CoRD: On-Policy Distillation via Correction Reconstruction

Anonymous authors
Paper under double-blind review

Abstract

On-policy distillation helps language models improve reasoning through teacher feedback on their own responses, but selecting fewer supervised tokens can distort the collective learning signal. We aim to preserve this signal under a limited token supervision budget by accounting for how individual corrections reinforce or cancel one another. We introduce Correction Reconstruction Distillation (CoRD), which preserves the contributions of repeated correction directions to the average update. CoRD selects complementary token corrections in an output-head gradient representation and assigns signed weights to reconstruct the dense mean within their span, with theoretical bounds on approximation error in this representation. Across three student models and six mathematical reasoning benchmarks, including MATH-500 and AIME 2024/2025, CoRD improves macro mean@16 by 1.2–1.3 percentage points and pass@8 by 1.0–2.3 points over the strongest compared baselines under matched GPU-hour caps, using a 25% supervised-token budget. Coupling complementary selection with mean reconstruction allows CoRD to exploit redundancy while accounting for omitted tokens' aggregate contributions.

1 Introduction

On-policy distillation (OPD) enables language models to learn from teacher feedback on their own generated responses, addressing the distribution mismatch between fixed training demonstrations and student-generated trajectories (Agarwal et al., 2024). In reasoning post-training, it supplies dense token-level supervision throughout a response, providing guidance at the intermediate steps of generation (Jin et al., 2026). For long reasoning responses, making effective use of this feedback requires understanding which supervision can be removed without weakening learning.

Recent sparse OPD methods show that the value of supervision varies across tokens. TIP identifies informative positions through student uncertainty and teacher–student disagreement (Xu et al., 2026). TA-OPD further asks whether the teacher's preferred continuations are locally accessible to the student (Wang et al., 2026), while IER-OPD accounts for the reliability of gradients estimated from sampled tokens (Sheng et al., 2026). These advances identify useful or reliable positions, but the learning effect of a selected set also depends on how its signals reinforce, repeat, or offset one another.

As illustrated in Figure 1, several tokens may convey similar feedback, yet simply removing repeats can change their relative influence on the full response's learning signal. The central tension is that *redundancy creates an opportunity to reduce supervision, yet redundant signals can still affect the aggregate update*. Gradient coresets provide a relevant precedent: CRAIG approximates full-data gradients through weighted representatives (Mirzasoleiman et al., 2020), and GRAD-MATCH explicitly optimizes gradient approximation (Killamsetty et al., 2021). The open question is when redundancy within a generated response permits sparse supervision while preserving its collective learning signal. Answering it requires accounting for the interactions among token corrections.

We study whether a small subset of token losses can faithfully approximate the learning signal supplied by dense on-policy supervision. Our preliminary study finds that overlapping token corrections can coexist with substantial mean distortion after uniform de-duplication (Section 3). Motivated by this distinction, we introduce *Correction Reconstruction Distillation* (CoRD). CoRD selects complementary token corrections and reweights them to reconstruct the dense mean within their span, accounting for the contributions of omitted tokens. Its approximation guarantees concern a tractable

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.

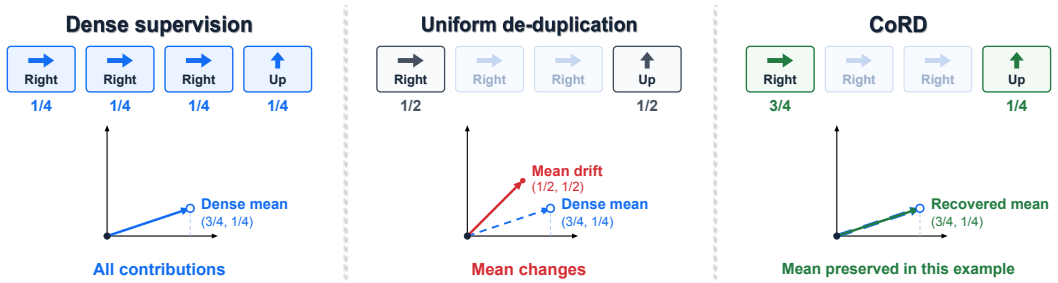

Figure 1: **Why selection needs reconstruction.** Uniform de-duplication can shift the mean by discarding repeated contributions. In this toy output-head example, CoRD recovers the dense mean by reweighting complementary token corrections.

output-head gradient proxy; fidelity to trainable LoRA gradients is evaluated empirically. The budget limits positions carrying weighted losses; all candidates still receive teacher and student scores.

