RETHINKING FIXED-HORIZON LATENT REASONING

Anonymous authors
Paper under double-blind review

ABSTRACT

Existing latent reasoners typically run for a preset latent horizon during inference. Intuitively, a longer horizon should provide more opportunity for reasoning. However, in this work, we find that extending latent Chain-of-Thought (CoT) to the full preset horizon may in fact degrade reasoning performance. We observe that latent CoT evolves through distinct functional stages, with earlier computation forming frontier reasoning states, while later computation mainly reorganizes existing candidates and can weaken gold candidate support. These findings motivate a rethink of fixed-horizon latent inference. To address this, we characterize the reasoning benefits of latent CoT by realized and remaining constructive progress. Empirically, we find that higher realized constructive progress corresponds to higher current reasoning accuracy, while more remaining progress corresponds to larger future gains from continued latent CoT. To this end, we introduce a lightweight method that estimates the remaining reasoning progress and adaptively adjusts the latent horizon during inference. Across representative latent CoT reasoners and multiple arithmetic benchmarks, our method improves reasoning accuracy by up to 7.2 percentage points while reducing latent computation by up to 29.0%, without retraining the LLM.

1 INTRODUCTION

Recent advances in latent reasoning (Deng et al., 2023; Hao et al., 2024) have established latent Chain-of-Thought (CoT) as an alternative to explicit CoT reasoning (Wei et al., 2022). Central to the appeal of latent CoT is its ability to carry intermediate reasoning directly in continuous representations, without realizing each step as discrete language tokens (Xu et al., 2025; Zhang et al., 2026). Prior work has demonstrated encouraging capabilities of this paradigm, including the maintenance of multiple reasoning alternatives (Zhu et al., 2026b) and the compression of explicit reasoning into continuous states (Cheng & Van Durme, 2024; Shen et al., 2025). Existing latent reasoners typically perform a fixed number of latent steps before generating an explicit answer (Wei et al., 2026). Intuitively, a longer latent horizon should provide more opportunity for latent CoT to advance its reasoning (Goyal et al., 2024). Yet, while prior work has established what latent CoT can achieve (Tan et al., 2026), how its reasoning benefits evolve across latent steps remains poorly understood.

In this work, we find that completing the full latent horizon may in fact degrade reasoning performance. To assess this, we compare the outcome at the preset full horizon with the best outcome reached earlier along the same latent trajectory. As shown in Figure 1, better reasoning outcomes can emerge before the preset horizon is exhausted and fail to be preserved at the full horizon. This suggests that reasoning benefits do not simply accumulate as latent CoT becomes deeper.

This key observation is closely related to how latent CoT evolves along the reasoning trajectory. We find that the latent trajectory can exhibit three distinct stages: frontier advancement, candidate organization, and answer revision. At the frontier stage, earlier evolution continues to form new task-relevant reasoning states, while later organization increasingly reorganizes candidates that have already emerged. As answer candidates become established, later latent steps mainly revise existing candidates and yield diminishing reasoning gains as latent evolution continues. Thus, the reasoning value of latent CoT is non-uniform along the reasoning trajectory.

The above observations motivate a rethink of fixed-horizon latent inference. Fixed-horizon inference executes the same preset depth regardless of how reasoning has progressed along the latent trajectory. Ideally, latent CoT should continue while further computation still benefits the reasoning

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.

process. Once later latent steps mainly revise gold candidates with diminishing additional gains, computation should be curtailed (Chen et al., 2024). However, this leaves a key challenge in deciding when to curtail latent CoT, before the gains from future latent steps become observable.

To this end, we introduce a lightweight method that estimates the remaining reasoning progress and adaptively adjusts the latent horizon during inference. Specifically, we estimate how much reasoning progress remains from the observed latent trajectory and use this estimate to determine the realized latent horizon. This allows the executed latent horizon to vary with the reasoning progress of the current trajectory. We evaluate trained latent CoT reasoners spanning CODI (Shen et al., 2025), Coconut (Hao et al., 2024), and SIM-CoT (Wei et al., 2026) variants on multiple arithmetic benchmarks. Empirically, our method improves final reasoning accuracy across all settings, with gains of up to 7.2 percentage points while reducing latent computation by up to 29.0%. Interested readers can refer to Appendix A for a thorough discussion of related work.

Statement of Contribution. Our contributions are threefold. First, we show that completing the full latent horizon can degrade reasoning performance and reveal three distinct stages of latent CoT. Second, we formulate latent continuation in terms of remaining reasoning progress and introduce a lightweight method that can adapt the latent horizon during inference. Third, we validate adaptive latent inference on multiple arithmetic benchmarks across CODI, Coconut, and SIM-CoT latent CoT reasoners, consistently improving reasoning accuracy while reducing latent computation.

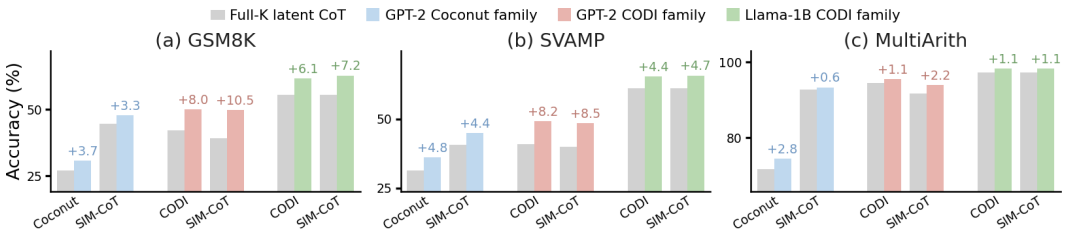

Figure 1: **The full latent horizon can bury better reasoning outcomes.** We evaluate six trained latent CoT reasoners across GPT-2 (Radford et al., 2019) and Llama-1B (Grattafiori et al., 2024) backbones, covering Coconut, CODI, and SIM-CoT variants. For each problem, we retrospectively stop latent reasoning at the first depth that produces a correct answer, if such a depth exists. Gray bars show accuracy at the preset full horizon, while colored bars show accuracy under this retrospective stopping point.

2 Problem Formulation

Latent CoT. Latent CoT carries the intermediate computations as a sequence of latent reasoning states. We index latent computation by $t \in \{1, \ldots, K\}$, where K denotes the preset latent horizon, and let z_t denote the reasoning state after the t-th latent step. Let x denote an input problem with gold answer $y^\star$, and let $y = (y_1, \ldots, y_L) \in \mathcal{Y}$ denote a generated answer. Given x and the preceding latent states $z_{<t}$, the model induces $z_t \sim \mathcal{T}_\theta(\cdot \mid x, z_{<t})$. After t latent steps, the model induces an answer distribution $p_\theta(y \mid x, z_{1:t})$. We use a task reward $r(x, y)$ to score answers generated at each depth. In our reasoning tasks, the reward is instantiated as answer correctness with respect to $y^\star$; implementation details are provided in Appendix B.1.

Adaptive Latent Horizon. Standard latent CoT inference executes all K preset latent steps before answer generation. Reasoning outcomes can vary across latent depths. In particular, an intermediate latent depth can yield a better reasoning outcome than the full horizon. This observation motivates adjusting the latent CoT budget during inference. Let $\tau \in \{1, \ldots, K\}$ denote the realized latent horizon, with the resulting answer generated as $y \sim p_\theta(\cdot \mid x, z_{1:\tau})$. At step t, continuation can depend only on the input x and the prefix latent states $z_{1:t}$. We study how to determine τ from the latent computation available during inference.

3 Functional Evolution in Latent Reasoning

Explicit CoT externalizes intermediate reasoning as a visible sequence of tokens (Nye et al., 2021). By contrast, continuous latent reasoning carries intermediate reasoning without explicitly decoding

each step, allowing multiple reasoning alternatives to remain unresolved (Gozeten et al., 2026). *However, how these alternatives evolve in the latent space remains unclear*. To examine this, we use reasoning candidates as a common reference and compare their support at successive steps.

3.1 Structured Progression along Latent Reasoning

To ground our investigation, we use ProsQA (Hao et al., 2024), where each problem is paired with an explicit reasoning graph. This structure allows us to associate reasoning candidates with different depths. We mainly analyze the symbolic Coconut model (Zhu et al., 2026b). Since intermediate reasoning remains latent, directly characterizing its evolution is difficult. Inspired by Hao et al. (2024), we decode reasoning candidates at each latent step and track how their probabilities change. For clarity, we present a selected example here and defer additional results to Appendix B.3.

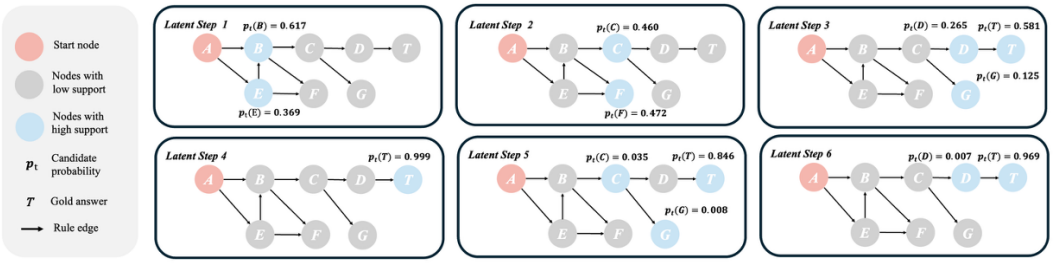

Figure 2: **An example of latent reasoning progress.** We use the symbolic Coconut model from Zhu et al. (2026b) and roll it out for six latent steps to trace the reasoning progress. We report the top-6 candidates at each step, with highlighted nodes indicating high support.

