% intervals resample questions, retaining their tokens together. Original CDF threshold: 1.0.

(a) Failure counts and recovery

Transition	Finite-step failures (%)	Quarter recovered (%)	Quadratic recovered (%)
$1 \to 2$	15657 (9.34)	12479 (79.70)	**15517** (99.11)
$2 \to 3$	16621 (9.92)	12924 (77.76)	**16377** (98.53)
$3 \to 4$	14473 (8.64)	10879 (75.17)	**14150** (97.77)

(b) Utility over all continuing prediction positions

Transition	Full	Quarter	Quadratic	Quadratic − full [95% CI]	Quadratic − quarter [95% CI]
$1 \to 2$	637.812	305.453	**651.649**	**13.837** [12.467, 15.184]	**346.195** [339.764, 352.695]
$2 \to 3$	142.276	66.697	**183.909**	**41.633** [40.670, 42.605]	**117.212** [114.934, 119.489]
$3 \to 4$	18.271	12.028	**57.044**	**38.773** [37.941, 39.571]	**45.016** [44.100, 45.946]

Table 14: Scale bounds on the complete mathematical benchmarks. Median and, in brackets, 75th percentile of each bound on $A > 0, Q < 0$, before the minimum with one in Theorem 3.6. Coverage counts bounds covering the original numerical scale-error upper endpoints.

Model / task	n	$M/(2\kappa)$	$\overline{C}/\kappa$	$\overline{S}/\kappa$	$\overline{C}$ covers	$\overline{S}$ covers
Ouro-1.4B MATH-500	315	4.6108 [8.1823]	0.0967 [0.1744]	0.0852 [0.1578]	315/315	315/315
Ouro-1.4B GSM8K	683	3.6661 [4.6757]	0.0583 [0.0974]	0.0485 [0.0900]	683/683	680/683
Ouro-2.6B MATH-500	266	2.8436 [4.5339]	0.0602 [0.1031]	0.0520 [0.0949]	266/266	265/266
Ouro-2.6B GSM8K	770	2.8903 [4.0521]	0.1136 [0.1734]	0.1042 [0.1677]	770/770	769/770

All three transitions have positive mean full-step gain, while 8.64–9.92% of their continuing prediction positions exhibit finite-step failure. The quadratic rule recovers 97.77–99.11% of these failures and has positive paired mean advantages over both full and quarter steps on the entire continuing population (Table 13). A descriptive margin check requiring $A > 10^{-4}$ and $\Delta U < -10^{-4}$ retains 13938, 13002, and 9706 failures, respectively, within the original populations. These results locate recoverable utility losses in early recurrent updates with positive aggregate progress, providing a further test of the direction–step distinction within the four-step inference budget.

Future direction: geometry-guided adaptive computation. These findings motivate using reference-defined geometry as offline supervision for learned exit and step-scale decisions, complementing task-loss-based depth allocation in Ouro (Zhu et al., 2026). Learning such decisions from inference-time states and evaluating them on generated trajectories under matched computation budgets would connect this mechanism to task-aware adaptive computation.

E.9 Curvature-Bound Evaluation

The bound evaluation covers both Ouro models on the complete MATH-500 test and GSM8K test sets, with 3638 model–question records and 2034 primary records satisfying $A > 0, Q < 0$. All four conditions use $4 \to 5$ and FP64 readout analysis with 257 partition nodes. These are separate computed paths from the original evaluation sets. The Ouro-1.4B/MATH-500 primary count of 315 equals that of the original evaluation set. Predictions, derivatives, and numerical optimum estimates are computed within this run. Table 14 compares three scale bounds for the same predictions: the global-third-derivative bound $M/(2\kappa)$, the integrated bound $\overline{C}/\kappa$, and the signed-residual bound $\overline{S}/\kappa$ defined below.

The integrated median lies between 0.0583 and 0.1136 and the signed median between 0.0485 and 0.1042, compared with 2.84–4.61 for the global bound. The integrated and signed bounds also have

lower 75th percentiles in every condition. The integrated bound covers the numerical scale-error interval of all 2034 primary records. The signed bound provides a supplementary tighter envelope with direct coverage for 2029 records. These comparisons give tighter bounds for the same predictions and do not change the prediction errors. The four cell summaries match the aggregate summary exactly.

With cellwise third-derivative bounds L_ℓ on the partition $0 = t_0 < \cdots < t_N = 1$, the upper sum of Proposition A.6 is evaluated as

$$\overline{C} = \sum_{\ell=0}^{N-1} \left[\frac{h_\ell}{2} \big(|g(t_\ell)| + |g(t_{\ell+1})| \big) + \frac{L_\ell h_\ell^2}{4} \right], \qquad g(s) = \phi''(s) - \phi''(0), \tag{60}$$

and enters the scale and regret bounds of Theorem 3.6.

Signed-residual envelope. Let $e(a) := \phi'(a) - q'(a)$ on $[0, 1]$. Under the C^3 setup, $e'' = \phi'''$. Use the partition and valid nonnegative constants L_ℓ of Proposition A.6, now obtained from $|\phi'''| \le L_\ell$ throughout each cell. Define

$$\overline{S} := \min \left\{ \overline{C}, \max_{0 \le \ell < N} \left[\max\{|e(t_\ell)|, |e(t_{\ell+1})|\} + \frac{L_\ell h_\ell^2}{8} \right] \right\}. \tag{61}$$

Proposition E.1 (Signed-Residual Approximation Bounds)**.** *Under these hypotheses and $Q < 0$, $|e(a)| \le \overline{S}$ for every $a \in [0, 1]$. For every true global maximiser $\alpha^\star$,*

$$|\widehat{\alpha} - \alpha^\star| \le \min\{1, \overline{S}/\kappa\}, \qquad 0 \le \phi(\alpha^\star) - \phi(\widehat{\alpha}) \le \frac{\overline{S}^2}{2\kappa}. \tag{62}$$

Proof. On a cell $[l, r]$ with $|e''| \le L$, let I_e be the linear interpolant of the endpoint values of e. The functions $e(a) + La^2/2$ and $-e(a) + La^2/2$ are convex. Bounding each by its endpoint interpolant and subtracting the quadratic gives

$$|e(a) - I_e(a)| \le \frac{L}{2}(a - l)(r - a) \le \frac{L(r - l)^2}{8}. \tag{63}$$

Since $|I_e(a)| \le \max\{|e(l)|, |e(r)|\}$, the cell expression bounds $|e|$. Every point is covered by a partition cell, so taking the maximum gives a uniform residual bound. Independently, $|e(a)| \le C(1) \le \overline{C}$, so their minimum remains valid. The variational location bound and the residual-increment/quadratic-gap argument in Appendix A.5.3 apply with $\varepsilon = \overline{S}$. □

F LIMITATIONS

The study examines individual recurrent updates in frozen looped Transformers, using reference utility under teacher forcing to measure task progress. Each intervention scales an already computed proposal and evaluates the resulting state before further recurrence. This isolates the immediate effect of update magnitude along the model's direction. Applying such interventions during generation would require evaluating their effects on subsequent states, answer quality, and computation cost. The present measurements identify recoverable losses within individual updates; how these losses accumulate over a recurrent trajectory remains to be established.