Our contributions are threefold:

- **We reveal that de-duplication can distort the mean despite directional redundancy.** Controlled analyses associate larger gains in gradient fidelity with responses whose corrections are both compressible and strongly canceling.
- **We propose CoRD for complementary token selection and mean reconstruction.** We establish correction-space error bounds and sufficient token budgets, and assess fidelity to LoRA gradients and optimizer updates.
- **CoRD achieves better reasoning performance under matched compute budgets.** Across three student backbones and six mathematical reasoning benchmarks, CoRD improves macro mean@16 by 1.2–1.3 percentage points and pass@8 by 1.0–2.3 points over the strongest compared baseline for each backbone and metric at matched GPU-hour caps.

2 Related Work

Distillation objectives and rollout distributions. Sequence-level distillation uses teacher-generated targets (Kim & Rush, 2016), whereas GKD supervises student-visited states to reduce exposure mismatch (Agarwal et al., 2024). MiniLLM and DistiLLM variants refine divergence objectives and data utilization (Gu et al., 2024; Ko et al., 2024; 2025); SKD interleaves teacher/student sampling (Xu et al., 2025), and EOPD adapts the objective to teacher uncertainty (Jin et al., 2026). These approaches improve the supervision or rollout distribution. CoRD addresses the complementary problem of redundancy among token corrections by approximating the dense mean correction on student-generated trajectories.

Selective token supervision. Selection has progressed from uncertainty and disagreement in TIP (Xu et al., 2026) to local support compatibility in TA-OPD (Wang et al., 2026) and sampled-gradient reliability in IER-OPD (Sheng et al., 2026). TSD-KD combines selective supervision with teacher-ranked token candidates (Kim & Baek, 2026), while FiRe-OPD couples trajectory filtering with soft token weighting (Li et al., 2026). These criteria assess individual signals. CoRD also accounts for their aggregate correction: it selects complementary directions and fits signed weights to preserve the dense mean within their span. Prefix-OPD controls supervision through early generation stopping (Zhang et al., 2026); CoRD selects across completed responses and scores all candidates.

Gradient coresets and correction geometry. CRAIG approximates full gradients through weighted representatives (Mirzasoleiman et al., 2020); GRAD-MATCH directly minimizes gradient mismatch using pursuit and supports last-layer proxies (Killamsetty et al., 2021). Studies of OPD also report concentrated cumulative updates (Shen et al., 2026; Yu et al., 2026). Our analysis examines redundancy at the finer scale of token corrections within a trajectory. Building on weighted-gradient approximation, CoRD couples residual-based coverage with signed mean reconstruction through

one pivoted-Cholesky factorization of an implicit token-correction kernel. This dynamic OPD construction gives an exact projection in the output-head proxy space; transfer to LoRA updates is evaluated empirically.

3 PRELIMINARY STUDY

Can dense OPD supervision be compressed without changing its aggregate correction? We examine 256 student-generated responses using a Qwen3-0.6B student and a frozen Qwen3-4B teacher. Each token correction c_t is the reverse-KL gradient of the student output head at fixed hidden states. The study diagnoses these corrections without applying CoRD or updating the student. Appendix B.4 provides the protocol and metric definitions; Appendix C.1 gives additional statistics.