Frontier advancement. Figure 2 (Steps 1–4) shows the support assigned to reasoning candidates at different graph depths over successive latent steps. Candidates at smaller graph depths receive support earlier, followed by candidates at larger depths, while answer candidates emerge later. Earlier intermediate candidates tend to receive support only briefly, whereas later candidates remain supported across more latent steps. Together, these patterns show an advancing reasoning frontier, with support moving through intermediate reasoning before answer candidates become established.

Candidate organization. As latent reasoning proceeds, as shown in Figure 2 (Steps 3–5), new intermediate reasoning candidates enter the high support region less frequently. In contrast, candidates that have already emerged continue receiving support, while answer candidates increasingly dominate the support. This suggests that latent reasoning is no longer continually forming new candidates with high support, but organizing candidates that have already emerged. We term this stage candidate organization.

Answer revision. Further, as shown in Figure 2 (Steps 4–6), once the gold answer candidate becomes the most supported, later latent steps still shift support among other candidates. In our tests, this shift can weaken the gold candidate's support and even cause it to lose the lead. Meanwhile, few new candidates enter, with the remaining changes concentrated among already formed answer candidates. Thus, we argue that the later latent reasoning enters an answer revision stage: the model mainly reorganizes the support among existing answer candidates, and this revision can overturn an already established gold answer.

3.2 Reasoning Gains Along the Progression

Building upon the progress identified above, we empirically test whether latent computation consistently yields further gains. Intuitively, while the solution remains unresolved, latent computation should continue to advance the reasoning frontier, repeatedly updating the top-k candidates. Once the gold answer begins to form, further computation should increasingly consolidate support around a stable solution. To make this intuition concrete, we consider two empirical implications: (1) While the reasoning frontier is still advancing, prematurely realizing the current latent state should reduce subsequent progress toward the gold answer. (2) Once the gold answer has formed, later latent computation should consolidate the gold result if it yields further gains.

To test the first empirical implication, we interrupt the ongoing latent CoT at the frontier stage by inserting a single explicit realization. Specifically, we decode the current state to its top token, add this token to the causal context, and continue latent computation. As shown in Figure 3(a), realization markedly delays subsequent progression toward the gold answer. To more rigorously assess the gain during the frontier stage, we introduce a matched control that perturbs the continuous state by a comparable magnitude. As shown in Figure 3(b), with perturbations, subsequent candidate progression remains largely preserved. Inspecting the top-k candidates further, we find that the gold answer already appears while the frontier is still advancing, suggesting that frontier stage gains can continue after the answer first emerges. Taken together, these results suggest that the function of the frontier stage is to sustain candidate progression toward the gold answer, thereby yielding further reasoning gains.

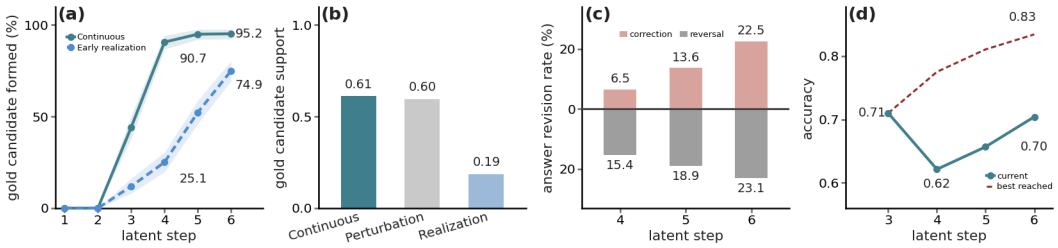

Figure 3: **Latent progression and revision.** **(a)** Gold candidate formation; **(b)** Gold candidate support. For trajectories whose answer is formed at the third latent step, **(c)** Correction and reversal rates, and **(d)** Accuracy over revision vs. the best accuracy reached along the trajectory.

To examine the second empirical implication, we assess whether later latent computation continues to consolidate the gold result after it has emerged. We quantify this by tracking later corrections toward the gold answer and reversals of an already established gold answer. As shown in Figure 3(c), later computation recovers the gold answer in some cases, while also overturning gold results that have already been established, leaving little net gain. This suggests that later revisions yield little stable additional reasoning gain. More importantly, as illustrated in Figure 3(d), some later revisions overturn an already established gold result, allowing gains reached earlier in the trajectory to be lost during continued computation. Consequently, continued latent evolution can weaken reasoning gains that have already emerged.

Together, these results show that the effect of latent CoT is non-uniform, giving rise to three functional stages. Specifically, frontier advancement continues to yield reasoning gains, candidate organization marks the transition, and later revision mainly reorganizes the formed candidates. In other words, once constructive progress has largely saturated, additional latent computation offers diminishing returns and even becomes detrimental.

4 Remaining Reasoning Progress in Latent Computation

So far, we have identified distinct reasoning stages in latent CoT, revealing non-uniform reasoning benefits across latent steps. In this section, we first examine the relation between latent CoT evolution and remaining progress (Section 4.1). Building on this, we propose a lightweight method that can estimate the remaining reasoning progress and adaptively adjust the latent horizon during inference to improve latent thinking (Section 4.2). Finally, we provide a formal characterization of progress estimation and allocation (Section 4.3).

4.1 Reasoning Gains Beyond the Current Frontier

The stage structure suggests a simple strategy: continue latent CoT during the frontier stage and stop once reasoning shifts to revision. This strategy, however, introduces two difficulties.

Ambiguity in latent gains. Intuitively, the extent of latent evolution may appear to provide a natural signal of reasoning progression (Geiping et al., 2026). Yet, we find that substantial latent evolution can persist during revision, making state evolution insufficient for distinguishing different stages. Moreover, the same ambiguity persists in realized reasoning gains, since a substantial local gain does

not necessarily establish a new frontier. For instance, later revision can induce large fluctuations that recover a previously attained value without advancing the frontier.

Dependence on future gains. Section 3 shows that latent evolution can persist even as genuinely new reasoning gains become sparse. Adapting the budget of latent CoT therefore requires judging whether further computation can still produce additional gain, which is realized in the remaining latent suffix $z_{t+1:K}$, while inference only has access to the realized computation $z_{1:t}$. The appropriate endpoint can therefore only be identified retrospectively from future gains, whereas continuation must be decided online from the current state.

4.2 From Remaining Progress to an Adaptive Latent Horizon

To resolve the ambiguity, we evaluate each latent depth against the best reasoning outcome reached so far, which defines the current reasoning frontier. Then, a latent step can establish constructive progress only when it advances this frontier. For online continuation, the core question is whether latent CoT is worth continuing. We address this in two steps: estimating how much reasoning progress remains and adapting the latent horizon accordingly.

Formulation. Consider an input x and a K-step latent CoT $z_1, \ldots, z_K$. After t latent steps, we define the current reasoning value as

$$G_t = \mathbb{E}_{y \sim p_\theta(\cdot|x, z_{1:t})}\big[r(x, y)\big]. \tag{1}$$

To quantify the remaining reasoning progress, we compare the reasoning frontier reached up to step t with the best frontier attainable within the full latent horizon, written as:

$$\Omega_t = \max_{1 \le j \le K} G_j - \max_{1 \le j \le t} G_j. \tag{2}$$

By this definition, we now assess whether latent reasoning is worth continuing by estimating how much further frontier advancement remains. At inference, Ω_t is not directly available because it depends on future reasoning gains. To estimate it from the current latent prefix, we represent the observed evolution of latent CoT with a causal trajectory representation X_t, which captures short-range changes in its predictive statistics and observed hidden state. We defer its detailed construction to Appendix B.2.

Estimation. The above formulation allows us to train a lightweight estimator f_ϕ to predict the remaining progress from X_t. To construct the training targets, we run the fixed latent reasoner over the full K step trajectory and roll out from each checkpoint to estimate its reasoning value $\hat{G}_t$, yielding the remaining progress target $\hat{\Omega}_t$ according to Eq. (2). We optimize f_ϕ by minimizing the squared error between prediction and the remaining progress target:

$$\mathcal{L}_{\text{prog}}(\phi) = \mathbb{E}_{(X_t, \widehat{\Omega}_t)}\left[\left(f_\phi(X_t) - \widehat{\Omega}_t\right)^2\right]. \tag{3}$$

Horizon adaptation. With this estimate, identifying an exact stopping point would still require predicting when future progress will occur. We aim for a common continuation mechanism across latent reasoners, even when progress is represented differently across backbones. Accordingly, we express each estimated remaining progress by its relative position in a frozen reference distribution. For each latent reasoner, let F_{ref} denote the cumulative distribution of its reference progress estimates. We define

$$u_t = F_{\text{ref}}[f_\phi(X_t)]\,. \tag{4}$$

Thus, u_t represents the relative continuation demand associated with the estimated remaining progress. We use u_t to control how much additional latent computation is allocated over the remaining horizon. Let a_t denote the additional latent computation allocated after step t. We write

$$a_t = (K - t)\, g(u_t), \tag{5}$$

where g is a bounded, non-decreasing continuation rule, with its concrete form deferred to Appendix B.2. After discretization, the allocation yields the realized horizon τ. In this way, we replace the preset horizon K with an adaptive horizon determined solely from the current state.

