GEOMETRIC DECOMPOSITION OF REASONING LOGIC IN LATENT SPACES

Anonymous authors
Paper under double-blind review

ABSTRACT

Latent chain-of-thought (CoT) reasoning methods train LLMs to reason with continuous latent states, reducing language CoT tokens to a small fixed budget. However, many latent reasoning methods rely solely on *outcome-rewarding objectives* which mainly supervise whether the latent chain reaches an answer state while giving little guidance to the reasoning space itself. In this paper, we study latent reasoning from the perspective of *global reasoning geometry*. Based on compact singular value decomposition, we introduce a **geometric decomposition** of the latent reasoning chain, separating the reasoning space into the principal direction and the orthogonal subspace. We empirically connect the two geometric objects with different aspects of reasoning logic: the principal direction is associated with the core logical form, while the orthogonal subspace is associated with semantic transitional logic in reasoning. Furthermore, we propose a **geometric decomposition-based auxiliary objective** for distillation-based latent reasoning. It aligns the latent reasoning chain with the language CoT's principal direction, and uniformly expands the latent reasoning chain in the orthogonal subspace. We theoretically and empirically show that outcome-rewarding models can lead to ill-conditioned reasoning as intermediate states converge to an "ideal exit" in the reasoning space, which harms the reasoning ability on difficult problems. In contrast, our geometric decomposition-based objective alleviates the latent chain convergence, expands the orthogonal subspace and provides richer reasoning information for answer decoding. Experiments on Llama 3 and Qwen 3 models show 12.7%–35.5% average Pass@1 relative improvements over outcome-rewarding methods on difficult mathematical reasoning settings, validating its effectiveness and geometric interpretability[1].

1 INTRODUCTION

Chain-of-thought (CoT) reasoning (Wei et al., 2022) has become one of the fundamental technologies for large language models (LLMs). A recent promising direction, *latent reasoning* (Chen et al., 2025; Li et al., 2025b), aims to internalize such reasoning steps into continuous hidden states, resembling the cognitive mode often described as "inner speech" (Alderson-Day & Fernyhough, 2015). A specific line of work totally removes reasoning text at inference time, and fine-tunes LLMs to reason with latent states supervised by both language CoT and answers (Deng et al., 2023; Hao et al., 2025; Shen et al., 2025; Wei et al., 2026; Kuzina et al., 2026; Shen et al., 2026; Zhao et al., 2026); we refer to this setting as latent supervised fine-tuning (latent SFT). This reduces the number of CoT tokens to a small fixed budget, achieving much higher token efficiency than language CoT.

However, a widespread concern is that many latent SFT methods work mainly on short and direct reasoning benchmarks, but struggle on more difficult reasoning problems. We argue that an important reason is their reliance solely on *outcome-rewarding objectives* (Thatikonda et al., 2025; Song et al., 2025), such as latent language modeling: they mainly supervise whether the latent chain reaches an answer state, but give little direct guidance to the reasoning space itself. As shown in Figure 1(a), we divide the latent space into (1) the *reasoning space* with the latent states of CoT steps, and (2) the *answer space* with the latent states of answer anchors, *i.e.*, tokens immediately before answers. Figure 1(e) visualizes real latent spaces of two mathematical reasoning subsets, in which the reasoning and answer spaces differ clearly in location and shape. We show both empirically and theoretically

[1]Code is available at https://anonymous.4open.science/r/gdF451.

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.

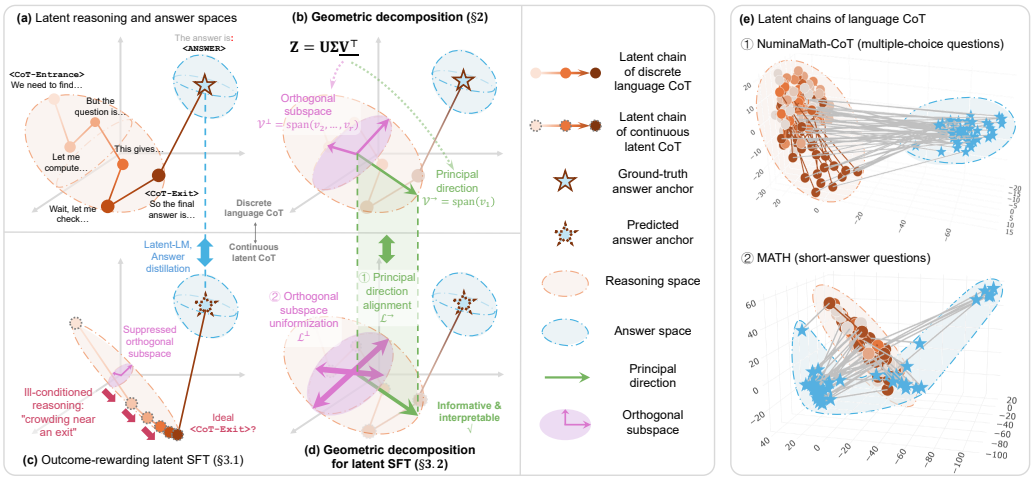


Figure 1: Overview of this paper. **(a)** The latent space contains the reasoning space and the answer space. Each point in the reasoning space denotes a sentence-level reasoning step, and the answer space contains the latent states of answer anchors. **(b)** Our geometric decomposition on the latent reasoning chain: the principal direction $\mathcal{V}^{\rightarrow}$ and the orthogonal subspace $\mathcal{V}^{\perp}$. **(c)** Outcome-rewarding latent SFT, including latent language modeling and answer distillation (Shen et al., 2025), induces ill-conditioned reasoning where intermediate states crowd near an "ideal exit". They also lead to a collapsed reasoning space as the orthogonal subspace is heavily suppressed. **(d)** Our geometric decomposition-based objective consists of ❶ principal direction alignment and ❷ orthogonal subspace uniformization. It alleviates the crowding phenomenon and expands the orthogonal subspace. **(e)** Reasoning and answer spaces (first 50 samples from NuminaMath-CoT and MATH; Llama 3.1-8B-Instruct). The answer space of MATH is more dispersed, as the answers of NuminaMath-CoT are mainly option letters while those of MATH are more diverse numbers and expressions.

that **outcome-rewarding objectives can make the latent chain converge toward a single direction**: intermediate hidden states crowd near an "ideal exit", as illustrated in Figure 1(c). This exit could be more of an uninformative conclusion state than a guaranteed mapping to the correct anchor in the answer space. Worse still, the reasoning chain is collapsed toward the direction rather than preserving a rich reasoning trajectory; this could harm the reasoning ability on difficult problems. See §D for more related work.

Our aim in this paper is to find *whether latent reasoning can be improved from a perspective of latent reasoning geometry.* Instead of the *local* geometric objects like step-wise distances (Sun et al., 2026) and curvature (Zhou et al., 2026), we study how *global* geometry reflects reasoning logic. Based on compact singular value decomposition (SVD), we propose a **geometric decomposition** of a latent reasoning chain into the **principal direction** and its **orthogonal** (complement) **subspace**, as shown in Figure 1(b). Our empirical analysis suggests that **the principal direction is associated with the logical form**: it stays stable across shallow perturbations and changes when the core logical form changes. In contrast, **the orthogonal subspace is associated with semantic transitional logic** beyond the main thread. These findings make global reasoning geometry more interpretable by connecting the principal and orthogonal components to different aspects of reasoning logic.

To ensure realizability, we further introduce **a geometric decomposition-based auxiliary objective for distillation-based latent SFT**, as shown in Figure 1(d). The objective has two aims: it aligns aligns the principal direction of the latent reasoning chain (the student) with that of the language CoT (the teacher), and uniformly expands the student chain in the teacher's orthogonal subspace. This yields a geometry-aware, process-rewarding supervision instead of only outcome rewards. Extensive experiments show that it alleviates the crowding phenomenon, expands the orthogonal subspace and provides richer reasoning information for answer decoding. Our contributions are as follows:

- We analyze the reasoning geometry by introducing a geometric decomposition of latent reasoning chains. We empirically show that the principal direction is linked to the main logical form (§2.1), while the orthogonal subspace is associated with semantic transitions in reasoning (§2.2).

- We geometrically revisit outcome-rewarding latent SFT objectives, and theoretically show that they can cause latent chain convergence and spectral collapse during training (§3.1).
- We propose a geometry-aware auxiliary loss, $\mathcal{L}_{\mathrm{GD}}$, which aligns the principal direction and uniformizes the orthogonal spectrum (§3.2). It mitigates the latent chain convergence and enriches the information carried by latent chains.
- On Llama 3 and Qwen 3 models, our method improves Pass@1 by relatively 12.7%–35.5% over outcome-rewarding latent SFT methods on difficult mathematical reasoning settings, demonstrating both the effectiveness and interpretability of geometric decomposition.

2 Geometric Decomposition of the Latent Reasoning Chain

Preliminary. Latent reasoning replaces the discrete language CoT with a sequence of continuous latent states. Starting from the question, the model performs fixed $K \geq 2$ latent forward passes and obtains an ordered sequence of latent reasoning states $\mathcal{Z} = (z_1, \ldots, z_K)$. We call this a *latent chain*. Let its row matrix form be $\mathbf{Z} = [z_1^\top; \ldots; z_K^\top] \in \mathbb{R}^{K \times d}$. We also define the latent chain for original language CoT following Li et al. (2025a); Wen et al. (2026); Alvarez & Baheri (2026): we split the CoT text into sentences, and take the mean-pooled hidden states of tokens in the k-th sentence as z_k. This treats natural language sentences as coherent logical units for the following empirical studies.

Geometric Decomposition. Prior work suggests that the link between reasoning logic and representation geometry, including the intrinsic dimension, curvature and spectrum (Lee et al., 2025; Zhou et al., 2026; Li et al., 2025c), is a central question for understanding and improving reasoning. Motivated by this view, we use the spectrum of a latent chain to analyze the geometry of latent reasoning, and introduce the following decomposition.

Main takeaway: geometric decomposition of the latent reasoning chain

By compact SVD on the centered row matrix of the latent chain $\mathbf{Z} - \mathbf{1}_K \bar{z}^\top = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top = \sum_{i=1}^{r} \sigma_i u_i v_i^\top$, where $\bar{z} = \sum_{k=1}^{K} z_k / K$, $\mathbf{1}_K$ is the all-one vector, r is the rank of the centered matrix, and $\sigma_1 \geq \cdots \geq \sigma_r > 0$ are the singular values, we decompose the reasoning space $\mathcal{V} \subset \mathbb{R}^d$ into the **principal direction** $\mathcal{V}^{\rightarrow}$ and its **orthogonal (complement) subspace** $\mathcal{V}^{\perp}$:

$$\mathcal{V} = \mathcal{V}^{\rightarrow} \oplus \mathcal{V}^{\perp}, \quad \mathcal{V}^{\rightarrow} := \mathrm{span}\{v_1\}, \quad \mathcal{V}^{\perp} := \mathrm{span}\{v_2, \ldots, v_r\} = \{a \in \mathcal{V} : v_1^\top a = 0\} \quad (1)$$

Next, we empirically analyze the connection between geometric decomposition and reasoning logic.

2.1 The Principal Direction

The principal direction $\mathcal{V}^{\rightarrow}$ captures the direction along which the latent chain moves the most in the reasoning space. We discuss some relative details in §E.1. We conjecture that the principal direction represents the "**main reasoning thread**", which is the overall reasoning logic that can be shared by a group of similar problems, and it is expected to be less sensitive to shallow, logic-unrelated perturbations such as scenarios, numbers and entities. At the same time, different reasoning threads can induce different principal directions, even when they solve the same set of questions.

We conduct a mathematical reasoning experiment to demonstrate the connection above. For each question sample, we first perturb its shallow semantics (scenarios, numbers and entities) and construct 10 questions that *have different answers but can be solved with the same deduction process*. We also design two different reasoning templates that can both solve these 10 questions correctly. An example is shown in Table 1. We construct three question groups from GSM8K (Cobbe et al., 2021), and mix them with another 100 unrelated questions from the dataset. Finally, we apply principal component analysis (PCA) to the embeddings of all samples jointly, and visualize the reasoning subspaces of the selected question groups. To compensate for possible bias introduced by dimensionality reduction, we also report the *mean resultant length* (MRL) (Rayleigh, 1880), which measures how strongly the principal vectors concentrate around a certain direction in the original high-dimensional space.

Figure 2(a) shows the latent chains and their principal directions under reasoning template 1. Within each question group, the principal directions are highly parallel, while the control group is scattered

Table 1: Example of a question group used in our empirical experiments. Only highlighted text in the reasoning templates change with different questions. See §B.1 for other groups.

Description	Question #1	Question #2	Reasoning template 1	Reasoning template 2
Question group #1 *Total sum of the first, second, and last (0) parts*	*"Billy sells CDs. He has 1,250 customers on Wednesday. His first 500 customers buy 3 CDs each. His next 450 customers buy 8 CDs each. His last 300 customers do not buy any CDs. How many CDs does Billy sell on Wednesday?"* ***Answer: 5100***	*"At an arcade counter, the first 125,000 kids win 8 tokens each. The next 175,000 kids win 3 tokens each. The last 120,000 kids win none. How many tokens are handed out in total?"* ***Answer: 1525000***	*"The first part contributes 500 * 3 = 1,500. The second part contributes 450 * 8 = 3,600. The remaining part contributes 0, so the total is 1,500 + 3,600 + 0 = 5,100."*	*"If all 1,250 cases contributed 3 each, the total would be 1,250 * 3 = 3,750. The second group differs from that baseline by 8 - 3 per case, so the adjustment is 450 * (8 - 3) = 2,250. The last group gives none, so remove 300 * 3 = 900, leaving 3,750 + 2,250 - 900 = 5,100."*

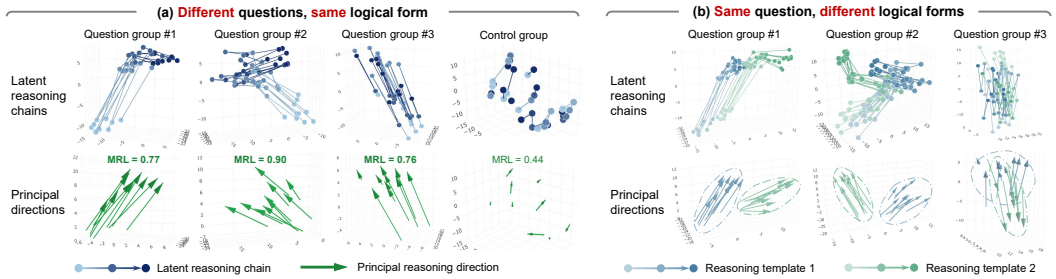


Figure 2: Connection between the principal direction and reasoning logic (GSM8K samples; Llama 3.1-8B-Instruct). **(a)** Latent chains and principal directions of question groups with different questions but the same reasoning template. MRL stands for mean resultant length. The control group contains 10 random samples from 100 jointly-PCAed questions. **(b)** Two groups with the same questions but different reasoning templates. Two principal-direction clusters point toward different directions whose angular separation varies across question groups. See §B.1 for MATH samples.

in the latent space. The MRL of each question group is also much higher than that of the control group. Figure 2(b) further adds the latent chains under reasoning template 2. Although the two sets of chains correspond to the same questions and answers, the two principal-direction clusters are clearly separated and point toward different directions.