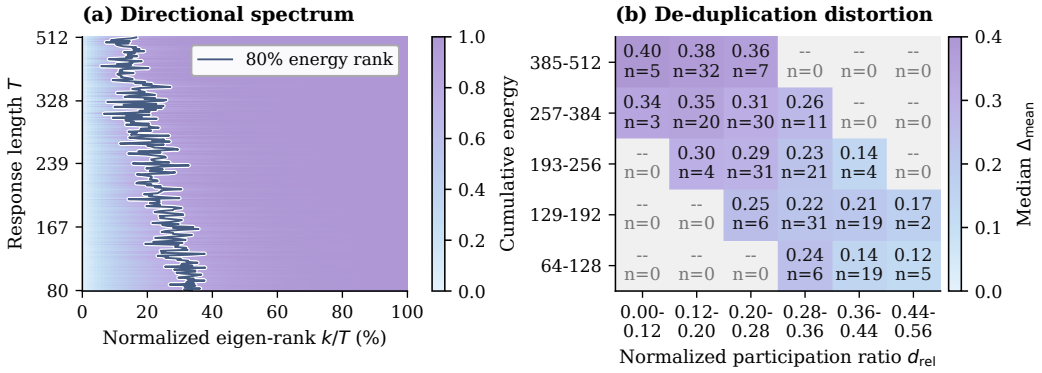

Figure 2: **Directional redundancy coexists with mean distortion under de-duplication. (a)** Cumulative spectral energy of unit-normalized corrections, with trajectories ordered by length; the curve marks the normalized rank explaining 80% of the energy. **(b)** Relative mean distortion after uniformly averaging cluster medoids, grouped by length and normalized participation ratio d_{rel}. Cells show median distortion and sample count; gray denotes no observations.

Directional redundancy. We normalize corrections to unit length and inspect their cosine Gram spectrum (Figure 2a). Its normalized participation ratio d_{rel} measures effective dimension relative to response length; smaller values indicate stronger spectral concentration. The median is 0.27 (IQR 0.20–0.34). The mean fraction of eigen-directions needed to explain 80% of the energy decreases from 32.6% for 64–128-token responses to 13.4% for 385–512-token responses. This concentration persists after removing correction magnitudes.

De-duplication changes the mean. As a diagnostic, we cluster correction directions and uniformly average one medoid per cluster, using the original correction magnitudes. Relative mean distortion Δ_{mean} measures the Euclidean change from the dense token-average correction, normalized by its norm with a stabilizer (Appendix B.4.3). Its median increases from 0.15 to 0.38 across the same short and long response groups (Figure 2b). Thus, directional concentration can coexist with substantial changes to the aggregate correction under uniform de-duplication.

These findings motivate CoRD's two design objectives: select complementary correction directions and preserve the dense trajectory's mean within their span. We therefore couple directional selection with mean reconstruction.

4 METHOD

Correction Reconstruction Distillation (CoRD) approximates a response's mean distillation correction with a weighted subset of token losses. It selects complementary corrections in an output-head gradient space, then fits the full mean within their span. Selection covers distinct directions; reconstruction accounts for repeated and opposing corrections. Both steps share one pivoted-Cholesky factorization (Figure 3).