4.3 Latent Trajectories and Returns from Additional Computation

One might wonder whether the latent trajectory provides sufficient information to adapt the remaining latent budget. To understand this, we examine how the latent CoT supports remaining progress estimation and how remaining progress affects the additional latent gain.

Estimation. Although Ω_t depends on later reasoning gains, online continuation only has access to X_t. We therefore characterize the optimal prediction of Ω_t from X_t under population squared risk.

Proposition 4.1 (Causal Estimation of Remaining Progress)**.** *Under the population squared risk $\mathcal{R}(f) = \mathbb{E}[(f(X_t) - \Omega_t)^2]$, the optimal estimator of remaining progress from the latent trajectory satisfies*

$$f^*(X_t) = \mathbb{E}[\Omega_t \mid X_t]\,. \tag{6}$$

The proof is provided in Appendix C.1. Thus, remaining progress can be predicted from the latent trajectory through its conditional expectation, which defines the population target learned by f_ϕ.

Allocation. For continuation, we consider how the reasoning gain from additional latent computation varies with remaining progress. Let ω denote the remaining progress and a the additional latent computation. We define $V(a, \omega)$ as the expected change in reasoning value after a additional latent steps when the remaining progress is ω. We consider the condition that a larger ω yields a larger marginal gain from additional latent computation. At latent depth t, let $A = K - t$ denote the maximum additional latent budget.

Theorem 4.2 (Monotone Allocation by Remaining Progress)**.** *Let $V(a, \omega)$ have increasing differences in (a, ω). For trajectories $i = 1, \ldots, n$ with remaining progress ω_i, consider integer allocations $a_i \in \{0, \ldots, A\}$ that maximize $\sum_i V(a_i, \omega_i)$ subject to a shared compute budget $\sum_i a_i \le B$. Then there exists an optimal allocation a^* such that $\omega_i > \omega_j$ implies $a_i^* \ge a_j^*$.*

Under these conditions, trajectories with more remaining progress receive no less additional latent computation. We defer the proof to Appendix C.3. Thus, the value of additional latent computation varies systematically with the remaining progress. Together, these results recast latent continuation in terms of online reasoning progress, providing a basis for estimating the remaining progress along the latent trajectory and adapting the latent horizon accordingly.

5 Experiments

In this section, we investigate how reasoning progress relates to both the outcomes of latent CoT and the gains from continued reasoning. We evaluate latent CoT reasoners spanning CODI, Coconut, and SIM-CoT variants on GSM8K (Cobbe et al., 2021), SVAMP (Patel et al., 2021), and MultiArith (Roy & Roth, 2015), with experimental settings provided in Appendix B. Our experimental design aims to examine three claims: (i) that adapting the latent horizon improves final reasoning outcomes while reducing latent computation, (ii) that higher realized constructive progress corresponds to higher current reasoning accuracy, and (iii) that more remaining progress corresponds to larger future gains from continued latent CoT.

5.1 Adaptive Latent Inference and Reasoning Outcomes

We first report the overall results of adaptive latent inference on final reasoning accuracy. As shown in Table 1, adaptive inference achieves higher final accuracy than the original Full-K inference across all settings. The largest gain is observed on MultiArith with GPT-2/Coconut, where accuracy improves from 71.7% to 78.9%. Moreover, the improvement persists even in settings where the original Full-K accuracy is already high and further improvement is more limited. For instance, on MultiArith with Llama-1B/CODI, accuracy still increases from 97.2% to 98.3%. Meanwhile, adaptive inference also reduces latent computation, with notable savings on latent steps. Consequently, we show that adaptive latent inference enables the same latent reasoner to achieve higher final reasoning accuracy while using less latent computation.

One might wonder whether, once the full preset horizon is no longer necessary, a shorter fixed-horizon would already suffice. To examine this, we compare fixed depth and adaptive inference under the same latent step budgets. As shown in Table 2, the advantage of adaptive inference is

already pronounced even under a limited latent step budget. For example, on GSM8K, the accuracy of GPT-2/CODI increases from 36.6% to 41.5% at four latent steps. The advantage persists as the latent CoT depth increases, remaining evident at both five and six steps. Delving into the fixed depth performance, we find that increasing the latent step budget can improve, leave unchanged, or even reduce reasoning accuracy. This pattern is consistent with our observation in Section 3 that additional latent computation yields distinct gains across three reasoning stages. Thus, these results support adapting the CoT depth according to the reasoning progress throughout inference.

Taken together, these results suggest that continued latent computation can alter whether reasoning gains established earlier are retained in the final prediction. Accordingly, with fixed-horizon inference, latent CoT proceeds to the preset depth even after useful reasoning gains have formed, and later revision can alter the resulting reasoning outcome. By contrast, adaptive latent inference adjusts the CoT depth based on reasoning progress, curtailing computation as further gains diminish. In this way, we can preserve constructive reasoning progress with less latent computation, resulting in better reasoning performance.

Table 1: **Adaptive latent CoT inference.** Full-K denotes the preset latent horizon while ours uses adaptive latent inference. Latent step saving (%) reports the average step reduction.

	GSM8K			SVAMP			MultiArith			Latent step
Method	Full-K	Ours	Gain	Full-K	Ours	Gain	Full-K	Ours	Gain	saving (%)
GPT-2										
CODI	42.1	**43.4**	+1.3	41.0	**43.1**	+2.1	94.4	**95.0**	+0.6	20.6
SIM-CoT	39.2	**41.5**	+2.3	40.0	**43.5**	+3.5	91.7	**92.2**	+0.5	29.0
Coconut	27.1	**27.7**	+0.6	31.4	**33.7**	+2.3	71.7	**78.9**	**+7.2**	11.0
SIM-CoT Coconut	44.7	44.7	+0.0	40.6	**42.2**	+1.6	92.8	**93.3**	+0.5	25.1
Llama-1B										
CODI	55.5	**55.7**	+0.2	61.1	**62.3**	+1.2	97.2	**98.3**	+1.1	26.9
SIM-CoT	55.6	**55.9**	+0.3	61.1	**62.2**	+1.1	97.2	**97.8**	+0.6	21.1

Table 2: **Comparison of latent CoT performance under adaptive and fixed-preset inference.** We report task accuracy (%) under maximum latent budgets of 4, 5, and 6 steps.

Model / Budget policy		Accuracy (%)								
		GSM8K			SVAMP			MultiArith		
(a) CODI family		4 steps	5 steps	6 steps	4 steps	5 steps	6 steps	4 steps	5 steps	6 steps
GPT-2										
CODI	Fixed depth	36.6	41.2	42.1	41.5	41.0	41.0	92.8	93.3	94.4
	Adaptive	**41.5**	**43.2**	**43.4**	**42.2**	**42.2**	**43.1**	92.8	93.3	**95.0**
SIM-CoT	Fixed depth	35.3	39.1	39.2	41.3	40.6	40.0	89.4	91.1	91.7
	Adaptive	**40.6**	**41.0**	**41.5**	**42.8**	**42.6**	**43.5**	89.4	91.1	**92.2**
Llama-1B										
CODI	Fixed depth	48.0	54.7	55.5	61.1	61.6	61.1	97.2	**97.8**	97.2
	Adaptive	**53.8**	**55.0**	**55.7**	**62.0**	**61.7**	**62.3**	97.2	97.2	**98.3**
SIM-CoT	Fixed depth	51.1	54.6	55.6	61.5	61.1	61.1	**97.8**	96.7	97.2
	Adaptive	**54.1**	**55.2**	**55.9**	**62.0**	**62.2**	**62.2**	97.2	**97.2**	**97.8**
(b) Coconut family		4 steps	5 steps	6 steps	4 steps	5 steps	6 steps	4 steps	5 steps	6 steps
GPT-2										
Coconut	Fixed depth	21.0	**25.8**	27.1	30.5	30.5	31.4	41.7	70.0	71.7
	Adaptive	21.0	25.6	**27.7**	**33.2**	**33.0**	**33.7**	**65.0**	**75.6**	**78.9**
SIM-CoT Coconut	Fixed depth	**27.9**	34.6	38.2	40.1	40.0	40.7	92.8	**92.2**	**92.8**
	Adaptive	26.5	34.6	**40.4**	40.1	**40.7**	**41.6**	92.8	91.1	91.1

5.2 Constructive Progress Across Latent CoT

Building on the advantages of adaptive latent CoT, we now examine constructive progress in trained latent reasoners. To understand why completing the full latent horizon can degrade reasoning performance, we first characterize how constructive progress evolves across latent depths, and then relate the progress to the reasoning performance.

Specifically, we characterize the distribution of constructive progress completion across latent steps. As shown in Figure 4, constructive progress is often established early, with a majority already completed by the third latent step across all six reasoners. Despite this common tendency, its progression differs across reasoners, with the CODI variants showing earlier completion while the Coconut variants retain more progress for later latent steps. These results indicate that trained latent reasoners can exhibit distinct constructive progress evolution.