These results empirically connect the principal direction with *logical form*, which we define here as the abstractable main reasoning logic shared across different instances. A classic example is Aristotle's syllogism: "*If A is predicated of all B, and B of all C, then A must be predicated of all C*", whose validity does not depend on the instantiation of A, B and C. This connection bridges the geometry and logic of our conjecture, and gives the proposed geometric decomposition a more interpretable role in latent reasoning.

2.2 The Orthogonal Subspace

The orthogonal subspace $\mathcal{V}^{\perp}$ is the orthogonal complement of $\mathcal{V}^{\rightarrow}$. While $\mathcal{V}^{\rightarrow}$ captures the core logical form of a latent chain, $\mathcal{V}^{\perp}$ can be viewed as carrying residual reasoning logic beyond it. It may include linguistic transitions between reasoning steps such as scenario grounding, strategy shifts and self-correction. Prior work suggests that such reflective ability is important for open-ended reasoning (Lee et al., 2026; Cheng et al., 2026). Despite not being included in the logical form, the semantic logic in the orthogonal subspace may also affect how the reasoning process unfolds.

To investigate which semantic logic contributes to the displacement in $\mathcal{V}^{\perp}$, we embed correct language CoTs generated by different models on MATH (Hendrycks et al., 2021), and measure the orthogonal displacement of each local reasoning step. Inspired by Dong et al. (2026); Wen et al. (2026); Lee et al. (2026), we manually group sentence-level logical pivots into five types:

- **Clarification**: steps that restate the goal or ground the current condition in the problem statement, including sentences with "*I need to*", "*the question is*", and "*looking at*".
- **Inference**: steps that derive a new intermediate consequence or carry out calculation, including sentences with "*now*", "*let me calculate*", "*so we have*", and "*this means*".
- **Transition**: steps that redirect the reasoning path, including sentences with "*but*" and "*however*".

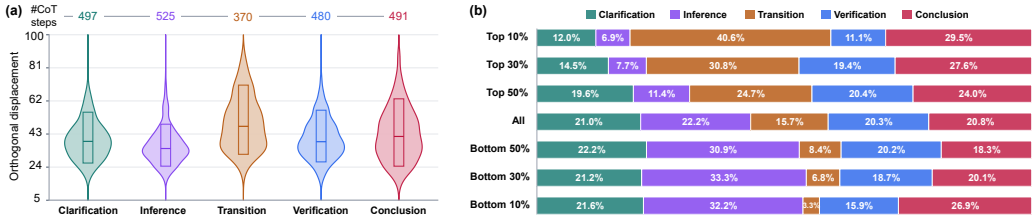


Figure 3: Different semantic logic types have different strengths in the orthogonal subspace (MATH; Llama 3.1-8B-Instruct). **(a)** Distribution of orthogonal displacement grouped by logical pivots. **(b)** Proportion of logical pivots among reasoning steps with different levels of orthogonal displacement. See §B.2 for more results.

- **Verification**: steps that check, verify, or correct a previous derivation, including sentences with "*check*", "*double-check*", "*verify*", and "*make sure*".
- **Conclusion**: the final steps that produce the answer, *e.g.*, "*I think that's it*", "*the answer should be*", and "*now produce final answer*".

Figure 3 reports the statistical results. It shows that different semantic logic types show different magnitudes of displacement in $\mathcal{V}^{\perp}$. For example, **Transition**, which often redirects to another reasoning branch, has the strongest orthogonal displacement; **Inference**, which is more tightly tied to the logical form, has the weakest displacement. This suggests **a positive association between the orthogonal subspace and semantic transitions in reasoning**. Although it is hard to assign each logic type to a specific orthogonal direction, this result indicates that orthogonal exploration in latent chains should not be overlooked for complex reasoning.

3 Geometric Decomposition as Latent Supervision

To show the practical value of geometric decomposition in latent reasoning training, we adapt it to the self-distillation-based latent SFT framework (Shen et al., 2025; Wei et al., 2026) to make it more informative and geometrically interpretable.

3.1 Outcome-rewarding Latent SFT

Traditional latent SFT trains the model by the language modeling loss on the final answer, with the latent states inserted into the forward pass before answer decoding. Given a question x and ground-truth answer tokens $a = (a_1, \ldots, a_T)$, Latent-LM loss is formalized as

$$\mathcal{L}_{\text{Latent-LM}}(x, \mathcal{Z}, a; \Theta) := -\frac{1}{T}\sum_{t=1}^{T} \log p_\Theta(a_t \mid x, \mathcal{Z}, a_{<t}) \tag{2}$$

For a self-distillation framework, the ground-truth language CoT gives a teacher latent chain $\mathcal{Z}_t = (z_{t,1}, \ldots, z_{t,K_t})$ and the latent CoT gives a student latent chain $\mathcal{Z}_s = (z_{s,1}, \ldots, z_{s,K_s})$. The same model is co-trained with both textual language modeling and the Latent-LM loss $\mathcal{L}_{\text{Latent-LM}}(x, \mathcal{Z}_s, a; \Theta)$. CODI (Shen et al., 2025) further uses an answer distillation loss $\mathcal{L}_{\text{answer-distill}} = \frac{1}{L}\sum_{l=1}^{L} \| \text{stopgrad}(z_t^{(l)}) - z_s^{(l)} \|^2$ to strengthen the teacher-student alignment in the latent answer space.

Unlike LM in language CoT, only the answer tokens are aligned in $\mathcal{L}_{\text{Latent-LM}}$. These objectives mainly supervise whether the latent chain reaches an answer state for better decoded outcomes, but *give little guidance on the latent reasoning process*. As a result, the model struggles to understand the reasoning space, which may end up taking the easy way out: **intermediate reasoning states crowd near an exit that leads to the answer space**.

We first theoretically show that any differentiable, convex outcome-based latent SFT objective pushes intermediate hidden states toward an "ideal exit". The latent chain converges near this point, which geometrically appears as a crowding of intermediate states around the exit.

Theorem 3.1 (Training-time convergence of latent chain)**.** *Assume for every sample x, there exists a unique ideal* `<CoT-Exit>` *state $e := \arg\min_z \mathcal{L}(z)$ with $\nabla_z \mathcal{L}(e) = 0$. That is, the training loss*

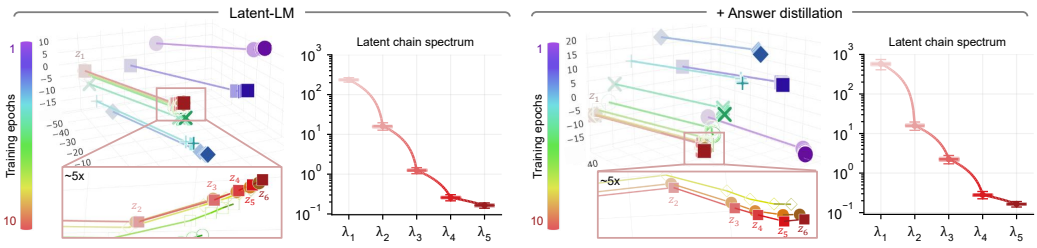


Figure 4: Convergence of latent chains in outcome-rewarding models (MATH; Llama 3.2-1B-Instruct). *Left*: latent chains at different training epochs. Intermediate states z_2–z_5 crowd around the exit z_6. *Right*: spectrum of the latent chain. For both Latent-LM with or without answer distillation, the spectral masses accumulate at λ_1 and exhibit a logarithmic decay for orthogonal directions λ_2–λ_5.

$\mathcal{L}(z)$ tends to make the latent chain $\mathcal{Z} = (z_i)_{i=1}^K$, $K \geq 2$ end at e before entering the answer space. If $\mathcal{L}(z)$ is differentiable and convex over each hidden state $z \in \mathbb{R}^d$, and its gradient is L_g-Lipschitz continuous for any $z_i, z_j \in \mathcal{Z} \cup \{e\}$, namely $\|\nabla_z \mathcal{L}(z_i) - \nabla_z \mathcal{L}(z_j)\| \leq L_g \|z_i - z_j\|$, then, under a first-order linear approximation to the Transformer architecture:

(1) For any $z_k \neq e$, $k = 2, \ldots, K$ we have

$$\|z_k^{(t+1)} - e\|^2 \leq \|z_k^{(t)} - e\|^2 - \underbrace{\eta \left(\frac{2}{L_g} - \eta \right)}_{\geq 0} \|\nabla_z \mathcal{L}(z_k^{(t)})\|^2 \tag{3}$$

where the learning rate $0 < \eta < 2/L_g$.

(2) The ℓ_2 distance from each state $z_k, k \neq 1$ to e converges to zero as training steps increase.

See §A.1 for the proof. Note that we do not include the <CoT-Entrance> state z_1 in the statements since z_1 is mainly determined by the question rather than the reasoning process. Under the condition in Theorem 3.1, both Latent-LM and answer distillation losses push hidden states toward e. This exit does not necessarily map to the correct answer anchor; see §B.3 for more analyses of this issue.

We capture the hidden state trajectories of models fine-tuned by (1) Latent-LM and (2) Latent-LM + answer distillation for fixed training steps in Figure 4. They indicate that, except for <CoT-Entrance>, the hidden states crowd inside a very small neighborhood.

Spectral collapse. The convergence of the latent chain indicates a collapsed reasoning space. It has large spectral mass along the entrance-to-exit direction, while the masses in the orthogonal directions are strongly suppressed. The theorem below quantifies the extremely long-tailed orthogonal spectrum.

Theorem 3.2 (Collapse of the orthogonal subspace)**.** *Let $\mathcal{Z}$ be a latent chain that converges to the ideal <CoT-Exit> state e, with $\|z_k - e\| \leq \varepsilon$ for $k = 2, \ldots, K$ where $0 \leq \varepsilon < \|z_1 - e\|$. Denote the average state by $\bar{z} := \frac{1}{K} \sum_{k=1}^K z_k$, the covariance matrix by $\mathrm{Cov}(\mathbf{Z}) := \frac{1}{K} \sum_{k=1}^K (z_k - \bar{z})(z_k - \bar{z})^\top$, and its eigenvalues by $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. Then we have*

$$\sum_{i \geq 2} \lambda_i \leq \frac{K\varepsilon^2}{(\|z_1 - e\| - \varepsilon)^2} \lambda_1 \tag{4}$$

See §A.2 for the proof. Figure 4 shows the spectrum of outcome-rewarding latent reasoning: the spectral masses accumulate at λ_1 and exhibit a logarithmic decay for the orthogonal directions. Geometrically, the chain is almost straightened, implying that the model tries to jump to a conclusion from the first latent step, instead of a step-by-step reasoning. This could harm the expressivity of reasoning models especially on difficult problems.

3.2 Geometric Decomposition for Latent SFT

To tackle the issues in outcome-rewarding SFT, we utilize geometric decomposition for a more geometrically interpretable training objective. We first discuss the two aims of the proposed method, and then show how they are naturally combined into a single spectral objective.

1st aim: principal direction alignment. This aim **aligns the student's spectral mass on the teacher's principal direction**. We first compute the centralized compact SVD as $\mathbf{Z}_t - \mathbf{1}_{K_t} \bar{z}_t^\top =$

$\mathbf{U}_t\mathbf{\Sigma}_t\mathbf{V}_t^\top$ and $\mathbf{Z}_s - \mathbf{1}_{K_s}\bar{z}_s^\top = \mathbf{U}_s\mathbf{\Sigma}_s\mathbf{V}_s^\top$. Then, the teacher's *principal spectral mass* is

$$m_t^{\rightarrow} := \frac{\lambda_{t,1}}{\mathrm{tr}(\mathbf{\Sigma}_t^2)/K_t + \varepsilon_1} \tag{5}$$

where $\lambda_{t,1}$ is the largest eigenvalue of teacher covariance and ε_1 is a small constant in case of very short CoT samples. Denote by $\lambda_{s,t}^{\rightarrow} := v_t^\top \frac{\mathbf{V}_s\mathbf{\Sigma}_s^2\mathbf{V}_s^\top}{K_s} v_t$ the eigenvalue of the student covariance projected to the teacher's principal direction. The *effective principal spectral mass* of $\mathcal{Z}_s$ projected to $\mathcal{Z}_t$ is

$$m_{s\to t}^{\rightarrow} := \frac{\lambda_{s,t}^{\rightarrow}}{\mathrm{tr}(\mathbf{\Sigma}_s^2)/K_s + \varepsilon_1} \tag{6}$$

And the alignment loss between the masses is

$$\mathcal{L}^{\rightarrow} := m_t^{\rightarrow} \log \frac{m_t^{\rightarrow}}{m_{s\to t}^{\rightarrow}} \equiv -m_t^{\rightarrow} \log m_{s\to t}^{\rightarrow} \tag{7}$$

where the equivalent form drops the teacher-only term $m_t^{\rightarrow} \log m_t^{\rightarrow}$ as teacher gradients are stopped.