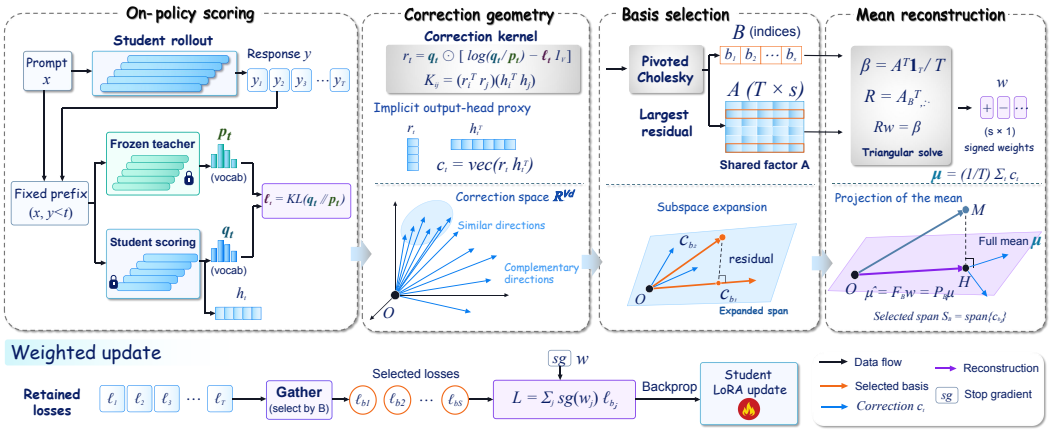

Figure 3: **Overview of CoRD.** Both models score fixed prefixes of a student-generated response for reverse-KL token losses. Detached logit gradients and hidden states define the output-head correction kernel. Pivoted Cholesky selects ordered positions B by residual energy and forms factor A; a triangular solve yields signed weights w and the proxy mean projection $\widehat{\mu} = F_B w = P_B \mu$. The bottom path applies detached w to selected losses in the retained scoring graph, updating only LoRA; the teacher, student base weights, and output head remain frozen.

4.1 On-policy correction geometry

Let π_θ be a student with trainable low-rank adaptation (LoRA) parameters θ in its attention and MLP projections (Hu et al., 2022) and fixed base weights. A frozen teacher π_{T} shares its vocabulary of size V. Following on-policy distillation (Agarwal et al., 2024), the student generates one response y per prompt x. Both models score the same prefixes: at each valid position $t \in \{1, \dots, T\}$, let $p_t = \pi_{\mathrm{T}}(\cdot \mid x, y_{<t})$ and $q_t = \pi_\theta(\cdot \mid x, y_{<t})$. The token loss and dense trajectory loss use full-vocabulary reverse KL at temperature one:

$$\ell_t(\theta) = \sum_{u=1}^{V} q_{t,u} \log \frac{q_{t,u}}{p_{t,u}}, \qquad \mathcal{L}_{\mathrm{full}}(\theta) = \frac{1}{T} \sum_{t=1}^{T} \ell_t(\theta). \tag{1}$$

Both distributions use finite logits and a common token-ID mapping, and generated prefixes stay fixed during differentiation. Valid positions include generated EOS tokens but exclude prompt tokens, padding, and appended special tokens. We discard empty responses.

The student's bias-free output head is $z_t = W h_t$, with input $h_t \in \mathbb{R}^d$ and fixed weights $W \in \mathbb{R}^{V \times d}$. We represent each token's correction by

$$r_{t,u} = \frac{\partial \ell_t}{\partial z_{t,u}} = q_{t,u} \left(\log \frac{q_{t,u}}{p_{t,u}} - \ell_t \right), \qquad c_t = \mathrm{vec}(r_t h_t^\top) \in \mathbb{R}^{Vd}. \tag{2}$$

Here $r_t \in \mathbb{R}^V$ is the logit gradient, and c_t is the direct head correction at fixed h_t. We use c_t as a proxy to select losses for adapter training. The correction Gram matrix requires no explicit outer products:

$$K_{ij} = c_i^\top c_j = (r_i^\top r_j)(h_i^\top h_j), \qquad K = F^\top F \succeq 0, \tag{3}$$

where $F = [c_1, \dots, c_T] \in \mathbb{R}^{Vd \times T}$. The diagonal measures correction energy; off-diagonal entries retain signed alignment. We preserve correction magnitudes in both selection and reconstruction. Appendix D.1.1 derives the gradient and kernel identity.

4.2 Selecting a correction basis

For retention fraction $\rho \in (0, 1]$, the per-trajectory budget is $m = \max\{1, \lfloor \rho T \rfloor\}$. Let $B = (b_1, \dots, b_s)$ list the selected positions in order, and let F_B contain their correction columns. The orthogonal projector onto their span is $P_B = F_B (K_{BB})^{-1} F_B^\top$ when these columns are linearly independent. We score each candidate by the correction energy outside this span:

$$v_t(B) = \| (I - P_B) c_t \|_2^2 = K_{tt} - K_{tB} (K_{BB})^{-1} K_{Bt}. \tag{4}$$

Initially, B is empty, $P_B = 0$, and $v_t = K_{tt}$. Each step selects the unselected position with the largest v_t, breaking ties by the smallest index. We stop after m selections or when the largest residual reaches the numerical-rank tolerance, giving $s \leq m$.