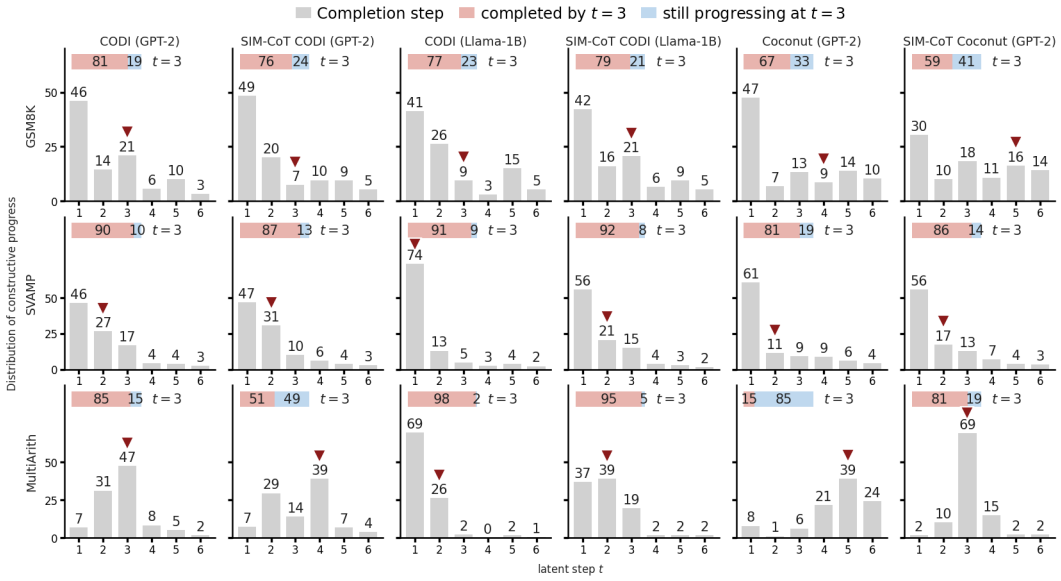

Figure 4: **Constructive progress completion across latent CoT**. Bars show how constructive progress completion is distributed. The top strips summarize completion status at latent step $t = 3$, contrasting completed and still progressing. Triangles mark the first latent step where at least 75% of examples have completed constructive progress. Completion marks the first latent step where the latent CoT realizes all constructive progress.

Table 3: **Reasoning performance across constructive progress.** At latent depth $t = 3$, we compare task accuracy (%) across different levels of realized constructive progress, as defined in Section 4.2.

Backbone	Method	GSM8K		SVAMP		MultiArith	
		Lower	Higher	Lower	Higher	Lower	Higher
GPT-2	CODI	29.3	45.0	16.6	70.8	81.0	96.7
	SIM-CoT	10.8	53.3	13.2	73.4	25.3	94.1
Llama-1B	CODI	24.9	69.8	54.4	77.0	95.1	99.0
	SIM-CoT	34.7	67.2	45.3	80.3	94.5	98.4
GPT-2	Coconut	11.4	32.8	21.1	45.3	31.4	38.7
	SIM-CoT Coconut	6.9	47.2	26.5	66.9	77.1	94.7

We further examine the relation between realized constructive progress and reasoning performance. At the third latent step, we compare task accuracy across different levels of realized constructive progress. As shown in Table 3, higher constructive progress consistently corresponds to higher reasoning accuracy. For instance, on SVAMP with GPT-2/SIM-CoT, accuracy increases from 13.2% under Lower progress to 73.4% under Higher progress. These results show that reasoning performance mainly depends on how much constructive progress has already been realized.

These results indicate that the same latent depth can correspond to different levels of constructive progress. At a fixed latent depth, reasoning performance varies systematically with the realized constructive progress. This relation is also consistent with the gains from adaptive inference in Table 1. Adaptive inference can adjust the executed horizon of CoT according to the evolution of constructive progress, while a fixed latent CoT cannot.

5.3 Remaining Progress and Computation Returns

So far, we have shown that latent reasoning outcomes vary with constructive progress along the reasoning frontier. Here, we test the relation between remaining progress and future gains from continued latent CoT. Recall from Section 4.2 that Ω_3 denotes the remaining constructive progress at the third latent step. We compare task accuracy at Steps 3 and 4 under lower and higher Ω_3. As shown in Table 4(a), task accuracy consistently improves more from the third to the fourth latent step under higher Ω_3. Interestingly, higher Ω_3 can coincide with lower accuracy at the third step while yielding substantially larger gains after continuation. For instance, with GPT-2/SIM-CoT, accuracy remains at 35.0% under lower Ω_3, whereas it increases from 18.5% to 57.3% under higher Ω_3. Therefore, we argue that more remaining constructive progress corresponds to a larger reasoning gain from continuing latent CoT.

We further extend this comparison over the subsequent latent steps, using Step 3 as the common starting point. As shown in Table 4(b), the future gain difference associated with higher Ω_3 grows over subsequent latent steps. Notably, this difference can remain small after one latent step but become larger as latent CoT continues. For instance, with Llama-1B/CODI, the difference is only 1.6% at Step 4, but later expands to 58.4%. These results indicate that the future gain indicated by remaining constructive progress can be realized cumulatively over subsequent latent CoT.

Table 4: **Reasoning gains across remaining constructive progress.** Ω_3 means the remaining constructive progress after 3 latent steps. Panel **(a)** reports task accuracy at Steps 3 and 4 under lower and higher Ω_3. Panel **(b)** reports future reasoning gain, measured by the change in $\hat{G}_t$ from Step 3 to Steps 4 and 5, under lower and higher Ω_3, with subscripts indicating the difference relative to lower Ω_3. Panel **(c)** reports the change in future reasoning gain as Ω_t decreases, with negative values indicating a decrease. We report GSM8K here and defer additional results to Appendix D.2.

		(a) Accuracy (%)				(b) Future gain				(c) Future gain change		
		Lower Ω_3		Higher Ω_3		Lower Ω_3		Higher Ω_3		Decrease in Ω_t		
Backbone / Method		Step 3	Step 4	Step 3	Step 4	Step 4	Step 5	Step 4	Step 5	Lower	Middle	Higher
CODI family												
GPT-2	CODI	46.8	46.3 $_{-0.5}$	34.0	35.9 $_{+1.9}$	-0.2	-0.2	$+4.9_{+5.1}$	$+44.0_{+44.2}$	$+0.5$	-3.7	-49.9
	SIM-CoT	35.0	35.0 $_{0.0}$	18.5	57.3 $_{+38.8}$	$+0.1$	$+0.2$	$+37.0_{+36.9}$	$+71.8_{+71.6}$	$+0.4$	-3.3	-47.5
Llama-1B	CODI	39.5	39.5 $_{0.0}$	23.9	31.0 $_{+7.1}$	-0.1	$+1.4$	$+1.5_{+1.6}$	$+59.8_{+58.4}$	-1.2	-9.7	-47.7
	SIM-CoT	51.6	50.5 $_{-1.1}$	23.7	56.3 $_{+32.6}$	-1.5	$+2.0$	$+25.7_{+27.2}$	$+53.1_{+51.1}$	-3.5	-10.7	-53.6
Coconut family												
GPT-2	Coconut	16.8	16.4 $_{-0.4}$	33.0	47.0 $_{+14.0}$	$+0.2$	$+0.3$	$+11.7_{+11.5}$	$+41.2_{+40.9}$	$+1.7$	$+1.8$	-28.3
	SIM-CoT Coconut	27.8	30.8 $_{+3.0}$	2.3	41.7 $_{+39.4}$	$+3.0$	$+2.6$	$+33.8_{+30.8}$	$+79.7_{+77.1}$	$+5.9$	-6.0	-66.4

As these constructive gains are realized, however, the remaining progress changes in turn. Further frontier advancement therefore leaves less further constructive progress. We therefore examine how the future gain changes as Ω_t decreases along this progression. As shown in Table 4(c), larger decreases in Ω_t are associated with larger reductions in the subsequent reasoning gain. The reduction becomes more pronounced as more constructive progress is realized. These results show that the future gain from continued latent CoT diminishes as constructive progress advances the reasoning frontier. Consequently, these results show that future gains from latent CoT evolve systematically with constructive progress along the reasoning frontier.

6 Conclusion

The intuition behind completing a preset latent horizon is that additional latent steps should provide more opportunity for reasoning. However, our study shows that reasoning gains do not simply accumulate with latent depth. Completing the full horizon can overwrite previously reached gold answers and diminish existing reasoning gains. More broadly, our results show that higher realized constructive progress corresponds to higher current reasoning accuracy, while more remaining progress corresponds to larger future gains from continued latent CoT. By rethinking the fixed-horizon of latent CoT, we develop a method that adaptively adjusts the executed horizon during inference. Across widely studied latent CoT approaches, including CODI, Coconut, and SIM-CoT variants, and multiple arithmetic benchmarks, our method consistently improves reasoning accuracy and also reduces latent computation.