2nd aim: orthogonal subspace uniformization (for $r \geq 2$ samples only). Unlike the 1st aim, we do not align the student's orthogonal subspace to the teacher's as its orthogonal masses are less stable and may introduce noise. Instead, we **uniformly expand the student chain over the effective dimensions in the teacher's orthogonal subspace**. Thus, the *uniform target spectral masses* are

$$m_{t,1}^{\perp} = m_{t,2}^{\perp} = \cdots = m_{t,r-1}^{\perp} := \frac{1 - m_t^{\rightarrow}}{r-1} \tag{8}$$

and the student's *effective orthogonal spectral masses* $m_{s\to t,i}^{\perp}$ are defined by the eigenvalues $\lambda_{s\to t,1}^{\perp}, \ldots, \lambda_{s\to t,r-1}^{\perp}$ of the student covariance projected to the teacher orthogonal subspace $(\mathbf{I} - v_t v_t^\top)\frac{\mathbf{V}_s\mathbf{\Sigma}_s^2\mathbf{V}_s^\top}{K_s}(\mathbf{I} - v_t v_t^\top)$, that is

$$m_{s\to t,i}^{\perp} := \frac{\lambda_{s\to t,i}^{\perp}}{\mathrm{tr}(\mathbf{\Sigma}_s^2)/K_s + \varepsilon_1} + \varepsilon_2 \tag{9}$$

where ε_2 is a small constant in case of zero $\lambda_{s\to t,i}^{\perp}$. The orthogonal uniformization loss is

$$\mathcal{L}^{\perp} := \sum_{i=1}^{r-1} m_{t,i}^{\perp} \log \frac{m_{t,i}^{\perp}}{m_{s\to t,i}^{\perp}} \equiv -\sum_{i=1}^{r-1} m_{t,i}^{\perp} \log m_{s\to t,i}^{\perp} \tag{10}$$

Overall geometric decomposition objective. We normalize the spectral masses to form two probability distributions in the effective dimensions of the teacher space:

$$q_t := [m_t^{\rightarrow}, m_{t,1}^{\perp}, \ldots, m_{t,r-1}^{\perp}], \quad q_s := [m_{s\to t}^{\rightarrow}, m_{s\to t,1}^{\perp}, \ldots, m_{s\to t,r-1}^{\perp}] \tag{11}$$

We then combine the two aims into a simple Kullback–Leibler distribution matching objective:

$$\mathcal{L}_{\mathrm{GD}} := \mathcal{L}^{\rightarrow} + \mathcal{L}^{\perp} = m_t^{\rightarrow} \log \frac{m_t^{\rightarrow}}{m_{s\to t}^{\rightarrow}} + \sum_{i=1}^{r-1} m_{t,i}^{\perp} \log \frac{m_{t,i}^{\perp}}{m_{s\to t,i}^{\perp}} = \underline{D_{\mathrm{KL}}(q_t \| q_s)} \tag{12}$$

This form follows the standard KL-based distillation (Hinton et al., 2015), so it can be directly used as a geometry-aware auxiliary loss without introducing new architectures.

Advantages, geometrically and informationally. Through $\mathcal{L}_{\mathrm{GD}}$ we inject global, geometric reasoning information into Latent SFT. *First*, **orthogonal uniformization greatly alleviates the convergence of the reasoning chain and the spectral collapse** in outcome-rewarding methods. Intermediate reasoning states now have a much greater impact on searching for the correct exit. *Second*, **the principal direction alignment prevents reasoning from drifting away from the logical form.** This gives a geometry-aware process-rewarding supervision: compared with step-level supervision, it is less sensitive to noise in reasoning steps and gives more freedom to latent reasoning.

Furthermore, we provide an information-theoretic view for the advantage of $\mathcal{L}_{\mathrm{GD}}$. For outcome-rewarding methods, the answer anchor receives inadequate information about the reasoning space as the hidden states carry almost only one dimension of it. Conversely, $\boldsymbol{\mathcal{L}_{\mathrm{GD}}}$ **expands the orthogonal directions which enrich the information carried by the chain**. To formalize, we introduce the *coding rate* (Yu et al., 2020) in information theory, which estimates the average number of bits needed to encode finite samples up to a distortion level ϵ. The following theorem shows that a uniform orthogonal spectrum maximizes the coding rate.

Table 2: Main results. Best latent reasoning results are in **bold**. See §C.2 for more results.

	MATH			NuminaMath-CoT (10%)			CommonsenseQA		
	Pass@1↑	Avg. #{T}↓	EFF↑	Pass@1↑	Avg. #{T}↓	EFF↑	Pass@1↑	Avg. #{T}↓	EFF↑
Llama 3.2-1B-Instruct									
CoT-SFT	7.10	109.2	0.07	7.69	103.5	0.07	68.06	41.9	1.62
Latent-LM	8.30	6	1.38	12.82	6	2.14	67.73	6	11.29
+ answer distill (Shen et al., 2025)	8.02	6	1.34	12.82	6	2.14	67.16	6	11.19
+ answer distill + step decode (Wei et al., 2026)	7.58	6	1.26	11.54	6	1.92	67.49	6	11.25
+ principal $\mathcal{L}^{\rightarrow}$ (7)	9.62	6	1.60	**14.10**	6	**2.35**	66.83	6	11.14
+ orthogonal $\mathcal{L}^{\perp}$ (10)	9.10	6	1.52	12.82	6	2.14	72.07	6	12.01
+ $\mathcal{L}_{\text{GD}}$ (12)	**9.74**	6	**1.62**	**14.10**	6	**2.35**	**73.14**	6	**12.19**
Llama 3.1-8B-Instruct									
CoT-SFT	23.80	157.7	0.15	20.51	143.7	0.14	52.25	53.2	0.98
Latent-LM	11.54	6	1.92	14.10	6	2.35	71.66	6	11.94
+ answer distill (Shen et al., 2025)	6.40	6	1.07	12.82	6	2.14	**83.87**	6	**13.98**
+ answer distill + step decode (Wei et al., 2026)	5.68	6	0.95	12.82	6	2.14	81.16	6	13.53
+ principal $\mathcal{L}^{\rightarrow}$ (7)	9.54	6	1.59	15.38	6	2.56	71.83	6	11.97
+ orthogonal $\mathcal{L}^{\perp}$ (10)	10.36	6	1.73	**19.23**	6	**3.21**	71.74	6	11.96
+ $\mathcal{L}_{\text{GD}}$ (12)	**12.92**	6	**2.15**	**19.23**	6	**3.21**	82.31	6	13.72
Qwen 3-0.6B									
CoT-SFT	17.50	185.5	0.09	12.82	227.7	0.06	64.37	58.8	1.09
Latent-LM	9.72	6	1.62	16.67	6	2.78	61.75	6	10.29
+ answer distill (Shen et al., 2025)	10.12	6	1.69	19.23	6	3.21	59.62	6	9.94
+ principal $\mathcal{L}^{\rightarrow}$ (7)	11.18	6	1.86	19.23	6	3.21	61.75	6	10.29
+ orthogonal $\mathcal{L}^{\perp}$ (10)	11.30	6	1.88	15.38	6	2.56	62.49	6	10.42
+ $\mathcal{L}_{\text{GD}}$ (12)	**11.62**	6	**1.94**	**20.51**	6	**3.42**	**63.64**	6	**10.61**
Qwen 3-8B									
CoT-SFT	22.46	205.0	0.11	15.38	236.1	0.07	80.92	55.9	1.45
Latent-LM	27.06	6	4.51	29.49	6	4.92	80.34	6	13.39
+ answer distill (Shen et al., 2025)	26.64	6	4.44	28.21	6	4.70	77.81	6	12.97
+ principal $\mathcal{L}^{\rightarrow}$ (7)	20.02	6	3.34	28.21	6	4.70	80.84	6	13.47
+ orthogonal $\mathcal{L}^{\perp}$ (10)	22.04	6	3.67	26.92	6	4.49	80.34	6	13.39
+ $\mathcal{L}_{\text{GD}}$ (12)	**27.62**	6	**4.60**	**34.62**	6	**5.77**	**81.24**	6	**13.54**

Theorem 3.3 (Uniform orthogonal spectrum maximizes coding rate)**.** *Define the coding rate of $\mathcal{Z}$ as*

$$R(\mathcal{Z}, \epsilon) := \frac{1}{2} \log \det \left(\mathbf{I}_r + \frac{d}{K\epsilon^2} \mathbf{\Sigma}^2 \right) \tag{13}$$

where the distortion $\epsilon > 0$. When the total spectral mass $\operatorname{tr}(\mathbf{\Sigma}^2)/K$ and the principal spectral mass $m^{\rightarrow} = \frac{\lambda_1}{\operatorname{tr}(\mathbf{\Sigma}^2)/K}$ are fixed, $R(\mathcal{Z}, \epsilon)$ is maximized if $\lambda_2 = \cdots = \lambda_r$.

See §A.3 for the proof. This indicates that $\mathcal{L}_{\text{GD}}$ makes the latent chain carry more information about the reasoning space by expanding the dimensions; it explains how our approach guides the model to perform latent reasoning in a more interpretable way.

4 Experiments

Setup. We mainly use MATH (Hendrycks et al., 2021) and the first 10% subset of NuminaMath-CoT (Li et al., 2024) for mathematical reasoning SFT. Both datasets contain more than $10\times$ longer language CoTs compared with GSM8K. We also evaluate CommonsenseQA (Talmor et al., 2019) with ground-truth CoT provided by (Shen et al., 2025) for non-mathematical scenarios. We conduct SFT and evaluation on the Llama 3 (Dubey et al., 2024) and Qwen 3 (Yang et al., 2025) families with a fixed number of latent states. Detailed setup and hyperparameters are given in §C.1.

Comparative results. We compare our method with language CoT fine-tuning (CoT-SFT), outcome-rewarding Latent-LM, answer distillation from CODI (Shen et al., 2025), and the process-rewarding step decoding method from SIM-CoT (Wei et al., 2026). We also evaluate the two aims of our objective, principal direction alignment $\mathcal{L}^{\rightarrow}$ and orthogonal subspace uniformization $\mathcal{L}^{\perp}$, separately. Table 2 reports Pass@1, average number of reasoning tokens $\#\{T\}$, and token efficiency (Cao et al., 2026) EFF $:=$ Pass@1$/\#\{T\}$ for CoT compression effectiveness. We observe the following:

(1) **Latent SFT with $\mathcal{L}_{\text{GD}}$ consistently outperforms outcome-rewarding methods on mathematical reasoning.** $\mathcal{L}_{\text{GD}}$ improves over Latent-LM and CODI's answer distillation by relatively $12.7\%/35.5\%$ on MATH and $21.7\%/22.3\%$ on NuminaMath-CoT on average. On CommonsenseQA, our method reaches the best performance for 3 out of 4 models, suggesting that its geometric advantage is not limited to mathematical reasoning.

(2) **Latent reasoning shows strong reasoning token compression**. Language CoT and CoT-SFT often require hundreds of reasoning tokens, which leads to much lower token efficiency. In contrast,

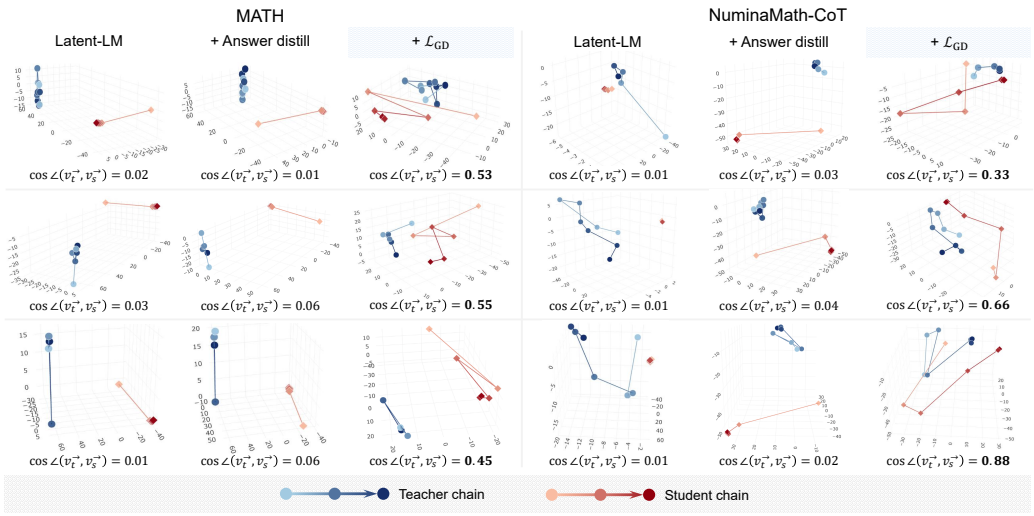


Figure 5: Visualization of latent reasoning spaces (MATH and NuminaMath-CoT; Llama 3.2-1B-Instruct). $\cos\angle(v_t^{\rightarrow}, v_s^{\rightarrow})$ denotes the absolute cosine similarity between the teacher and student principal directions. Compared to answer distillation as an auxiliary, $\mathcal{L}_{\text{GD}}$ exhibits a much higher principal direction similarity and more diverse chain geometry. See §C.4 for more results.

$\mathcal{L}_{\text{GD}}$ consistently surpasses CoT-SFT in EFF with only 6 latent forward passes. This suggests that latent reasoning can improve token efficiency while retaining competitive accuracy.

(3) **Two aims are better than one.** Only using $\mathcal{L}^{\rightarrow}$ may leave latent chain convergence unaddressed, and only using $\mathcal{L}^{\perp}$ may not provide a stable reasoning progress. As a result, neither separate component gives consistent improvements like $\mathcal{L}_{\text{GD}}$ does.

4.1 Visualization of the Latent Reasoning Space

To further inspect the reasoning geometry, we apply PCA to the teacher and student chains jointly and visualize them. Some examples are shown in Figure 5. We also report the absolute cosine similarity between the teacher and student principal directions. Latent-LM and answer distillation both show a clear convergence pattern: the latent states collapse into an almost straight line which is nearly orthogonal to the teacher principal direction. Instead, $\mathcal{L}_{\text{GD}}$ produces latent chains with richer orthogonal structure, as well as making the student principal direction more aligned with the teacher. Figure 6 further compares the overall spectrum of latent chains. With orthogonal subspace uniformization, $\mathcal{L}_{\text{GD}}$ gives up to 267× stronger spectral mass in the orthogonal directions than Latent-LM, greatly alleviating spectral collapse. These geometric effects are important reasons for $\mathcal{L}_{\text{GD}}$'s effectiveness.

See §C for more experiment results.

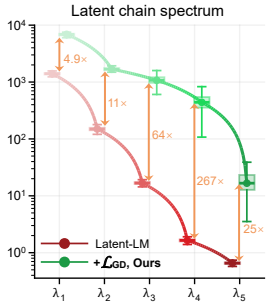


Figure 6: Comparison of the latent chain spectrum (MATH; Llama 3.1-8B-Instruct).

5 Conclusion

We revisit latent reasoning through global reasoning geometry, and decompose the latent reasoning space into the principal direction and the orthogonal subspace. We empirically show that the principal direction is connected to the main logical form, while the orthogonal subspace is associated with semantic transitions. We then argue that outcome-rewarding methods can make intermediate states crowd near an "ideal exit", causing spectral collapse and reducing the information carried by the chain. To address this, we propose a geometric decomposition objective with principal direction alignment and orthogonal subspace uniformization, which preserves the logical form while expanding the orthogonal dimensions in a geometrically interpretable way. Extensive experiments show that our approach consistently improves latent reasoning on difficult mathematical reasoning settings. See §E for discussions and limitations of our work.