AI USE STATEMENT

In this work, we did not use generative AI tools for tasks requiring mandatory disclosure under the ICLR 2027 AI Policy. Additionally, we used generative AI tools to create or edit software code and to edit the research paper to improve readability. We have reviewed all AI-assisted work. AI-assisted code was verified and tested by the authors, and AI-assisted edits to the manuscript were reviewed by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The source code used for our experiments is provided in the supplementary materials and will be released in a public repository. Appendix B provides the experimental details for task-level evaluation, adaptive latent inference, and latent progression analysis. Appendix C provides the assumptions and complete proofs for the theoretical results. Appendix D reports additional experimental results.

REFERENCES

Pranjal Aggarwal, Aman Madaan, Yiming Yang, et al. Let's sample step by step: Adaptive-consistency for efficient reasoning and coding with llms. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 12375–12396, 2023.

Andrea Banino, Jan Balaguer, and Charles Blundell. Pondernet: Learning to ponder. *arXiv preprint arXiv:2107.05407*, 2021.

Xinghao Chen, Anhao Zhao, Heming Xia, Xuan Lu, Hanlin Wang, Yanjun Chen, Wei Zhang, Jian Wang, Wenjie Li, and Xiaoyu Shen. Reasoning beyond language: A comprehensive survey on latent chain-of-thought reasoning. *arXiv preprint arXiv:2505.16782*, 2025.

Xingyu Chen, Jiahao Xu, Tian Liang, Zhiwei He, Jianhui Pang, Dian Yu, Linfeng Song, Qiuzhi Liu, Mengfei Zhou, Zhuosheng Zhang, et al. Do not think that much for 2+ 3=? on the overthinking of o1-like llms. *arXiv preprint arXiv:2412.21187*, 2024.

Jeffrey Cheng and Benjamin Van Durme. Compressed chain of thought: Efficient reasoning through dense representations. *arXiv preprint arXiv:2412.13171*, 2024.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, and Łukasz Kaiser. Universal transformers. *arXiv preprint arXiv:1807.03819*, 2018.

Yuntian Deng, Kiran Prasad, Roland Fernandez, Paul Smolensky, Vishrav Chaudhary, and Stuart Shieber. Implicit chain of thought reasoning via knowledge distillation. *arXiv preprint arXiv:2311.01460*, 2023.

Connor Dilgren and Sarah Wiegreffe. Are latent reasoning models easily interpretable? *arXiv preprint arXiv:2604.04902*, 2026.

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. *Advances in Neural Information Processing Systems*, 38: 41340–41391, 2026.

Sachin Goyal, Ziwei Ji, Ankit Singh Rawat, Aditya Krishna Menon, Sanjiv Kumar, and Vaishnavh Nagarajan. Think before you speak: Training language models with pause tokens. In *International Conference on Learning Representations*, volume 2024, pp. 27896–27923, 2024.

Alperen Gozeten, Muhammed Ildiz, Xuechen Zhang, Hrayr Harutyunyan, Ankit Singh Rawat, and Samet Oymak. Continuous chain of thought enables parallel exploration and reasoning. In *International Conference on Learning Representations*, volume 2026, pp. 82881–82914, 2026.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Alex Graves. Adaptive computation time for recurrent neural networks. *arXiv preprint arXiv:1603.08983*, 2016.

Shibo Hao, Sainbayar Sukhbaatar, DiJia Su, Xian Li, Zhiting Hu, Jason Weston, and Yuandong Tian. Training large language models to reason in a continuous latent space. *arXiv preprint arXiv:2412.06769*, 2024.

Maxwell Nye, Anders Johan Andreassen, Guy Gur-Ari, Henryk Michalewski, Jacob Austin, David Bieber, David Dohan, Aitor Lewkowycz, Maarten Bosma, David Luan, et al. Show your work: Scratchpads for intermediate computation with language models. *arXiv preprint arXiv:2112.00114*, 2021.

Arkil Patel, Satwik Bhattamishra, and Navin Goyal. Are nlp models really able to solve simple math word problems? In *Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: human language technologies*, pp. 2080–2094, 2021.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

Michael Rizvi-Martel, Guillaume Rabusseau, and Marius Mosbach. The illusion of superposition? a principled analysis of latent thinking in language models. *arXiv preprint arXiv:2604.06374*, 2026.

Subhro Roy and Dan Roth. Solving general arithmetic word problems. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pp. 1743–1752, 2015.

Zhenyi Shen, Hanqi Yan, Linhai Zhang, Zhanghao Hu, Yali Du, and Yulan He. Codi: Compressing chain-of-thought into continuous space via self-distillation. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 677–693, 2025.

Charlie Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. Scaling llm test-time compute optimally can be more effective than scaling model parameters. *arXiv preprint arXiv:2408.03314*, 2024.

Wenhui Tan, Jiaze Li, Jianzhong Ju, Zhenbo Luo, Ruihua Song, and Jian Luan. Think silently, think fast: Dynamic latent compression of llm reasoning chains. *Advances in Neural Information Processing Systems*, 38:4646–4668, 2026.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.

Xilin Wei, Xiaoran Liu, Yuhang Zang, Xiaoyi Dong, Yuhang Cao, Jiaqi Wang, Xipeng Qiu, and Dahua Lin. Sim-cot: Supervised implicit chain-of-thought. In *International Conference on Learning Representations*, volume 2026, pp. 56721–56742, 2026.

Yige Xu, Xu Guo, Zhiwei Zeng, and Chunyan Miao. Softcot: Soft chain-of-thought for efficient reasoning with llms. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 23336–23351, 2025.

Chenxu Yang, Qingyi Si, Yongjie Duan, Zheliang Zhu, Chenyu Zhu, Qiaowei Li, Minghui Chen, Zheng Lin, and Weipinng Wang. Dynamic early exit in reasoning models. In *International Conference on Learning Representations*, volume 2026, pp. 88170–88210, 2026.

Zhen Zhang, Xuehai He, Weixiang Yan, Ao Shen, Chenyang Zhao, and Xin Wang. Soft thinking: Unlocking the reasoning potential of llms in continuous concept space. *Advances in Neural Information Processing Systems*, 38:168990–169012, 2026.

Hanlin Zhu, Shibo Hao, Zhiting Hu, Jiantao Jiao, Stuart Russell, and Yuandong Tian. Emergence of superposition: Unveiling the training dynamics of chain of continuous thought. In *International Conference on Learning Representations*, volume 2026, pp. 62257–62288, 2026a.

Hanlin Zhu, Shibo Hao, Zhiting Hu, Jiantao Jiao, Stuart J Russell, and Yuandong Tian. Reasoning by superposition: A theoretical perspective on chain of continuous thought. *Advances in Neural Information Processing Systems*, 38:79931–79963, 2026b.

Appendix

A. Related Work

Recent work has developed several formulations of latent CoT, ranging from implicit reasoning through distillation (Deng et al., 2023) to direct reasoning in continuous latent space (Hao et al., 2024). Meanwhile, latent CoT has also been studied through self-distillation (Shen et al., 2025), soft representations (Xu et al., 2025), and supervised implicit CoT (Wei et al., 2026). These studies establish latent CoT as a broad family of reasoning methods acquired through distinct training schemes. One common structure is that intermediate reasoning unfolds through successive latent states before explicit answer generation. Such latent trajectories have been considered a common structure for multi-step inference across different latent CoT methods (Chen et al., 2025).

While early work primarily focused on training latent reasoners, recent studies have increasingly examined the mechanisms underlying latent CoT. Specifically, at the representation level, some works characterize latent states as compressed representations of explicit reasoning (Dilgren & Wiegreffe, 2026). Meanwhile, some studies show that continuous reasoning can preserve multiple reasoning alternatives in superposition (Zhu et al., 2026b), enabling parallel exploration over candidate reasoning traces (Gozeten et al., 2026). More recently, the community has also begun to examine how latent reasoning evolves across successive steps, both through its training dynamics (Zhu et al., 2026a) and through probing of internal representations (Rizvi-Martel et al., 2026). Taken together, existing work has predominantly focused on characterizing the structure and evolution of latent reasoning, leaving less examined how its reasoning benefits evolve along the latent trajectory. Our work complements these studies by characterizing how the reasoning value of latent CoT changes along this progression.

The core insight of adaptive reasoning methods is to vary the reasoning depth across different instances (Graves, 2016; Dehghani et al., 2018). Early approaches often set the reasoning depth according to problem difficulty (Banino et al., 2021; Snell et al., 2024), while recent methods increasingly adapt it using confidence or consistency in intermediate predictions (Aggarwal et al., 2023; Yang et al., 2026). Yet, these criteria primarily characterize problem difficulty or the current prediction, leaving the value of further reasoning less directly examined. Dynamic compression has further extended this adaptive view to latent reasoning chains (Tan et al., 2026). Nevertheless, these methods generally assume that additional latent steps can yield further reasoning gain. In this paper, we revisit this assumption, characterize where latent CoT gains arise, and develop a method that adapts its depth accordingly to improve reasoning accuracy.

B. Experimental Details

B.1. Task-Level Evaluation

Latent depth evaluation. For task-level evaluation, we decode a complete answer from each evaluated latent prefix. Within each reasoner, intermediate depths and Full-K use the same answer-decoding procedure, differing only in the latent prefix length.