AI USE STATEMENT

This work used the GPT-5 family (OpenAI, 2025), mainly including GPT-5.4, 5.5, and 5.6, as research assistants. We list the respective roles of the authors and AI in Table 3. We declare that none of the contributions in this work were made solely by AI; all AI-assisted outputs were reviewed, verified and revised by the authors; the authors take full responsibility for every part of this paper.

Table 3: Use of generative AI in this work.

Activity	What the authors **MAINLY** did	What AI **ONLY** did	What AI did **NOT** do in the slightest
Research ideation	• Origination of all insights stated in the contributions • Detailed implementation schemes of ideas	• Aid in searches and summaries of relevant literature • Discussion of the auxiliary loss design	• Origination of the ideas
Implementation	• Provision of detailed function-wise implementation instructions • Review, modification, and debugging of code • Code execution and result organization	• Aid in coding	• Execution of main experiments, apart from visualization and content-formatting scripts on provided data
Experiment	• Core design and execution of all experiments	• Aid in generation of datasets for the empirical studies in §2.1, see §B.1 for details • Aid in coding visualization scripts • Aid in identifying appropriate evaluation metrics	• Generation of results without experimental evidence
Mathematical proof	• Provision of detailed blueprints and proof sketches • Review, reconstruction, and formatting of proofs • Summarization and formalization of all theorems and conclusions	• Aid in drafting of mathematical proofs • Aid in checking details in manually written parts	• Generation of theorem statements
Paper writing	• Conception and organization of the narrative • Paragraph-wise drafting of the paper • Formatting of BibTeX references • Word-wise proofreading and substantial revisions of AI-assisted content	• Paraphrasing and polishing, strictly following the authors' draft paper • Aid in typesetting refinement • Aid in paper double-checking	• Generation of original content that depart from the authors' intent • Generation of BibTeX entries • Generation of contents in figures or tables

REFERENCES

Ben Alderson-Day and Charles Fernyhough. Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, 141(5):931–965, 2015.

Tyler Alvarez and Ali Baheri. Where does reasoning break? step-level hallucination detection via hidden-state transport geometry. *arXiv:2605.13772*, 2026.

Jean-Bernard Baillon and Georges Haddad. Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones. *Israel Journal of Mathematics*, 26(2):137–150, 1977.

Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michal Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, et al. Graph of thoughts: Solving elaborate problems with large language models. In *AAAI*, volume 38, pp. 17682–17690, 2024.

Apoorva Bhandari, Haley Keglovits, Defne Buyukyazgan, and David Badre. Task structure tailors the geometry of neural representations in human lateral prefrontal cortex. *bioRxiv*, 2025.

Jie Cao, Tianwei Lin, Zhenxuan Fan, Bo Yuan, Ziyuan Zhao, Rolan Yan, Wenqiao Zhang, and Siliang Tang. Draft-thinking: Learning efficient reasoning in long chain-of-thought LLMs. *arXiv:2603.00578*, 2026.

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:2505.16782*, 2025.

Xinghao Chen, Chak Tou Leong, Wenjin Guo, Jian Wang, Wenjie Li, and Xiaoyu Shen. What makes effective supervision in latent chain-of-thought: An information-theoretic analysis. In *ICML*, 2026.

Daixuan Cheng, Shaohan Huang, Xuekai Zhu, Bo Dai, Xin Zhao, Zhenliang Zhang, and Furu Wei. Reasoning with exploration: An entropy perspective. In *AAAI*, volume 40, pp. 30377–30385, 2026.

Jeffrey Cheng and Benjamin Van Durme. Compressed chain of thought: Efficient reasoning through dense representations. *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:2110.14168*, 2021.

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

Yuntian Deng, Yejin Choi, and Stuart Shieber. From explicit CoT to implicit CoT: Learning to internalize CoT step by step. *arXiv:2405.14838*, 2024.

Yubo Dong, Hehe Fan, Linchao Zhu, and Yi Yang. Structured reasoning for LLMs: A unified framework for efficiency and explainability. In *ICLR*, 2026.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Amy Yang, Angela Fan, et al. The Llama 3 herd of models. *arXiv:2407.21783*, 2024.

Timo Flesch, Keno Juechems, Tsvetomira Dumbalska, Andrew Saxe, and Christopher Summerfield. Orthogonal representations for robust context-dependent task performance in brains and neural networks. *Neuron*, 110(7):1258–1270, 2022.

Luyu Gao, Aman Madaan, Shuyan Zhou, Uri Alon, Pengfei Liu, Yiming Yang, Jamie Callan, and Graham Neubig. PAL: Program-aided language models. In *ICML*, pp. 10764–10799, 2023.

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. In *COLM*, 2025.

Yuwen Hao and Menglin Yang. Think wider: Mitigating latent rank collapse in implicit chain-of-thought reasoning. In *EMNLP Findings*, 2026.

Yinhan He, Wendy Zheng, Yaochen Zhu, Zaiyi Zheng, Lin Su, Sriram Vasudevan, Qi Guo, Liangjie Hong, and Jundong Li. SemCoT: Accelerating chain-of-thought reasoning through semantically-aligned implicit tokens. In *NeurIPS*, pp. 43455–43485, 2025.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the MATH dataset. In *NeurIPS*, 2021.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv:1503.02531*, 2015.

Eghbal Hosseini and Evelina Fedorenko. Large language models implicitly learn to straighten neural sentence trajectories to construct a predictive representation of natural language. In *NeurIPS*, volume 36, pp. 43918–43930, 2023.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *ICLR*, 2022.

Jack King, Evelina Fedorenko, and Eghbal A Hosseini. Representational curvature modulates behavioral uncertainty in large language models. In *ICML*, 2026.

Anna Kuzina, Maciej Pioro, and Babak Ehteshami Bejnordi. KaVa: Latent reasoning via compressed KV-cache distillation. In *ICLR*, 2026.

Andrew Lee, Melanie Weber, Fernanda Viégas, and Martin Wattenberg. Shared global and local geometry of language model embeddings. In *COLM*, 2025.

Daniel Lee, Owen Queen, and James Zou. ReasonOps: Operator segmentation for LLM reasoning traces. In *ICML Workshop on Mechanistic Interpretability*, 2026.

Bo Li, Guanzhi Deng, Ronghao Chen, Junrong Yue, Shuo Zhang, Qinghua Zhao, Linqi Song, and Lijie Wen. REMA: A unified reasoning manifold framework for interpreting large language model. *arXiv:2509.22518*, 2025a.

Jia Li, Edward Beeching, Lewis Tunstall, Ben Lipkin, Roman Soletskyi, Shengyi Huang, Kashif Rasul, Longhui Yu, Albert Q Jiang, Ziju Shen, et al. NuminaMath: The largest public dataset in ai4maths with 860k pairs of competition math problems and solutions. 2024. URL https://huggingface.co/datasets/AI-MO/NuminaMath-CoT.

Jindong Li, Yali Fu, Li Fan, Jiahong Liu, Yao Shu, Chengwei Qin, Menglin Yang, Irwin King, and Rex Ying. Implicit reasoning in large language models: A comprehensive survey. *arXiv:2509.02350*, 2025b.

Melody Li, Kumar Krishna Agrawal, Arna Ghosh, Komal Teru, Adam Santoro, Guillaume Lajoie, and Blake Richards. Tracing the representation geometry of language models from pretraining to post-training. In *NeurIPS*, pp. 54691–54724, 2025c.

Tomáš Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *NAACL*, pp. 746–751, 2013.

Alexander Modell, Patrick Rubin-Delanchy, and Nick Whiteley. The origins of representation manifolds in large language models. *arXiv:2505.18235*, 2025.

OpenAI. Introducing GPT-5. https://openai.com/index/introducing-gpt-5/, 2025.

Tianyu Pan and Damon L Woodard. GeoLAN: Geometric learning of latent explanatory directions in large language models. In *ACL Findings*, pp. 5493–5521, 2026.

Kiho Park, Yo Joong Choe, and Victor Veitch. The linear representation hypothesis and the geometry of large language models. In *ICML*, pp. 39643–39666, 2024.

Samyam Rajbhandari, Jeff Rasley, Olatunji Ruwase, and Yuxiong He. ZeRO: Memory optimizations toward training trillion parameter models. In *SC20*, pp. 1–16. IEEE, 2020.

Lord Rayleigh. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10 (60):73–78, 1880.

Hangliang Ren. LSRL: Process-supervised GRPO on latent recurrent states improves mathematical reasoning. In *EMNLP Findings*, pp. 12534–12545, 2025.

Xuan Shen, Yizhou Wang, Yufa Zhou, Xiangxi Shi, Pu Zhao, Yanzhi Wang, and Jiuxiang Gu. Efficient reasoning with hidden thinking. In *ICML*, 2026.

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 *EMNLP*, pp. 677–693, 2025.

Dachuan Shi, Abedelkadir Asi, Keying Li, Xiangchi Yuan, Leyan Pan, Wenke Lee, and Wen Xiao. SwiReasoning: Switch-thinking in latent and explicit for pareto-superior reasoning LLMs. In *ICLR*, 2026.

Ryan Solgi, Jiayi Tian, and Zheng Zhang. LoRi: Low-rank distillation for implicit reasoning. *arXiv:2606.05315*, 2026.

Yuda Song, Julia Kempe, and Remi Munos. Outcome-based exploration for LLM reasoning. In *NeurIPS Workshop on Aligning Reinforcement Learning Experimentalists and Theorists*, 2025.

Lihao Sun, Hang Dong, Bo Qiao, Qingwei Lin, Dongmei Zhang, and Saravan Rajmohan. LLM reasoning as trajectories: Step-specific representation geometry and correctness signals. In *ACL*, pp. 26872–26887, 2026.

Alon Talmor, Jonathan Herzig, Nicholas Lourie, and Jonathan Berant. CommonsenseQA: A question answering challenge targeting commonsense knowledge. In *NAACL*, pp. 4149–4158, 2019.

Wenhui Tan, Jiaze Li, Jianzhong Ju, Zhenbo Luo, Ruihua Song, and Jian Luan. Think silently, think fast: Dynamic latent compression of LLM reasoning chains. In *NeurIPS*, pp. 4646–4668, 2025.

Ramya Keerthy Thatikonda, Wray Buntine, and Ehsan Shareghi. Logical reasoning with outcome reward models for test-time scaling. In *EMNLP*, pp. 26102–26112, 2025.

Hong Wang, Zhezheng Hao, Jian Luo, Chenxing Wei, Yao Shu, Lei Liu, Qiang Lin, Hande Dong, and Jiawei Chen. Scheduling your LLM reinforcement learning with reasoning trees. In *ICLR*, 2026.

Jan Weber, Gabriela Iwama, Anne-Kristin Solbakk, Alejandro O Blenkmann, Pal G Larsson, Jugoslav Ivanovic, Robert T Knight, Tor Endestad, and Randolph Helfrich. Subspace partitioning in the human prefrontal cortex resolves cognitive interference. *Proceedings of the National Academy of Sciences*, 120(28):e2220523120, 2023.

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. In *NeurIPS*, pp. 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 *ICLR*, 2026.

Xiangyu Wen, Junhua Huang, Zeju Li, Min Li, Jianyuan Zhong, Zhijian Xu, Mingxuan Yuan, Yongxiang Huang, and Qiang Xu. Reasoning scaffolding: Distilling the flow of thought from LLMs. In *ICLR*, 2026.

Haoyi Wu, Zhihao Teng, and Kewei Tu. Parallel continuous chain-of-thought with Jacobi iteration. In *EMNLP*, pp. 914–926, 2025.

Yige Xu, Xu Guo, Zhiwei Zeng, and Chunyan Miao. SoftCoT: Soft chain-of-thought for efficient reasoning with LLMs. In *ACL*, pp. 23336–23351, 2025.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv:2505.09388*, 2025.

Yixin Yang, Qingxiu Dong, and Zhifang Sui. Decoding in geometry: Alleviating embedding-space crowding for complex reasoning. *arXiv:2601.22536*, 2026.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. In *NeurIPS*, pp. 11809–11822, 2023.

Yaodong Yu, Kwan Ho Ryan Chan, Chong You, Chaobing Song, and Yi Ma. Learning diverse and discriminative representations via the principle of maximal coding rate reduction. In *NeurIPS*, pp. 9422–9434, 2020.

Zhenrui Yue, Bowen Jin, Huimin Zeng, Honglei Zhuang, Zhen Qin, Jinsung Yoon, Lanyu Shang, Jiawei Han, and Dong Wang. Hybrid latent reasoning via reinforcement learning. In *NeurIPS*, pp. 5501–5530, 2025.

Yuyi Zhang, Boyu Tang, Tianjie Ju, Sufeng Duan, and Gongshen Liu. Do latent tokens think? a causal and adversarial analysis of chain-of-continuous-thought. *arXiv:2512.21711*, 2025a.

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. In *NeurIPS*, pp. 168990–169012, 2025b.

Dayuan Zhao, Shengcao Cao, Yu-Xiong Wang, and Liang-Yan Gui. Think in latent, explain in language: Self-explainable latent reasoning. In *ICML*, 2026.

Yufa Zhou, Yixiao Wang, Xunjian Yin, Shuyan Zhou, and Anru Zhang. The geometry of reasoning: Flowing logics in representation space. In *ICLR*, 2026.

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. In *NeurIPS*, pp. 79931–79963, 2025.

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 *ICLR*, 2026.