Benchmarks and scoring. For arithmetic reasoning tasks, we evaluate on the official GSM8K test split, the full SVAMP benchmark, and the MultiArith test split. We use greedy decoding throughout and follow the original evaluation protocol of each latent reasoner family for final-answer extraction and matching. We instantiate $r(x, y)$ as binary answer correctness. Each output receives $r(x, y) = 1$ if the extracted answer matches the ground-truth answer, and 0 otherwise. We report accuracy as the mean binary correctness over the full evaluation set, with outputs lacking a valid extracted answer counted as incorrect.

Horizon configuration. Throughout our evaluation, we use a continuous thought, or latent state, as the basic unit of latent depth. For CODI and SIM-CoT CODI, one continuous thought corresponds to one latent step. For Coconut and SIM-CoT Coconut, each latent stage contains two continuous thoughts, corresponding to two latent steps. The native preset horizon is six latent steps for CODI, SIM-CoT CODI, and Coconut, and ten latent steps (five stages) for SIM-CoT Coconut. We use

Full-K to denote the native preset horizon of each latent reasoner. For analyses that require a shared depth axis, we use the first six latent steps as a common analysis window.

B.2. Adaptive Latent Inference

Trajectory representation. For adaptive latent inference, X_t summarizes the observed evolution of the latent trajectory up to step t. We construct X_t from predictive statistics, the latent state, and their short-range changes over the observed prefix. The predictive statistics capture confidence, distributional change, and prediction stability, while the latent state is projected to a 64-dimensional representation together with its first-order change. We use observations from the current and two preceding latent steps, with unavailable history zero-filled together with an explicit missing indicator. The PCA projection is fitted on offline training trajectories.

Remaining-progress estimation. We construct the remaining-progress targets from offline trajectories of the fixed latent reasoner. At each latent depth, we estimate $\hat{G}_t$ as the fraction of sampled continuations that produce the correct answer. We then construct $\hat{\Omega}_t$ from these reasoning-value estimates according to Eq. (2) and fit f_ϕ to predict $\hat{\Omega}_t$ from X_t by squared-error regression. The estimator is fitted separately for each model–task configuration using its offline trajectories. We implement f_ϕ with a histogram-based gradient-boosting regressor.

Horizon adaptation. For each model configuration, we construct F_{ref} as the empirical cumulative distribution of reference progress estimates and keep it fixed when evaluating the current examples. With u_t defined in Eq. (4), we instantiate the continuation rule as

$$g(u_t) = \mathrm{clip}\left(\rho_0 + 0.5(2u_t - 1),\, 0,\, 1\right),$$

where ρ_0 specifies the baseline fraction of the remaining latent horizon. The additional latent budget is $a_t = (K - t)g(u_t)$, following Eq. (5). We apply the estimator and continuation rule once at the current adaptation step and use randomized rounding to convert the resulting budget into an integer realized horizon $\tau \leq K$. For a nonnegative allocation a, randomized rounding returns $\lfloor a \rfloor + 1$ with probability $a - \lfloor a \rfloor$, and $\lfloor a \rfloor$ otherwise.

B.3. Latent Progression Analysis

For the analysis in ProsQA, each node represents a reasoning candidate, and the target node denotes the gold answer. At each latent step, the continuous thought induces a vocabulary distribution, from which we record the probability of each reasoning candidate. We use these probabilities together with graph depth to trace candidate progression over latent steps. Across the full ProsQA set, candidates at larger graph depths tend to receive support at later latent steps, with a Spearman correlation of 0.833. After the gold answer becomes the most supported candidate, later latent steps continue to change the leading candidate, with correction and reversal rates of 16.1% and 15.6%, respectively.

Across the six arithmetic reasoners, we characterize progression directly from task outcomes over latent depth. Constructive progress is realized at different latent depths across reasoners, occurring earlier for the CODI variants and later for the Coconut variants. Consistent with Figure 1, the correctness trajectories also show that outcomes reached at intermediate depths can change under continued latent computation.

C. Proofs and Derivations

C.1. Population Estimation of Remaining Progress

For completeness, we restate Proposition 4.1, which characterizes the optimal population prediction of remaining progress from X_t under squared risk, in its precise form and prove it.

Proposition 4.1 (Causal Estimation of Remaining Progress). *For the population squared risk $\mathcal{R}(f) = \mathbb{E}[(f(X_t) - \Omega_t)^2]$ over square-integrable predictors $f(X_t)$, the minimizer satisfies*

$$f^*(X_t) = \mathbb{E}[\Omega_t \mid X_t] \qquad \textit{a.s.}$$

Proof. Fix a latent depth t and write $m_t(X_t) := \mathbb{E}[\Omega_t \mid X_t]$. For any square-integrable predictor $f(X_t)$,

$$\begin{aligned}\mathcal{R}(f) &= \mathbb{E}\Big[\big(\Omega_t - m_t(X_t) + m_t(X_t) - f(X_t)\big)^2\Big] \\ &= \mathbb{E}\Big[\big(\Omega_t - m_t(X_t)\big)^2\Big] + \mathbb{E}\Big[\big(m_t(X_t) - f(X_t)\big)^2\Big] \\ &\quad + 2\mathbb{E}\big[\big(m_t(X_t) - f(X_t)\big)\mathbb{E}[\Omega_t - m_t(X_t) \mid X_t]\big] .\end{aligned}$$

Since $\mathbb{E}[\Omega_t - m_t(X_t) \mid X_t] = 0$, the cross term vanishes. Hence,

$$\mathcal{R}(f) = \mathcal{R}(m_t) + \mathbb{E}\Big[\big(f(X_t) - m_t(X_t)\big)^2\Big] . \tag{7}$$

The second term is nonnegative and vanishes if and only if $f(X_t) = m_t(X_t)$ almost surely. Therefore, $f^*(X_t) = m_t(X_t) = \mathbb{E}[\Omega_t \mid X_t]$ almost surely, which proves Proposition 4.1. □

At the optimum, $\mathcal{R}(f^*) = \mathbb{E}[\mathrm{Var}(\Omega_t \mid X_t)]$. Equation (7) therefore separates the minimum prediction risk given X_t from the excess risk of a predictor relative to the conditional mean.

C.2. Frontier Progress and Additional Reasoning Gains

We make explicit the relation between the remaining frontier progress and the reasoning gain realized by additional latent computation. Fix a latent depth t and write $M_t := \max_{1\le j\le t} G_j$, so that $\Omega_t = M_K - M_t$.

Frontier realization. For an integer $a \in \{0, \dots, K-t\}$, define the frontier advancement realized within a additional latent steps as

$$P_t(a) := M_{t+a} - M_t = \max_{1\le j\le t+a} G_j - \max_{1\le j\le t} G_j. \tag{8}$$

Lemma C.1 (Frontier realization over the remaining horizon). *For any integers $0 \le a \le b \le K - t$,*

$$0 = P_t(0) \le P_t(a) \le P_t(b) \le \Omega_t, \qquad P_t(K-t) = \Omega_t.$$

Proof. The index sets are nested as

$$\{1, \dots, t\} \subseteq \{1, \dots, t+a\} \subseteq \{1, \dots, t+b\} \subseteq \{1, \dots, K\}.$$

Hence, the corresponding frontier values are nondecreasing, which gives $0 = P_t(0) \le P_t(a) \le P_t(b) \le \Omega_t$. Finally, setting $a = K - t$ in Eq. (8) gives

$$P_t(K-t) = M_K - M_t = \Omega_t,$$

which proves the result. □

Signed reasoning gain. Frontier advancement does not by itself determine whether the reasoning value after continuation improves over its current value. We therefore define the signed reasoning gain from a additional latent steps as $\Delta G_t(a) := G_{t+a} - G_t$. Let $d_t := M_t - G_t \ge 0$ denote the gap between the current reasoning value and the best frontier reached so far.

Lemma C.2 (Decomposition of additional reasoning gain). *For any $a \in \{0, \dots, K-t\}$,*

$$\Delta G_t(a) = P_t(a) + d_t - d_{t+a}. \tag{9}$$

Proof. By definition, $G_t = M_t - d_t$ and $G_{t+a} = M_{t+a} - d_{t+a}$. Therefore,

$$\begin{aligned}\Delta G_t(a) &= G_{t+a} - G_t \\ &= \big(M_{t+a} - d_{t+a}\big) - \big(M_t - d_t\big) \\ &= P_t(a) + d_t - d_{t+a},\end{aligned}$$

which proves Eq. (9). □

In particular, over the full remaining horizon,

$$\Delta G_t(K-t) = \Omega_t + d_t - d_K.$$

Thus, remaining frontier progress and the endpoint gain need not coincide: whether the final reasoning value improves also depends on how the trajectory moves relative to the frontier already attained.

Moreover, if $\Omega_t = 0$, Lemma C.1 gives $P_t(a) = 0$ for every $a \in \{0, \dots, K-t\}$, and Eq. (9) reduces to

$$\Delta G_t(a) = d_t - d_{t+a}.$$

In this case, additional computation cannot establish a new frontier; its signed gain is determined entirely by how the subsequent state moves relative to the existing frontier.

Expected continuation return. For ω in the support of Ω_t, define

$$V_t(a, \omega) := \mathbb{E}[\Delta G_t(a) \mid \Omega_t = \omega]. \tag{10}$$

At a fixed depth t, we suppress the subscript and write $V(a, \omega)$. By construction, $V_t(0, \omega) = 0$. Unlike the frontier realization $P_t(a)$, both $\Delta G_t(a)$ and $V_t(a, \omega)$ are signed quantities and therefore capture whether additional latent computation improves or degrades the reasoning value.

C.3. Monotone Allocation by Remaining Progress

This subsection proves Theorem 4.2, which characterizes how optimal additional computation varies with remaining progress.

Discrete allocation. Fix a latent depth t and let $A := K - t$. For trajectories $i = 1, \dots, n$ with remaining progress ω_i, define the feasible allocation set

$$\mathcal{A}_B := \left\{ a \in \{0, \dots, A\}^n : \sum_{i=1}^{n} a_i \le B \right\}.$$

Assume that $V(a, \omega)$ has increasing differences in (a, ω). Thus, for integers $0 \le a \le a' \le A$ and $\omega \le \omega'$,

$$V(a', \omega') - V(a, \omega') \ge V(a', \omega) - V(a, \omega).$$

Proof of Theorem 4.2. Since $\mathcal{A}_B$ is finite and nonempty, an optimal allocation exists. Let $\bar{a} \in \mathcal{A}_B$ be an optimal allocation. Relabel the trajectories so that $\omega_1 \ge \cdots \ge \omega_n$. If there exist $i < j$ with $\omega_i > \omega_j$ but $\bar{a}_i < \bar{a}_j$, exchange $\bar{a}_i$ and $\bar{a}_j$. The exchange preserves feasibility. By increasing differences,

$$V(\bar{a}_j, \omega_i) + V(\bar{a}_i, \omega_j) \ge V(\bar{a}_i, \omega_i) + V(\bar{a}_j, \omega_j).$$

Hence, removing such an inversion does not decrease the objective. Applying these exchanges until no inversion remains yields an optimal allocation a^* satisfying

$$\omega_i > \omega_j \quad \Longrightarrow \quad a_i^* \ge a_j^*,$$

which proves Theorem 4.2. □

C.4. Marginal Returns and Remaining Progress

We now make explicit the marginal-return interpretation of the increasing-differences condition in Theorem 4.2. Fix a latent depth t. For integers $0 \le a < a' \le K - t$, define the incremental return from extending the additional latent budget from a to a' as

$$R_t(a, a') := \Delta G_t(a') - \Delta G_t(a) = G_{t+a'} - G_{t+a}.$$

Lemma C.3 (Decomposition of incremental continuation return). *For any integers $0 \le a < a' \le K - t$,*

$$R_t(a, a') = P_t(a') - P_t(a) + d_{t+a} - d_{t+a'}. \tag{11}$$

Moreover, for ω in the support of Ω_t,

$$V_t(a', \omega) - V_t(a, \omega) = \mathbb{E}[R_t(a, a') \mid \Omega_t = \omega]. \tag{12}$$

Proof. Subtracting the decomposition in Eq. (9) at budgets a and a' gives

$$\begin{aligned} R_t(a, a') &= \Delta G_t(a') - \Delta G_t(a) \\ &= \big(P_t(a') + d_t - d_{t+a'}\big) - \big(P_t(a) + d_t - d_{t+a}\big) \\ &= P_t(a') - P_t(a) + d_{t+a} - d_{t+a'}, \end{aligned}$$

which proves Eq. (11). Taking conditional expectation given $\Omega_t = \omega$ and using the definition of V_t in Eq. (10) gives Eq. (12). □

Lemma C.3 gives a trajectory-level interpretation of the return structure used in Theorem 4.2. For the same additional-compute interval $a \to a'$, the incremental return consists of the new frontier advancement over that interval together with the change in the state's distance from the attained frontier. Consequently, at a fixed latent depth t, the increasing-differences condition takes the form

$$\omega' > \omega \quad \Longrightarrow \quad \mathbb{E}[R_t(a, a') \mid \Omega_t = \omega'] \geq \mathbb{E}[R_t(a, a') \mid \Omega_t = \omega]\,.$$

Thus, the condition states that the same additional latent-compute interval yields no smaller expected signed incremental return when more constructive progress remains.

Corollary C.4 (Return ordering by remaining progress). *Suppose that $V_t(a, \omega)$ has increasing differences in (a, ω). For any feasible additional latent budget $a \in \{0, \ldots, K - t\}$ and any $\omega' > \omega$,*

$$V_t(a, \omega') \geq V_t(a, \omega).$$

Proof. By increasing differences, comparing the additional budgets 0 and a gives

$$V_t(a, \omega') - V_t(0, \omega') \geq V_t(a, \omega) - V_t(0, \omega).$$

Since $V_t(0, \omega') = V_t(0, \omega) = 0$, it follows that

$$V_t(a, \omega') \geq V_t(a, \omega),$$

which proves the result. □

Hence, under a common additional latent budget, larger remaining progress corresponds to no smaller expected signed reasoning gain.

D. Additional Experiments

D.1. Marginal Continuation Returns

Appendix C.4 characterizes increasing differences through the signed marginal return over a common interval of additional latent computation. Here, we examine this return ordering beyond the one-step comparison in Table 4, using the same lower Ω_t and higher Ω_t groups and signed future gains. For each latent depth, we compare the marginal return of the two groups over subsequent adjacent compute intervals.

Table 5: **Marginal return ordering over continued latent CoT.** Each entry reports the mean signed marginal-return gap between higher- and lower-remaining-progress trajectories over the subsequent additional compute intervals, together with the fraction of intervals with a nonnegative gap.

Backbone	Method	GSM8K		SVAMP		MultiArith	
		Mean return gap (pp)	Nonnegative intervals (%)	Mean return gap (pp)	Nonnegative intervals (%)	Mean return gap (pp)	Nonnegative intervals (%)
GPT-2	CODI	+13.1	100	+4.1	80	+4.7	60
	SIM-CoT	+14.7	100	+4.2	60	+15.2	100
Llama-1B	CODI	+15.1	90	+2.9	80	+5.2	70
	SIM-CoT	+13.6	100	+2.7	80	+6.2	60
GPT-2	Coconut	+14.5	100	+6.5	100	+17.9	80
	SIM-CoT Coconut	+20.9	100	+6.9	100	+4.4	70

As shown in Table 5, the mean return gap is positive in all settings, and the ordering remains nonnegative for most intervals in every setting. The pattern is most consistent on GSM8K, while SVAMP

and MultiArith show greater variation across intervals. These results show that higher remaining progress is generally associated with larger signed marginal continuation returns over subsequent latent CoT.

D.2. Additional Results for Remaining Progress and Computation Returns

Tables 6 and 7 report the corresponding analyses for SVAMP and MultiArith, both of which exhibit the same overall pattern. Higher remaining constructive progress is associated with larger future gains from continued latent CoT, while larger decreases in Ω_t correspond to larger reductions in subsequent reasoning gain. The relation between remaining progress and future reasoning gains identified in the main analysis is therefore consistent across benchmarks.

Table 6: **Reasoning gains across remaining constructive progress.** We report SVAMP. Ω_3 means the remaining constructive progress after 3 latent steps. Panel **(a)** reports task accuracy at Steps 3 and 4 under lower and higher Ω_3. Panel **(b)** reports future reasoning gain, measured by the change in $\hat{G}_t$ from Step 3 to Steps 4 and 5, under lower and higher Ω_3, with subscripts indicating the difference relative to lower Ω_3. Panel **(c)** reports the change in future reasoning gain as Ω_t decreases, with negative values indicating a decrease.

		(a) Accuracy (%)				(b) Future gain				(c) Future gain change		
		Lower Ω_3		Higher Ω_3		Lower Ω_3		Higher Ω_3		Decrease in Ω_t		
Backbone / Method		Step 3	Step 4	Step 3	Step 4	Step 4	Step 5	Step 4	Step 5	Lower	Middle	Higher
CODI family												
GPT-2	CODI	45.2	44.1 $_{-1.1}$	51.7	73.3 $_{+21.6}$	-0.9	$+0.5$	$+12.3_{+13.2}$	$+21.2_{+20.7}$	-1.9	-9.3	-39.4
	SIM-CoT	37.5	36.4 $_{-1.1}$	26.5	70.6 $_{+44.1}$	$+0.3$	-0.6	$+31.7_{+31.4}$	$+40.8_{+41.4}$	-2.1	-6.1	-38.5
Llama-1B	CODI	53.3	53.3 $_{0.0}$	42.6	48.9 $_{+6.3}$	$+1.0$	$+1.6$	$+2.3_{+1.3}$	$+15.4_{+13.8}$	-0.6	-3.5	-22.9
	SIM-CoT	61.8	61.8 $_{0.0}$	73.0	75.7 $_{+2.7}$	-0.4	-0.7	$+6.5_{+6.9}$	$+7.8_{+8.5}$	-1.9	-7.2	-26.2
Coconut family												
GPT-2	Coconut	27.0	28.8 $_{+1.8}$	46.3	65.3 $_{+19.0}$	$+0.7$	$+0.1$	$+19.2_{+18.5}$	$+26.7_{+26.6}$	$+0.9$	-0.1	-23.5
	SIM-CoT Coconut	42.4	42.4 $_{0.0}$	41.9	56.5 $_{+14.6}$	$+1.3$	$+2.4$	$+14.3_{+13.0}$	$+20.2_{+17.8}$	$+1.0$	-6.8	-52.9

Table 7: **Reasoning gains across remaining constructive progress.** We report MultiArith. Ω_3 means the remaining constructive progress after 3 latent steps. Panel **(a)** reports task accuracy at Steps 3 and 4 under lower and higher Ω_3. Panel **(b)** reports future reasoning gain, measured by the change in $\hat{G}_t$ from Step 3 to Steps 4 and 5, under lower and higher Ω_3, with subscripts indicating the difference relative to lower Ω_3. Panel **(c)** reports the change in future reasoning gain as Ω_t decreases, with negative values indicating a decrease.