APPENDIX CONTENTS

A Proofs **16**

A.1 Proof of Theorem 3.1 16

A.2 Proof of Theorem 3.2 17

A.3 Proof of Theorem 3.3 18

B Additional Empirical Study Results **19**

B.1 The Principal Direction (§2.1) 19

B.2 The Orthogonal Subspace (§2.2) 20

B.3 Why Can't We Only Align Answer Anchors? 20

C Additional Experiment Results **21**

C.1 Experiment Setup 21

C.2 Results on GSM8K 22

C.3 Parameter Analyses 22

C.4 Visualization of the Latent Reasoning Space (§4.1) 23

D Related Work **23**

E Discussion and Limitations **26**

E.1 Stability and Directionality of the Principal Direction 26

E.2 Broader Applications 27

E.3 Limitations 27

A PROOFS

A.1 PROOF OF THEOREM 3.1

Theorem 3.1 (Training-time convergence of latent chain). *Assume for every sample x, there exists a unique ideal <CoT-Exit> state $\mathrm{e} := \arg\min_z \mathcal{L}(z)$ with $\nabla_z \mathcal{L}(\mathrm{e}) = 0$. That is, the training loss $\mathcal{L}(z)$ tends to make the latent chain $\mathcal{Z} = (z_i)_{i=1}^K$, $K \geq 2$ end at e before entering the answer space. If $\mathcal{L}(z)$ is differentiable and convex over each hidden state $z \in \mathbb{R}^d$, and its gradient is L_g-Lipschitz continuous for any $z_i, z_j \in \mathcal{Z} \cup \{\mathrm{e}\}$, namely $\|\nabla_z \mathcal{L}(z_i) - \nabla_z \mathcal{L}(z_j)\| \leq L_g \|z_i - z_j\|$, then, under a first-order linear approximation to the Transformer architecture:*

(1) For any $z_k \neq \mathrm{e}$, $k = 2, \ldots, K$ we have

$$\|z_k^{(t+1)} - \mathrm{e}\|^2 \leq \|z_k^{(t)} - \mathrm{e}\|^2 - \underbrace{\eta \left(\frac{2}{L_g} - \eta\right)}_{\geq 0} \|\nabla_z \mathcal{L}(z_k^{(t)})\|^2 \tag{14}$$

where the learning rate $0 < \eta < 2/L_g$.

(2) The ℓ_2 distance from each state $z_k, k \neq 1$ to e converges to zero as training steps increase.

Proof. (1) By the first-order property of convex functions:

$$\mathcal{L}(\mathrm{e}) \geq \mathcal{L}(z_k) + \langle \nabla_z \mathcal{L}(z_k), \mathrm{e} - z_k \rangle \Rightarrow \langle \nabla_z \mathcal{L}(z_k), z_k - \mathrm{e} \rangle \geq \mathcal{L}(z_k) - \mathcal{L}(\mathrm{e}) > 0 \tag{15}$$

Thus, the descent direction $-\nabla_z \mathcal{L}(z_k)$ always has a component pointing from z_k toward e.

Since $\mathcal{L}$ is convex in z_k and $\nabla_z \mathcal{L}(z_k)$ is L_g-Lipschitz continuous, the Baillon-Haddad theorem (Baillon & Haddad, 1977) gives

$$\langle \nabla_z \mathcal{L}(z_k) - \nabla_z \mathcal{L}(\mathrm{e}), z_k - \mathrm{e} \rangle \geq \frac{1}{L_g} \|\nabla_z \mathcal{L}(z_k) - \nabla_z \mathcal{L}(\mathrm{e})\|^2 \tag{16}$$

$$(\nabla_z \mathcal{L}(\mathrm{e}) = 0) \Rightarrow \qquad \langle \nabla_z \mathcal{L}(z_k), z_k - \mathrm{e} \rangle \geq \frac{1}{L_g} \|\nabla_z \mathcal{L}(z_k)\|^2 \tag{17}$$

Assume that each hidden state is updated by $z_k^{(t+1)} \leftarrow z_k^{(t)} - \eta \nabla_z \mathcal{L}(z_k^{(t)})$. Then

$$\|z_k^{(t+1)} - \mathrm{e}\|^2 = \|z_k^{(t)} - \mathrm{e} - \eta \nabla_z \mathcal{L}(z_k^{(t)})\|^2 \tag{18}$$

$$= \|z_k^{(t)} - \mathrm{e}\|^2 - 2\eta \left\langle \nabla_z \mathcal{L}(z_k^{(t)}), z_k^{(t)} - \mathrm{e} \right\rangle + \eta^2 \|\nabla_z \mathcal{L}(z_k^{(t)})\|^2 \tag{19}$$

$$\leq \|z_k^{(t)} - \mathrm{e}\|^2 - \eta \left(\frac{2}{L_g} - \eta\right) \|\nabla_z \mathcal{L}(z_k^{(t)})\|^2 \tag{20}$$

which means that $\|z_k^{(t)} - \mathrm{e}\|^2$ is non-increasing when $0 < \eta < 2/L_g$.

(2) Summing the inequality (20) over training steps gives

$$\eta \left(\frac{2}{L_g} - \eta\right) \sum_{t=0}^{T-1} \|\nabla_z \mathcal{L}(z_k^{(t)})\|^2 \leq \|z_k^{(0)} - \mathrm{e}\|^2 \tag{21}$$

for any T. Since the sum is bounded for arbitrarily large T, we have $\lim_{t\to\infty} \|\nabla_z \mathcal{L}(z_k^{(t)})\| = 0$. By the Bolzano-Weierstrass theorem, there exists a sub-chain $(z_k^{(t_j)}), t_1 < t_2 < \ldots$ with average state $\bar{z}_k$ such that $\lim_{j\to\infty} z_k^{(t_j)} = \bar{z}_k$. The sub-chain also satisfies $\lim_{j\to\infty} \nabla_z \mathcal{L}(z_k^{(t_j)}) = 0$. We know from the gradient L_g-Lipschitz continuity that

$$\|\nabla_z \mathcal{L}(z_k^{(t_j)}) - \nabla_z \mathcal{L}(\bar{z}_k)\| \leq L_g \|z_k^{(t_j)} - \bar{z}_k\| \tag{22}$$

So when $t \to \infty$, $\lim_{j\to\infty} \|\nabla_z \mathcal{L}(z_k^{(t_j)}) - \nabla_z \mathcal{L}(\bar{z}_k)\| \to 0$ and thus $\lim_{j\to\infty} \nabla_z \mathcal{L}(z_k^{(t_j)}) = \nabla_z \mathcal{L}(\bar{z}_k) = 0$. This marks $\bar{z}_k$ the ideal <CoT-Exit> so we have $\mathrm{e} = \bar{z}_k$.

Now we look at the convergence of the distance sequence $\|z_k^{(t)} - \mathrm{e}\|$. We know that $0 \leq \|z_k^{(t+1)} - \mathrm{e}\|^2 \leq \|z_k^{(t)} - \mathrm{e}\|^2$ when $0 < \eta < 2/L_g$, so it is convergent. Taking the subsequence $(\|z_k^{(t_j)} - \mathrm{e}\|)_{t_j}$, we have $\lim_{j\to\infty} \|z_k^{(t_j)} - \mathrm{e}\| = \|\mathrm{e} - \mathrm{e}\| = 0$; so $\lim_{t\to\infty} \|z_k^{(t)} - \mathrm{e}\| = 0$ holds for the entire sequence.

□

A.2 Proof of Theorem 3.2

Theorem 3.2 (Collapse of the orthogonal subspace). *Let $\mathcal{Z}$ be a latent chain that converges to the ideal* `<CoT-Exit>` *state* $\mathtt{e}$*, with* $\|z_k - \mathtt{e}\| \le \varepsilon$ *for* $k = 2, \dots, K$ *where* $0 \le \varepsilon < \|z_1 - \mathtt{e}\|$*. Denote the average state by* $\bar{z} := \frac{1}{K}\sum_{k=1}^{K} z_k$*, the covariance matrix by* $\mathrm{Cov}(\mathbf{Z}) := \frac{1}{K}\sum_{k=1}^{K}(z_k - \bar{z})(z_k - \bar{z})^\top$*, and its eigenvalues by* $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$*. Then we have*

$$\sum_{i \ge 2} \lambda_i \le \frac{K\varepsilon^2}{\left(\|z_1 - \mathtt{e}\| - \varepsilon\right)^2}\lambda_1 \tag{23}$$

Proof. Let $\bar{z}_{2:K} := \frac{1}{K-1}\sum_{k=2}^{K} z_k$ be the average state excluding `<CoT-Entrance>`. The covariance of $\mathcal{Z}$ can be decomposed as

$$\mathrm{Cov}(\mathbf{Z}) = \underbrace{\frac{K-1}{K^2}(z_1 - \bar{z}_{2:K})(z_1 - \bar{z}_{2:K})^\top}_{\texttt{<CoT-Entrance>,rank=1}} + \underbrace{\frac{1}{K}\sum_{k=2}^{K}(z_k - \bar{z}_{2:K})(z_k - \bar{z}_{2:K})^\top}_{\mathcal{Z}\setminus\texttt{<CoT-Entrance>}} \tag{24}$$

Both terms are positive semi-definite. Thus,

$$\lambda_1 \ge \left(\frac{z_1 - \bar{z}_{2:K}}{\|z_1 - \bar{z}_{2:K}\|}\right)^\top \mathrm{Cov}(\mathbf{Z}) \left(\frac{z_1 - \bar{z}_{2:K}}{\|z_1 - \bar{z}_{2:K}\|}\right) \tag{25}$$

$$\ge \frac{K-1}{K^2}\left(\frac{z_1 - \bar{z}_{2:K}}{\|z_1 - \bar{z}_{2:K}\|}\right)^\top (z_1 - \bar{z}_{2:K})(z_1 - \bar{z}_{2:K})^\top \left(\frac{z_1 - \bar{z}_{2:K}}{\|z_1 - \bar{z}_{2:K}\|}\right) \tag{26}$$

$$= \frac{K-1}{K^2}\|z_1 - \bar{z}_{2:K}\|^2 \tag{27}$$

So

$$\sum_{i \ge 2} \lambda_i = \mathrm{tr}(\mathrm{Cov}(\mathbf{Z})) - \lambda_1 \tag{28}$$

$$= \underbrace{\frac{K-1}{K^2}\|z_1 - \bar{z}_{2:K}\|^2 - \lambda_1}_{\le 0} + \frac{1}{K}\sum_{k=2}^{K}\|z_k - \bar{z}_{2:K}\|^2 \tag{29}$$

$$\le \frac{1}{K}\sum_{k=2}^{K}\|z_k - \bar{z}_{2:K}\|^2 \tag{30}$$

Next we discuss the distance between z_k and $\mathtt{e}$. We can first obtain that

$$\sum_{k=2}^{K}\|z_k - e\|^2 = \sum_{k=2}^{K}\|(z_k - \bar{z}_{2:K}) + (\bar{z}_{2:K} - \mathtt{e})\|^2 \tag{31}$$

$$= \sum_{k=2}^{K}\|z_k - \bar{z}_{2:K}\|^2 + 2\,\langle \underbrace{\sum_{k=2}^{K}(z_k - \bar{z}_{2:K})}_{0}, \bar{z}_{2:K} - \mathtt{e}\rangle + (K-1)\|\bar{z}_{2:K} - \mathtt{e}\|^2 \tag{32}$$

$$= \sum_{k=2}^{K}\|z_k - \bar{z}_{2:K}\|^2 + (K-1)\|\bar{z}_{2:K} - \mathtt{e}\|^2 \tag{33}$$

$$\ge \sum_{k=2}^{K}\|z_k - \bar{z}_{2:K}\|^2 \tag{34}$$

As $\|z_k - e\| \le \varepsilon$, $\sum_{k=2}^{K}\|z_k - \bar{z}_{2:K}\|^2 \le \sum_{k=2}^{K}\varepsilon^2 = (K-1)\varepsilon^2$. Substituting into (30) gives

$$\sum_{i \ge 2}\lambda_i \le \frac{K-1}{K}\varepsilon^2 \tag{35}$$

On the other hand,

$$\|\bar{z}_{2:K} - \mathtt{e}\| = \left\|\frac{1}{K-1}\sum_{k=2}^{K}(z_k - \mathtt{e})\right\| \le \frac{1}{K-1}\sum_{k=2}^{K}\|z_k - \mathtt{e}\| \le \varepsilon \tag{36}$$

So

$$\|z_1 - \bar{z}_{2:K}\| \geq \|z_1 - \mathrm{e}\| - \|\bar{z}_{2:K} - \mathrm{e}\| \geq \|z_1 - \mathrm{e}\| - \varepsilon \tag{37}$$

Substituting this into (27) gives

$$\lambda_1 \geq \frac{K-1}{K^2} \left(\|z_1 - \mathrm{e}\| - \varepsilon\right)^2 \tag{38}$$

Combining (35) and (38) to obtain the final result:

$$\sum_{i \geq 2} \lambda_i \leq \frac{K\varepsilon^2}{\left(\|z_1 - \mathrm{e}\| - \varepsilon\right)^2} \lambda_1 \tag{39}$$

□

A.3 Proof of Theorem 3.3

Theorem 3.3 (Uniform orthogonal spectrum maximizes coding rate). *Define the coding rate of $\mathcal{Z}$ as*

$$R(\mathcal{Z}, \epsilon) := \frac{1}{2} \log\det\left(\mathbf{I}_r + \frac{d}{K\epsilon^2}\mathbf{\Sigma}^2\right) \tag{40}$$

where the distortion $\epsilon > 0$. When the total spectral mass $\mathrm{tr}(\mathbf{\Sigma}^2)/K$ and the principal spectral mass $m^{\rightarrow} = \frac{\lambda_1}{\mathrm{tr}(\mathbf{\Sigma}^2)/K}$ are fixed, $R(\mathcal{Z}, \epsilon)$ is maximized if $\lambda_2 = \cdots = \lambda_r$.

Proof. Since $\lambda_i = \sigma_i^2/K$, the coding rate can be written as

$$R(\mathcal{Z}, \epsilon) = \frac{1}{2} \log\det\left(\mathbf{I}_r + \frac{d}{K\epsilon^2}\mathbf{\Sigma}^2\right) = \frac{1}{2} \log \prod_{i=1}^{r} \left(1 + \frac{d\lambda_i}{\epsilon^2}\right) \tag{41}$$

Because both the total spectral mass and the principal spectral mass $m^{\rightarrow}$ are fixed, λ_1 and the residual spectral sum $\sum_{i=2}^{r} \lambda_i = \frac{1-m^{\rightarrow}}{m^{\rightarrow}}\lambda_1$ are also fixed. Thus, we can decompose the coding rate into the principal and the orthogonal parts:

$$R(\mathcal{Z}, \epsilon) = \underbrace{\frac{1}{2} \log\left(1 + \frac{d\lambda_1}{\epsilon^2}\right)}_{\text{constant}} + \frac{1}{2} \sum_{i=2}^{r} \log\left(1 + \frac{d\lambda_i}{\epsilon^2}\right) \tag{42}$$

So maximizing $R(\mathcal{Z}, \epsilon)$ is equivalent to optimizing the second term of the RHS. In this case, $R(\mathcal{Z}, \epsilon)$ is strictly concave with respect to the orthogonal spectrum, since

$$\frac{\partial^2}{\partial \lambda_i^2}\left(\frac{1}{2} \log\left(1 + \frac{d\lambda_i}{\epsilon^2}\right)\right) = -\frac{d^2}{2(\epsilon^2 + d\lambda_i)^2} < 0 \tag{43}$$

for all $\lambda_i \geq 0$. By Jensen's inequality,

$$\frac{1}{r-1} \sum_{i=2}^{r} \frac{1}{2} \log\left(1 + \frac{d\lambda_i}{\epsilon^2}\right) \leq \frac{1}{2} \log\left(1 + \frac{d}{(r-1)\epsilon^2} \sum_{i=2}^{r} \lambda_i\right) \tag{44}$$

$$\left(\sum_{i=2}^{r} \lambda_i = \frac{1-m^{\rightarrow}}{m^{\rightarrow}}\lambda_1\right) \Rightarrow \qquad = \frac{1}{2} \log\left(1 + \frac{d(1-m^{\rightarrow})\lambda_1}{(r-1)m^{\rightarrow}\epsilon^2}\right) \tag{45}$$

Therefore, $R(\mathcal{Z}, \epsilon)$ is maximized when $\lambda_2 = \cdots = \lambda_r$, which gives

$$\max R(\mathcal{Z}, \epsilon) = \frac{1}{2} \log\left(1 + \frac{d\lambda_1}{\epsilon^2}\right) + \frac{r-1}{2} \log\left(1 + \frac{d(1-m^{\rightarrow})\lambda_1}{(r-1)m^{\rightarrow}\epsilon^2}\right) \tag{46}$$

□

Table 4: Full version of Table 1: question examples and reasoning templates of every question groups used in our empirical experiments.

Description	Question #1	Question #2	Reasoning template 1	Reasoning template 2
			GSM8K samples	
Question group #1 *Total sum of the first, second, and last (0) parts*	*"Billy sells CDs. He has 1,250 customers on Wednesday. His first 500 customers buy 3 CDs each. His next 450 customers buy 8 CDs each. His last 300 customers do not buy any CDs. How many CDs does Billy sell on Wednesday?"* ***Answer: 5100***	*"At an arcade counter, the first 125,000 kids win 8 tokens each. The next 175,000 kids win 3 tokens each. The last 120,000 kids win none. How many tokens are handed out in total?"* ***Answer: 1525000***	*"The first part contributes 500 * 3 = 1,500. The second part contributes 450 * 8 = 3,600. The remaining part contributes 0, so the total is 1,500 + 3,600 + 0 = 5,100."*	*"If all 1,250 cases contributed 3 each, the total would be 1,250 * 3 = 3,750. The second group differs from that baseline by 8 - 3 per case, so the adjustment is 450 * (8 - 3) = 2,250. The last group gives none, so remove 300 * 3 = 900, leaving 3,750 + 2,250 - 900 = 5,100."*
Question group #2 *Overall rate across three groups (with a derived 3rd group rate)*	*"Janet is collecting the results of Herbert Hoover Elementary School's annual standardized test. 340 out of 500 third-graders passed, along with 40 out of 100 fourth graders. The 400 fifth graders had a pass rate that was twice the fourth grades' pass rate. What is the school's overall pass rate?"* ***Answer: 70***	*"A library tracks on-time returns in three branches. The first branch has 75 on-time returns out of 150 loans, and the second branch has 80 on-time returns out of 400 loans. The third branch has 450 loans with an on-time rate three times the second branch's rate. What is the overall on-time return rate?"* ***Answer: 42.5***	*"First find the rate for the second group: 40 / 100 = 0.4. Then multiply that rate by 2 to find the third group's rate: 0.4 * 2 = 0.8. Then multiply the third group's rate by its size: 0.8 * 400 = 320. Now add the successful cases from each group: 340 + 40 + 320 = 700. Add the sizes of all groups: 500 + 100 + 400 = 1,000. Finally, divide successful cases by total cases and multiply by 100%: 700 / 1,000 * 100% = 70%."*	*"The first group has rate 340 / 500 = 0.68. The second group has rate 40 / 100 = 0.4. The third group has 2 times the second group's rate, so its rate is 0.4 * 2 = 0.8. The total size is 500 + 100 + 400 = 1,000. The weighted overall rate is (500/1,000) * 0.68 + (100/1,000) * 0.4 + (400/1,000) * 0.8 = 0.7. Converting this to a percent gives 70%."*
Question group #3 *Adding a second type to a final fixed ratio*	*"Bob had 7 fish in his ornamental fish pond. 3 were orange, and 4 were white. He bought 17 fish from a mixed tank. After adding them to the pond, he had twice as many orange fish as white fish. How many white fish did Bob buy?"* ***Answer: 4***	*"A laboratory tray holds 54,321 clear tubes and 1,234 amber tubes. Another 44,444 tubes of the two types are placed on the tray. The final number of clear tubes is twice the final number of amber tubes. How many amber tubes were added?"* ***Answer: 32099***	*"The final total is 3 + 4 + 17 = 24. The final total consists of 2 + 1 = 3 equal groups because the first group is 2 times the second group. Each group therefore contains 24 / 3 = 8 items. The number added to the second group is 8 - 4 = 4."*	*"Let x be the number of items from the second group that were added, so 17 - x items from the first group were added. The final counts are 3 + 17 - x for the first group and 4 + x for the second group. Because the first group is 2 times the second group, 3 + 17 - x = 2(4 + x). Solving this equation gives x = 4, so 4 items from the second group were added."*
			MATH samples	
Question group #1 *Four ordered integers and six pairwise sums*	*"Four distinct integers a, b, c and d have the property that when added in pairs, the sums 10, 18, 19, 20, 21, and 29 are obtained. What are the four integers in increasing order? (place a comma and then a space between each integer)"* ***Answer: 4,6,14,15***	*"A delivery center dispatches four drones for a test run. Each drone carries a distinct positive integer number of payload units, represented by a, b, c and d in increasing order. Because individual telemetry is unavailable, only the combined payloads for each pair are recorded: 1323, 2010, 2517, 3158, 3665, and 4352. What are the four payload values in increasing order? (place a comma and then a space between each integer)"* ***Answer: 408,915,1602,2750***	*"Let the integers satisfy $a < b < c < d$. The smallest pair sum is $a + b = 10$, and the second-smallest pair sum is $a+c = 18$. The largest pair sum is $c + d = 29$, and the second-largest pair sum is $b + d = 21$. The two unused sums must be $a + d$ and $b + c$. First suppose $a + d = 19$ and $b + c = 20$. Then $(a + b) + (b + c) - (a + c) = 10 + 20 - 18 = 12$, so $2b = 12$ and $b = 6$. This gives $a = 10 - 6 = 4$, $c = 18 - 4 = 14$, and $d = 29 - 14 = 15$. Next suppose $a + d = 20$ and $b + c = 19$. Then $(a + b) + (b + c) - (a + c) = 10 + 19 - 18 = 11$, so $2b = 11$ and $b = 5.5$. This case is impossible because b must be an integer. Therefore the four integers are $\boxed{4, 6, 14, 15}$."*	*"Let the integers satisfy $a < b < c < d$. The smallest pair sum is $a + b = 10$, and the second-smallest pair sum is $a + c = 18$. Subtracting these two equations gives $c - b = 18 - 10 = 8$. Since $a < b$ and $a + b = 10$, b must be at least 6. Since $c < d$ and $c + d = 29$, c can be at most 14. But $c = b + 8$, so $b + 8 \leq 14$ and $b \leq 6$. Therefore b is an integer between 6 and 6. Checking these candidates against the remaining pair sums gives $b = 6$. Then $c = b + 8 = 14$. From $a + b = 10$, we get $a = 4$. From $c + d = 29$, we get $d = 15$. Thus the four integers are $\boxed{4, 6, 14, 15}$."*
Question group #2 *Extremum of a quadratic*	*"In an equation of the form $k = ax^2 + bx + c$ with $a > 0$, the least possible value of k occurs at $x = -b/(2a)$. In the equation $k = (6x + 12)(x - 8)$, what is the least possible value for k?"* ***Answer: -150***	*"A graphic filter assigns an intensity index $k = (-2x + 3)(x - 5)$ to each real setting x. What is the greatest possible value of the intensity index?"* ***Answer: 6.125***	*"Let $k = (6x + 12)(x - 8)$. Expanding the product gives $k = 6x^2 - 36x - 96$. The coefficient of x^2 is 6. The coefficient of x is -36. The vertex formula gives $x = -(-36)/(2 \times 6)$. Therefore $x = 3$. At this value, $x^2 = 3^2 = 9$. Substituting into the expanded equation gives $k = 6 \times 9 - 36 \times 3 - 96 = \boxed{-150}$."*	*"Let $k = (6x + 12)(x - 8)$. The first factor is zero when $6x + 12 = 0$. Thus the first root is $x = -2$. The second factor is zero when $x - 8 = 0$. Thus the second root is $x = 8$. The vertex lies halfway between the two roots. Therefore $x = ((-2) + 8)/2 = 3$. At this value, the first factor is $6 \times 3 + 12 = 30$. The second factor is $3 - 8 = -5$. Multiplying the two factors gives $k = 30 \times (-5) = \boxed{-150}$."*
Question group #3 *Circumcircle through the pairwise intersections of three lines*	*"The vertices of a triangle are the points of intersection of the line $y = -x - 1$, the line $x = 2$, and $y = \frac{1}{5}x + \frac{13}{5}$. Find an equation of the circle passing through all three vertices."* ***Answer:*** $x^2 + y^2 = 13$	*"On a large-scale logistics map with axes g and t, three shipping corridors follow $t = 5g - 91$, $t = 84$, and $5g + t = -91$. Their pairwise crossings form a triangular zone. Find an equation of the circle through the three crossings."* ***Answer:*** $g^2 + t^2 = 8281$	*"Write the three given lines as $L_1 = x + y + 1$, $L_2 = x - 2$, and $L_3 = -x + 5y - 13$, each equal to 0. Let A, B, and C be the intersections of L_1, L_2, L_2, L_3, and L_1, L_3, respectively. Solving $L_1 = 0$ and $L_2 = 0$ gives $A = (2, -3)$. Solving $L_2 = 0$ and $L_3 = 0$ gives $B = (2, 3)$. Solving $L_1 = 0$ and $L_3 = 0$ gives $C = (-3, 2)$. The perpendicular bisector of AB is $y = 0$. The perpendicular bisector of AC is $x - y = 0$. The two perpendicular bisectors intersect at $(0, 0)$, so this is the center of the circle. The squared radius is $2^2 + (-3)^2 = 13$, so the circle is $\boxed{x^2 + y^2 = 13}$."*	*"Write the three given lines as $L_1 = x + y + 1$, $L_2 = x - 2$, and $L_3 = -x + 5y - 13$, each equal to 0. Any equation of the form $Q = \alpha L_1 L_2 + \beta L_2 L_3 + \gamma L_3 L_1 = 0$ passes through all three pairwise intersections. Expanding gives $L_1 L_2 = x^2 + xy - x - 2y - 2$. Also, $L_2 L_3 = -x^2 + 5xy - 11x - 10y + 26$. Finally, $L_3 L_1 = -x^2 + 4xy - 14x + 5y^2 - 8y - 13$. For Q to be a circle, its xy coefficient must be 0 and its x^2 and y^2 coefficients must be equal, giving $\alpha + 5\beta + 4\gamma = 0$ and $\alpha - \beta - 6\gamma = 0$. Choosing $\gamma = 3$ gives $\alpha = 13$ and $\beta = -5$. Thus $Q = 13L_1 L_2 - 5L_2 L_3 + 3L_3 L_1 = 0$. Simplifying this equation gives $\boxed{x^2 + y^2 = 13}$."*

B Additional Empirical Study Results

B.1 The Principal Direction (§2.1)

Setup. For each reasoning sample, we use GPT-5.6-Luna to generate nine perturbed variants. We then construct two reasoning templates for each group, and instantiate them into CoTs. Finally, we

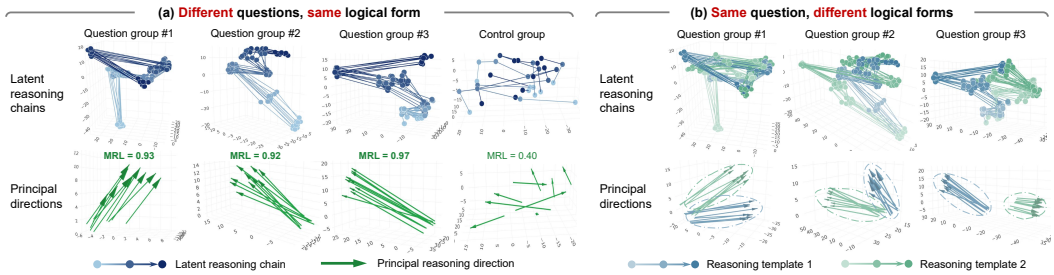


Figure 7: Connection between the principal direction and reasoning logic (MATH samples; Llama 3.1-8B-Instruct). **(a)** Latent chains and principal directions of question groups with different questions but the same reasoning template. MRL stands for mean resultant length. The control group contains 10 random samples from 100 jointly-PCAed questions. **(b)** Two groups with the same questions but different reasoning templates. Two principal-direction clusters point toward different directions whose angular separation varies across question groups.

manually revise and correct the generated reasoning steps, like format errors, logical errors and hallucinations. In total, each question group contains 20 samples. The questions differ in scenarios, numbers and entities, but they can be solved correctly by the same two reasoning templates. Examples of the generated questions and reasoning templates are shown in Table 4. Full questions are provided in our repository.

For a group of latent chains $\{\mathcal{Z}_i\}$ and their principal directions $\{v_i^{\rightarrow}\}$, we define the mean resultant length as

$$\mathrm{MRL} = \frac{\|\sum_i a_i v_i^{\rightarrow}\|}{\sum_i a_i}, \quad a_i = \max_t \langle z_{i,t} - \bar{z}_i,\ v_i^{\rightarrow} \rangle \tag{47}$$

where a_i is the scale of chain $\mathcal{Z}_i$ along $v_i^{\rightarrow}$.

Additional visualization on MATH samples. We also construct question groups from MATH to verify the conclusion on longer and more complex CoTs. Figure 7 shows a consistent pattern: the principal directions are only mildly affected by shallow perturbations, but change clearly with the main reasoning thread. This further supports the connection between the principal direction and the logical form in reasoning.