		(a) Accuracy (%)				(b) Future gain				(c) Future gain change		
		Lower Ω_3		Higher Ω_3		Lower Ω_3		Higher Ω_3		Decrease in Ω_t		
Backbone / Method		Step 3	Step 4	Step 3	Step 4	Step 4	Step 5	Step 4	Step 5	Lower	Middle	Higher
CODI family												
GPT-2	CODI	87.5	87.5 $_{0.0}$	61.5	76.9 $_{+15.4}$	0.0	$+0.4$	$+24.3_{+24.3}$	$+41.8_{+41.4}$	-2.2	-28.8	-64.7
	SIM-CoT	87.2	87.2 $_{0.0}$	2.7	94.6 $_{+91.9}$	$+1.0$	$+1.1$	$+83.3_{+82.3}$	$+86.7_{+85.6}$	-0.3	-7.9	-60.2
Llama-1B	CODI	85.7	85.7 $_{0.0}$	75.0	75.0 $_{0.0}$	0.0	-0.9	$+3.1_{+3.1}$	$+32.8_{+33.7}$	-1.3	-6.4	-27.2
	SIM-CoT	84.6	84.6 $_{0.0}$	85.7	85.7 $_{0.0}$	-0.5	-0.7	$+2.2_{+2.7}$	$+0.9_{+1.6}$	-2.5	-9.0	-42.0
Coconut family												
GPT-2	Coconut	50.0	51.8 $_{+1.8}$	10.6	36.2 $_{+25.6}$	$+9.7$	$+13.2$	$+19.4_{+9.7}$	$+77.7_{+64.5}$	$+7.6$	-6.0	-56.3
	SIM-CoT Coconut	92.3	88.5 $_{-3.8}$	62.5	100.0 $_{+37.5}$	0.0	$+2.2$	$+35.7_{+35.7}$	$+37.5_{+35.3}$	$+20.2$	-22.2	-84.4

D.3. Continuation Signals for Adaptive Latent Inference

Here we examine how the choice of continuation signal affects adaptive inference under the same controller. Specifically, we keep the continuation rule and final prediction procedure fixed, and substitute the online signal used to allocate the remaining latent budget. We consider remaining progress, answer stability, prediction confidence, and state convergence, with average latent computation matched across criteria.

As shown in Table 8, remaining progress yields higher accuracy than state convergence across all six settings, while answer stability and prediction confidence achieve similar or higher accuracy in several settings. Relative to signal free allocation under the same controller, remaining progress improves average accuracy by 0.98 percentage points. Remaining progress therefore provides a

direct quantity for future continuation opportunity and a useful signal for adaptive allocation, while simpler online signals can yield similar continuation performance in some settings.

Table 8: **Continuation signals under matched latent computation.** All criteria use the same adaptive controller and differ only in the online signal used to allocate the remaining latent budget. We report final task accuracy under matched average latent computation, with subscripts indicating the difference relative to remaining progress.

			Accuracy (%) by continuation signal				Compute
Dataset	Backbone	Method	Progress	Stability	Confidence	Convergence	Avg. steps
GSM8K	*GPT-2*	CODI	41.1	$40.5_{-0.6}$	$40.2_{-0.9}$	$40.0_{-1.1}$	4.37
		SIM-CoT	39.2	$37.9_{-1.3}$	$36.8_{-2.4}$	$37.5_{-1.7}$	4.28
	Llama-1B	SIM-CoT	54.7	$53.3_{-1.4}$	$54.2_{-0.5}$	$53.4_{-1.3}$	4.72
SVAMP	*GPT-2*	SIM-CoT	40.2	$40.7_{+0.5}$	$40.7_{+0.5}$	$40.1_{-0.1}$	4.68
	Llama-1B	CODI	60.7	$60.9_{+0.2}$	$61.0_{+0.3}$	$60.5_{-0.2}$	4.85
MultiArith	*GPT-2*	SIM-CoT	89.4	$88.9_{-0.5}$	$91.7_{+2.3}$	$88.3_{-1.1}$	4.61

Table 9: **Future reasoning gains under estimated remaining progress.** $f_\phi(X_3)$ denotes the estimated remaining progress after 3 latent steps. Panel **(a)** reports task accuracy at Steps 3 and 4 under lower and higher $f_\phi(X_3)$. Panel **(b)** reports future reasoning gain, measured by the change in $\hat{G}_t$ from Step 3 to Steps 4 and 5, under lower and higher $f_\phi(X_3)$, with subscripts indicating the difference relative to lower $f_\phi(X_3)$.

			(a) Accuracy (%)				(b) Future gain			
			Lower $f_\phi(X_3)$		Higher $f_\phi(X_3)$		Lower $f_\phi(X_3)$		Higher $f_\phi(X_3)$	
Dataset	Backbone	Method	Step 3	Step 4	Step 3	Step 4	Step 4	Step 5	Step 4	Step 5
GSM8K	*GPT-2*	CODI	46.4	47.6 $_{+1.2}$	27.1	24.6 $_{-2.5}$	$+0.2$	$+0.3$	$-0.6_{-0.8}$	$+13.0_{+12.7}$
		SIM-CoT	43.6	44.8 $_{+1.2}$	21.6	38.3 $_{+16.7}$	$+0.1$	-0.1	$+16.4_{+16.3}$	$+29.2_{+29.3}$
	Llama-1B	CODI	61.1	61.4 $_{+0.3}$	36.2	37.1 $_{+0.9}$	-0.3	$+0.9$	$0.0_{+0.3}$	$+16.1_{+15.2}$
		SIM-CoT	61.1	61.6 $_{+0.5}$	39.4	47.8 $_{+8.4}$	$+0.2$	$+0.1$	$+7.3_{+7.1}$	$+17.5_{+17.4}$
SVAMP	*GPT-2*	CODI	53.0	53.6 $_{+0.6}$	31.2	32.1 $_{+0.9}$	-0.1	-0.4	$+0.6_{+0.7}$	$+0.1_{+0.5}$
		SIM-CoT	53.3	53.0 $_{-0.3}$	23.7	27.0 $_{+3.3}$	-1.0	-1.7	$+2.3_{+3.3}$	$+2.6_{+4.3}$
	Llama-1B	CODI	78.4	78.4 $_{0.0}$	36.6	37.5 $_{+0.9}$	-0.3	-0.4	$+0.2_{+0.5}$	$+0.8_{+1.2}$
		SIM-CoT	76.0	76.0 $_{0.0}$	49.5	49.2 $_{-0.3}$	0.0	-0.5	$+0.3_{+0.3}$	$-0.5_{0.0}$
MultiArith	*GPT-2*	CODI	96.7	96.7 $_{0.0}$	86.7	88.3 $_{+1.6}$	-0.3	$+0.1$	$+3.7_{+4.0}$	$+5.3_{+5.2}$
		SIM-CoT	77.8	84.4 $_{+6.6}$	55.6	97.8 $_{+42.2}$	$+7.5$	$+9.0$	$+45.5_{+38.0}$	$+46.4_{+37.4}$
	Llama-1B	CODI	100.0	100.0 $_{0.0}$	93.3	93.3 $_{0.0}$	-1.0	$+0.1$	$+0.3_{+1.3}$	$+3.0_{+2.9}$
		SIM-CoT	98.3	98.3 $_{0.0}$	95.0	96.7 $_{+1.7}$	$+0.1$	$+0.2$	$-0.5_{-0.6}$	$-1.2_{-1.4}$

D.4. Estimated Remaining Progress and Future Reasoning Gains

Section 5.3 shows that more remaining constructive progress corresponds to larger reasoning gains from continued latent CoT. Here, we examine whether this relation is retained when remaining progress is estimated from the current latent trajectory. Using out-of-sample estimates $f_\phi(X_3)$, we form lower progress and higher progress groups and evaluate the same reasoning gain and task accuracy outcomes.

As shown in Table 9(b), the higher estimate group obtains larger future reasoning gains at Step 4 and Step 5, with average gaps of 5.9 and 10.4 percentage points, respectively. Table 9(a) shows the same pattern in task accuracy, with the higher estimate group improving more from Step 3 to Step 4. In several settings, we observe that the higher estimate group starts from lower accuracy at Step 3 yet gains more after continuation, consistent with remaining progress characterizing future reasoning opportunity. Together, these results show that estimated remaining progress preserves the future gain separation identified in Section 5.3, indicating that the current latent trajectory carries information relevant to future continuation opportunity.