B.2 The Orthogonal Subspace (§2.2)

Setup. We conduct the experiment using the reasoning CoT provided by ReasonOps (Lee et al., 2026). They are generated by different models and sizes, including Qwen3, Claude-4.5, Kimi-K2.5 and more. See Table 4 in Lee et al. (2026) for details.

We define the reported orthogonal displacement as

$$\|\Delta_k^{\perp}\| = \|\Delta_k - (\Delta_k^{\top} v^{\rightarrow}) v^{\rightarrow}\| \tag{48}$$

where $\Delta_k = z_{k+1} - z_k$ is the difference of two adjacent latent states and $v^{\rightarrow}$ is the principal direction vector. Note that this is a local step-level measure; it does not fully describe the global orthogonal subspace, but it provides a simple way to interpret the connection between semantic logic and reasoning geometry.

Additional results. Figure 8 shows the same statistics using the latent space of Qwen3-8B. The result is consistent with the main figure: **Transition** and **Inference** have the strongest and weakest orthogonal displacement, respectively.

B.3 Why Can't We Only Align Answer Anchors?

Answer distillation aligns the hidden representations of the predicted and the ground-truth answer anchor in the answer space. However, as discussed in §3.1, it gives little direct guidance to the reasoning space, and intermediate states can crowd near an "ideal <CoT-Exit>". With a fixed

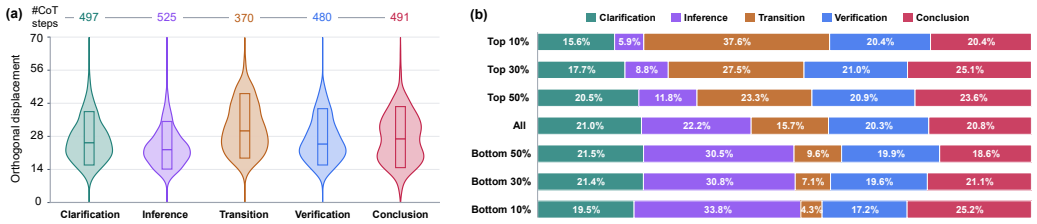


Figure 8: Semantic logic have different strengths in the orthogonal subspace (MATH; Qwen 3-8B). **(a)** Distribution of orthogonal displacement grouped by logical pivots. **(b)** Proportion of logical pivots among reasoning steps with different levels of orthogonal displacement.

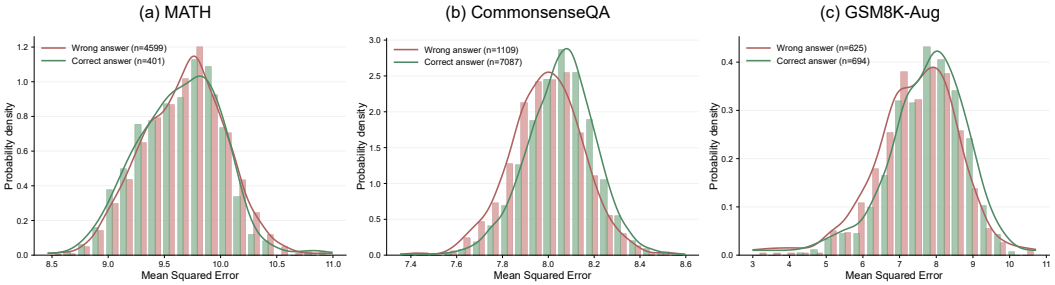


Figure 9: Probability distribution of the Mean Squared Error between the predicted and the ground-truth answer anchor (Llama 3.2-1B-Instruct trained with Latent-LM + answer distillation). The relative offsets of correct and wrong answer bars do not affect the distribution curves.

number of latent forward passes, this exit can be understood as an uninformative conclusion state: once the latent chain reaches this state, the model tends to leave the reasoning space and move toward the answer space. It does not necessarily mean that the chain has reached the correct answer anchor, as **the distance between predicted and ground-truth answer anchors does not reliably distinguish correct and wrong answers**.

To inspect this issue, we compare the Mean Squared Error (MSE) between the predicted answer anchor and the ground-truth answer anchor for correct and wrong predictions, from the model trained with Latent-LM + answer distillation (Shen et al., 2025). As shown in Figure 9, correct predictions do not always have a clearly smaller anchor MSE than wrong predictions; in some cases, their distributions even shift toward larger values. This indicates that it is still possible to obtain the correct answer even if the anchor latent states are not perfectly aligned. We suggest that one source of this ambiguity is the difference of context between language reasoning and latent reasoning. Therefore, despite that outcome-rewarding objectives align states in the answer space, preserving the reasoning information remains crucial for obtaining the correct answer, especially on complex problems.

C Additional Experiment Results

C.1 Experiment Setup

Datasets. MATH (Hendrycks et al., 2021) is a competition-level mathematical reasoning benchmark with multi-step problems and full step-by-step solutions across several mathematical subjects. We keep samples with fewer than 2,048 total tokens. NuminaMath-CoT (Li et al., 2024) is a large collection of CoT-formatted mathematical solutions built from sources including AoPS forum problems, AMC/AIME, Chinese K-12 exercises, GSM8K, MATH, olympiad problems, OrcaMath and synthetic math subsets. We only use its first 10% subset to train the models, and proof-style questions are filtered during evaluation. CommonsenseQA (Talmor et al., 2019) is a multiple-choice commonsense reasoning benchmark built from conceptual relations, and we use the CoT annotations provided by CODI (Shen et al., 2025). We also evaluate on grade-school-level mathematical reasoning datasets GSM8K-Aug (Cobbe et al., 2021; Deng et al., 2023) and GSM8K-Hard (Gao et al., 2023). We

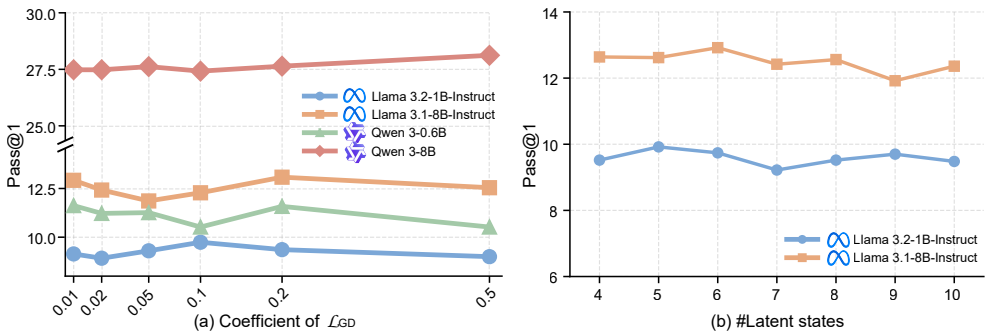


Figure 10: Parameter analyses on MATH. **(a)** Effect of coefficient of $\mathcal{L}_{\text{GD}}$ to $\mathcal{L}_{\text{Latent-LM}}$. **(b)** Effect of the number of hidden states.

keep samples with fewer than 200 total tokens following CODI (Shen et al., 2025). We provide the statistics of all datasets in Table 5.

Table 5: Statistics of datasets.

Dataset	Category	Usage	#Questions		CoT type	Avg. #CoT tokens		Answer type
			Train	Test		Train	Test	
GSM8K-Aug	Mathematical reasoning	Train / Test	385,620	1,319	Augmented CoT	20.49	24.93	Numerical
GSM8K-Hard	Mathematical reasoning	Test	–	1,319	–	–	–	Numerical
MATH	Mathematical reasoning	Train / Test	7,000	5,000	Human-labeled CoT	218.76	206.11	Numerical, formula & choice
NuminaMath-CoT (10%)	Mathematical reasoning	Train / Test	65,315	78	Human-labeled CoT	311.98	305.28	Numerical, formula & choice
CommonsenseQA-CoT	Commonsense logical reasoning	Train / Test	8,196	1,221	Generated CoT	84.00	–	Choice

Training setup. We evaluate our geometric decomposition-based objective under the self-distillation framework from CODI (Shen et al., 2025) and SIM-CoT (Wei et al., 2026). We follow their configurations unless otherwise stated, including LoRA fine-tuning (Hu et al., 2022), the colon token in "*The answer is:*" as the answer anchor, and a two-layer MLP projector with layer normalization that maps each output hidden state to the next latent input. During post-training, we exclude the last step of the teacher chain from geometric supervision because it is usually closer to the answer space than the reasoning space.

Hardware and Implementation. Experiments are run on up to 8× 80GB NVIDIA H100 GPUs. For long-CoT datasets such as MATH and NuminaMath-CoT, we use the DeepSpeed implementation of ZeRO-3 (Rajbhandari et al., 2020), which partitions model parameters, gradients and optimizer states across GPUs to reduce per-GPU memory usage. Table 6 lists the training hyperparameters. It is noteworthy that, except for the coefficient of $\mathcal{L}_{\text{GD}}$, **we do not search over any hyperparameters**.

C.2 Results on GSM8K

Tables 7 and 8 report quantitative results trained on GSM8K-Aug and evaluated on GSM8K-Aug and GSM8K-Hard, respectively. Unlike our results in Table 2, Latent-LM with only $\mathcal{L}_{\text{GD}}$ performs worse than with outcome-rewarding objectives. We attribute this to two factors. First, the training set of GSM8K has only $K = 2.6$ reasoning steps on average, leaving $\mathcal{L}_{\text{GD}}$ skipped on many short samples as the orthogonal subspace is only defined on chains with $r \geq 2$, *i.e.,* $K \geq 3$. Second, GSM8K-Aug compresses semantic logic in language CoT and only keeps arithmetic expressions, further limiting the ability of the proposed method to capture reasoning logic. Despite the two limitation factors, they mainly affect short, simple samples. With the compensation of answer distillation, $\mathcal{L}_{\text{GD}}$ remains effective on longer and more complex samples. This complementarity explains why their combination achieves the best Pass@1. The out-of-distribution results on GSM8K-Hard support the same conclusion.

C.3 Parameter Analyses

Figure 10(a) shows the effect of the coefficient of $\mathcal{L}_{\text{GD}}$ to $\mathcal{L}_{\text{Latent-LM}}$. Pass@1 remains largely stable as the coefficient increases, and Qwen 3-8B also shows a small gain. This indicates that $\mathcal{L}_{\text{GD}}$ has

Table 6: Hyperparameters in our quantitative experiments.

	Llama 3.2-1B -Instruct	Llama 3.1-8B -Instruct	Qwen 3-0.6B	Qwen 3-8B
Training epochs*	10	6	10	6
Training batch size	8	1	4	1
Gradient accumulation step	4	8	8	8
Training batch size (GSM8K-Aug)	32	8	32	8
Gradient accumulation step (GSM8K-Aug)	2	2	2	2
Learning rate	8e-4	1e-4	8e-4	1e-4
Weight decay	0.1	0.1	0.1	0.1
Precision	BF16	BF16	BF16	BF16
LoRA rank	128	128	128	128
LoRA α	32	32	32	32
MLP projector dimension	2,048	4,096	1,024	4,096
Warm-up ratio for $\mathcal{L}_{\text{Latent-LM}}$	0.03	0.03	0.03	0.03
Warm-up ratio for $\mathcal{L}_{\text{GD}}$	0.03	0.03	0.03	0.03
Coefficient for $\mathcal{L}_{\text{GD}}$	0.1	0.01	0.01	0.05
ε_1	1e-8	1e-8	1e-8	1e-8
ε_2	1e-8	1e-8	1e-8	1e-8

* We train all models for 3 epochs on CommonsenseQA.

Table 7: Results on GSM8K-Aug. Best latent reasoning results are in **bold**. "OOM" stands for out-of-memory on 8× 80 GB GPUs.

	Llama 3.2-1B-Instruct			Llama 3.1-8B-Instruct			Qwen 3-8B		
	Pass@1↑	Avg. #{T}↓	EFF↑	Pass@1↑	Avg. #{T}↓	EFF↑	Pass@1↑	Avg. #{T}↓	EFF↑
CoT-SFT	47.38	15.8	3.00	61.33	14.8	4.14	81.65	152.7	0.53
Latent-LM	35.56	6	5.93	43.59	6	7.26	44.73	6	7.46
+ answer distill (Shen et al., 2025)	52.62	6	8.77	58.98	6	9.83	62.09	6	10.35
+ answer distill + step decode (Wei et al., 2026)	51.63	6	8.60	OOM	–	–	OOM	–	–
+ $\mathcal{L}_{\text{GD}}$ (12)	34.57	6	5.76	44.50	6	7.42	50.79	6	8.46
+ answer distill + principal $\mathcal{L}^{\rightarrow}$ (7)	53.60	6	8.93	60.12	6	10.02	60.12	6	10.02
+ answer distill + orthogonal $\mathcal{L}^{\perp}$ (10)	54.66	6	9.11	59.06	6	9.84	48.07	6	8.01
+ answer distill + $\mathcal{L}_{\text{GD}}$ (12)	**55.65**	6	**9.28**	**62.77**	6	**10.46**	**64.25**	6	**10.71**

Table 8: Pass@1 on GSM8K-Hard (trained on GSM8K-Aug). Best latent reasoning results are in **bold**. "OOM" stands for out-of-memory on 8× 80 GB GPUs.

	Llama 3.2-1B-Instruct	Llama 3.1-8B-Instruct	Qwen 3-8B
CoT-SFT	9.40	17.74	46.93
Latent-LM	8.42	11.30	18.06
+ answer distill (Shen et al., 2025)	12.66	13.87	**21.00**
+ answer distill + step decode (Wei et al., 2026)	11.90	OOM	OOM
+ answer distill + principal $\mathcal{L}^{\rightarrow}$ (7)	**12.89**	15.31	20.43
+ answer distill + orthogonal $\mathcal{L}^{\perp}$ (10)	12.74	15.16	18.95
+ answer distill + $\mathcal{L}_{\text{GD}}$ (12)	**12.89**	**16.76**	**21.00**

little effect on the training stability of latent SFT. Together with Figure 12 below, we find that even a coefficient as small as 0.01 can align the principal direction and expand the orthogonal subspace pretty well. Figure 10(b) shows the effect of latent chain length. Consistent with Shen et al. (2025); Wei et al. (2026), a longer latent chain does not necessarily yield better reasoning performance. This does not mean that latent reasoning is limited to simple, short reasoning tasks; from the superposition theory (Zhu et al., 2025; 2026), it prefers to generate more informative latent states rather than extending the latent chain for more efficient complex reasoning.

C.4 Visualization of the Latent Reasoning Space (§4.1)

Figures 11 and 12 show more visualization results of the teacher and student chains in Figure 5. Interactive 3-D plots can be found in our repository.

D Related Work

Reasoning in latent spaces. Latent reasoning replaces part or all of language CoT with continuous latent states. They do not directly participate in decoding or next-token prediction. Some existing methods employ token-level latent states or alternate between latent and language CoT (Zhang et al.,

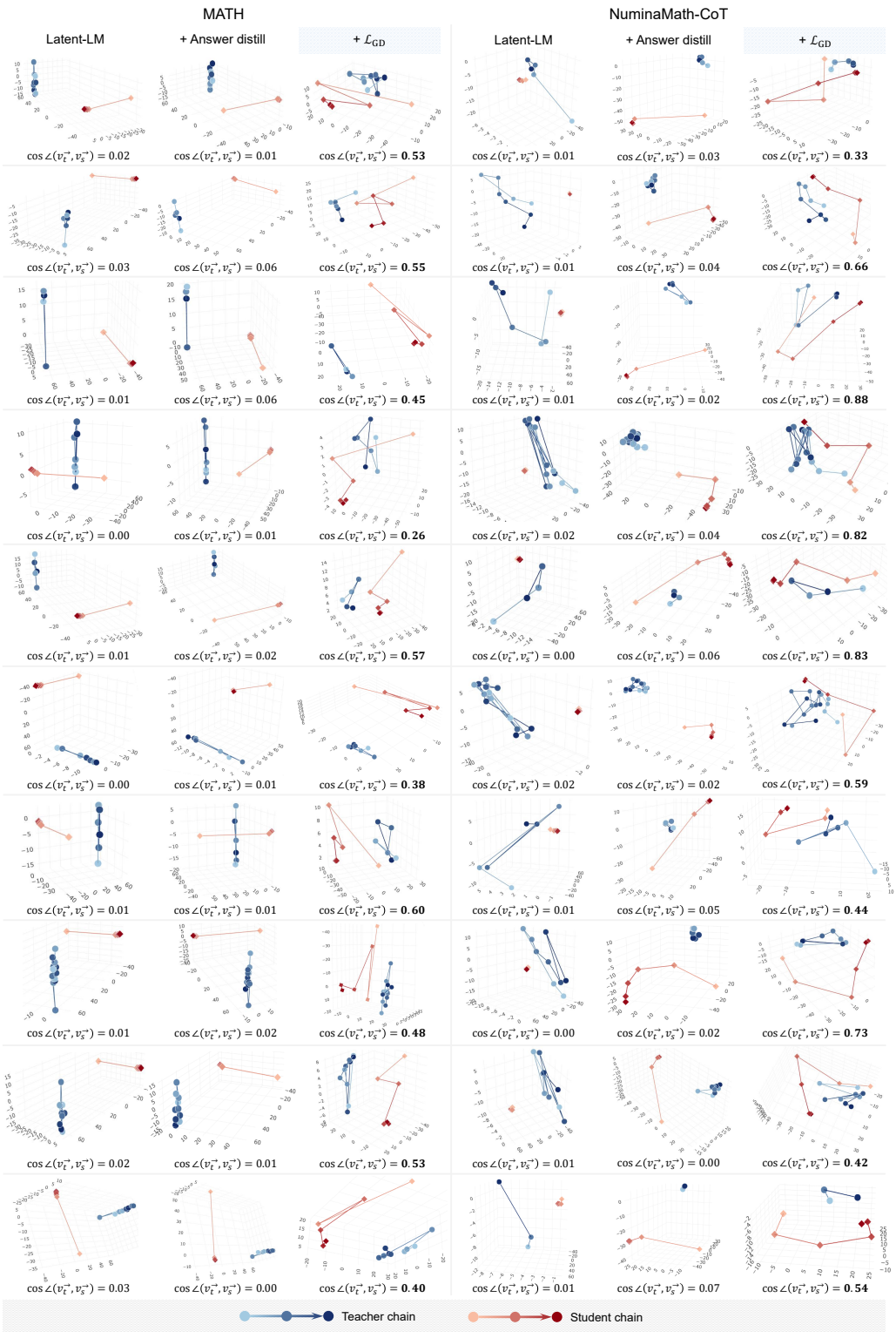


Figure 11: Visualization of latent reasoning spaces (Llama 3.2-1B-Instruct). $\cos\angle(v_t^{\rightarrow}, v_s^{\rightarrow})$ denotes the absolute cosine similarity between the teacher and student principal directions.

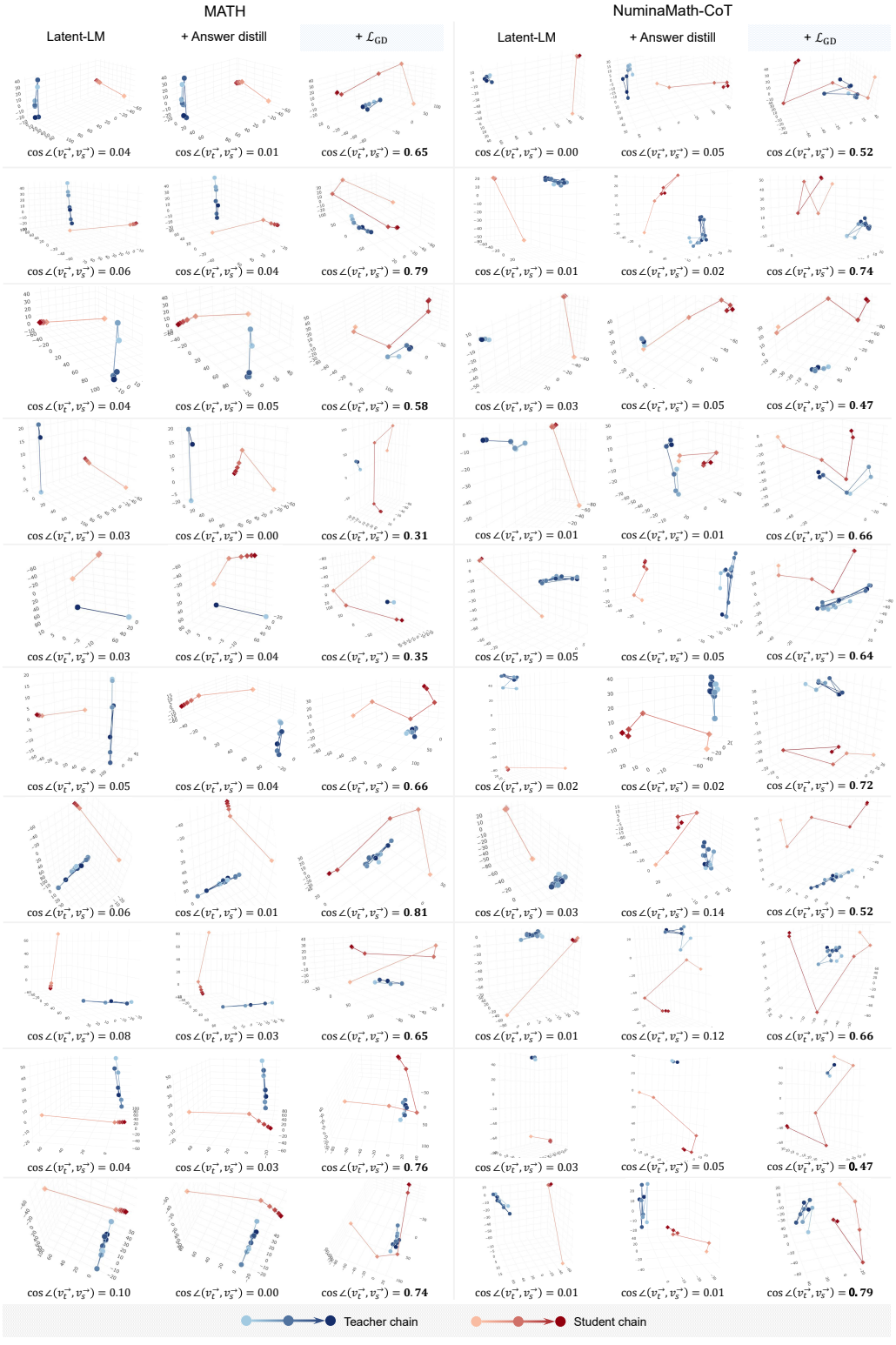


Figure 12: Visualization of latent reasoning spaces (Llama 3.1-8B-Instruct). $\cos\angle(v_t^{\rightarrow}, v_s^{\rightarrow})$ denotes the absolute cosine similarity between the teacher and student principal directions.

2025b; Xu et al., 2025; Shi et al., 2026; Yue et al., 2025). Another line performs compressed latent reasoning with a much shorter latent chain and no language CoT at inference (Hao et al., 2025; Cheng & Van Durme, 2024; Shen et al., 2025; Wei et al., 2026; Wu et al., 2025; Kuzina et al., 2026; He et al., 2025; Zhao et al., 2026). We focus on this line as it achieves better token efficiency.

Latent reasoning can be further divided by training strategies and objectives. (1) *Outcome-rewarding models* supervise whether the latent chain reaches a correct answer, including SFT objectives like latent language modeling and answer distillation (Shen et al., 2025; Wu et al., 2025). Some reinforcement learning objectives like variants of group relative policy optimization (Yue et al., 2025; Tan et al., 2025) are also outcome-rewarding but they are out of scope in this paper. As these objectives give limited guidance to the reasoning space, recent analyses have pointed out their limitations such as homogeneous latent states, semantic drift, and weak causal contribution of latent states (Zhang et al., 2025a; Li et al., 2025c; Song et al., 2025; Wei et al., 2026; Chen et al., 2026; Hao & Yang, 2026). (2) *Process-rewarding models* actively guide the latent reasoning process during training. For example, curriculum learning methods (Hao et al., 2025; Deng et al., 2024) gradually replace language CoT with latent states to explicitly and sequentially learn the role of each latent state. Alignment methods align the intermediate latent states with those from language CoT (Cheng & Van Durme, 2024; He et al., 2025; Kuzina et al., 2026), or decode the latent states into text and match the language CoT with language modeling (Wei et al., 2026; Ren, 2025; Zhao et al., 2026). However, few process-rewarding methods provide effective supervision from the perspective of reasoning geometry. A similar recent work LoRi (Solgi et al., 2026) uses SVD to align the dominant subspace for distillation-based latent SFT. However, it does not establish a connection between the reasoning geometry and reasoning logic. We advocate that the principal direction and the orthogonal subspace should play different supervisory roles in latent reasoning, as they are associated with different types of reasoning logic.

Reasoning geometry and reasoning logic. The awareness and utilization of reasoning logic is a key factor in enhancing reasoning ability. Early reasoning methods introduce search, branching and backtracking logic via structured CoT reasoning, such as trees (Yao et al., 2023; Wang et al., 2026) and graphs (Besta et al., 2024). A related motivation also comes from cognitive neuroscience: task-relevant abstract variables in the prefrontal cortex and hippocampus are often organized into low-dimensional, generalizable and nearly orthogonal geometries (Flesch et al., 2022; Weber et al., 2023; Bhandari et al., 2025). Recent studies on LLM representation geometry further connect reasoning and semantic logic with linear directions (Mikolov et al., 2013; Park et al., 2024; Modell et al., 2025), step-wise distances (Sun et al., 2026), trajectory curvature (Hosseini & Fedorenko, 2023; Pan & Woodard, 2026; Zhou et al., 2026; King et al., 2026), etc. Other works show that tokens with related semantics or similar functionality can exhibit shared relative directions or cluster patterns in the high-dimensional space (Lee et al., 2025; Yang et al., 2026). Most of these studies focus on *local geometry* of the latent reasoning chain, while our work studies the global geometry of a latent reasoning chain and its connection with reasoning logic through a geometric decomposition.

E Discussion and Limitations

E.1 Stability and Directionality of the Principal Direction

We use compact SVD to decompose the reasoning space $\mathcal{V} := \mathrm{span}\{v_1, \ldots, v_r\} \subset \mathbb{R}^d$ into the principal direction $\mathcal{V}^{\rightarrow} := \mathrm{span}\{v_1\}$ and its orthogonal (complement) subspace $\mathcal{V}^{\perp} := \mathrm{span}\{v_2, \ldots, v_r\} = \{a \in \mathcal{V} : v_1^{\top} a = 0\}$ in (12). The stability and directionality of our principal direction are guaranteed because of two reasons. (1) The compact SVD determines $\mathcal{V}^{\rightarrow}$ as an unoriented 1-D subspace. When $\sigma_1 > \sigma_2$, $\mathcal{V}^{\rightarrow}$ is unique, and its robustness to perturbations increases with the leading spectral gap $\sigma_1 - \sigma_2$. For the vast majority of samples in our analyses, $\sigma_1 > \sigma_2$ and often $\sigma_1 \gg \sigma_2$; see Figure 4. Thus, the principal direction is stable in most cases. For the few chains with $\sigma_1 \approx \sigma_2$, principal directions can be less reliable, but they still carry information of the reasoning threads, inferring negligible influence to latent SFT. (2) The directionality of v_1 stems from that of the reasoning sequence $\mathcal{Z}$ itself, which uniquely determines our principal direction; and thus, it is not necessary to strictly fix the orientation of $\mathcal{V}_t^{\rightarrow}$ and $\mathcal{V}_s^{\rightarrow}$ in our $\mathcal{L}_{\mathrm{GD}}$.

E.2 Broader Applications

Although we validate the proposed geometric decomposition only in distillation-based latent SFT, it has great potential to also be adapted to latent pretraining, reinforcement learning-based post-training and on-policy distillation across different base models. Related concentration phenomena have been observed in language model pretraining and post-training as representation collapse (Li et al., 2025c), in outcome-based reinforcement learning as a loss of reasoning diversity (Song et al., 2025), etc. Our strategy may provide an interpretable way to analyze such effects and guide related improvements.

E.3 Limitations

A key limitation of this work is that the understanding of reasoning logic in latent reasoning models remains difficult to verify directly in text. To the best of our knowledge, existing stepwise decoding methods such as Wei et al. (2026) decode only mathematical expressions and therefore do not reveal semantic logic. The two methods that decode latent states into language CoTs (Shen et al., 2026; Zhao et al., 2026) are designed mainly in vision-language settings. More importantly, these decoding methods do not guarantee that CoT text decoded from adjacent latent states forms coherent reasoning logic. Thus, decoding approaches that disentangle logical form from semantic logic are still underexplored for latent reasoning. Consequently, the geometric decomposition currently serves as a geometric interpretation of our proposed method; our empirical studies indicate its logic-preserving property rather than providing explicit textual evidence.