Selection scores corrections individually to cover large directions hidden by mean cancellation; reconstruction captures their signed aggregate. Pivoted Cholesky (Harbrecht et al., 2012) forms $A \in \mathbb{R}^{T\times s}$ with one new kernel column per pivot; row t gives c_t's coordinates along the selected orthonormal directions. The factor supplies residual scores and reconstruction coefficients without storing F or the basis. Appendix D.1.2 gives the recurrence; Appendix D.1.4 contrasts coverage with direct mean-error minimization.

4.3 Reconstruction and the student update

Reconstruction preserves the contributions of all tokens within the selected span. For example, averaging representatives of c_a, c_a, c_b changes $(2c_a + c_b)/3$ to $(c_a + c_b)/2$. We therefore fit the full mean correction $\mu = F\mathbf{1}_T/T$, where $\mathbf{1}_T$ is the all-ones vector. The selected independent columns have unique least-squares coefficients:

$$w = \arg\min_{u\in\mathbb{R}^s} \|F_B u - \mu\|_2^2, \qquad Rw = A^\top \mathbf{1}_T/T, \qquad R = A^\top_{B,:}. \tag{5}$$

Since $R \in \mathbb{R}^{s\times s}$ is upper triangular in pivot order, reconstruction needs one triangular solve. The right-hand side averages coordinates over all original positions, retaining repeated corrections' contributions. Appendix D.1.3 proves equivalence to the least-squares problem.

We compute B, w at the current parameters θ_0 and hold them fixed for one update. With sg denoting stop-gradient, the trajectory loss is

$$\mathcal{L}_{\mathrm{CoRD}}(\theta;\theta_0) = \sum_{j=1}^{s} \mathrm{sg}(w_j)\ell_{b_j}(\theta). \tag{6}$$

These signed weights enter Equation (6) directly, without positivity or unit-sum constraints. For a batch of nonempty responses indexed by a, we use $\mathcal{L}_{\mathrm{batch}} = \sum_a T_a \mathcal{L}_{\mathrm{CoRD},a} / \sum_a T_a$ to preserve the dense reference's token normalization. We clip the gradient norm to 1 and take one AdamW step (Loshchilov & Hutter, 2019) to update θ, then generate fresh responses and recompute the bases and weights.

At full budget ($m = T$), every position receives weight $1/T$, recovering the dense loss. For a zero kernel, we sample a uniform subset of size m with weights $1/m$. Appendix A gives the training procedure and numerical rules.

In exact arithmetic, the reconstructed correction $\widehat{\mu} = F_B w$ satisfies

$$\widehat{\mu} = P_B \mu, \qquad \|\mu - \widehat{\mu}\|_2^2 \leq \frac{1}{T}\sum_{t=1}^{T} v_t(B). \tag{7}$$

Spanning all output-head corrections recovers their mean exactly; Subsection 4.4 sharpens the bound using the kernel spectrum, and Appendix D.2.5 derives sufficient token budgets. Agreement of the adapter gradient $\sum_j w_j \nabla_\theta \ell_{b_j}$ with the dense gradient depends on position-specific parameter Jacobians and coefficient magnitudes (Appendix D.1.6); Appendix D.1.5 covers single-trajectory head descent and batching effects on alignment.

Like CRAIG and GRAD-MATCH (Mirzasoleiman et al., 2020; Killamsetty et al., 2021), CoRD uses weighted gradient subsets. One factorization supports coverage selection ($O(Ts(V+d)+Ts^2)$) and mean reconstruction ($O(Ts + s^2)$), excluding model evaluation. Both models score all candidates; only weighted-loss positions are budgeted (storage and backward costs: Appendix A.5).

4.4 Compressibility of the correction space

The kernel spectrum bounds how many tokens suffice to reconstruct the mean output-head correction. For a fixed trajectory, let P_m project onto the span after m greedy selections and write $\mathcal{E}_m = \|\mu - P_m\mu\|_2^2$.

Theorem 1 (Spectral compression). *Let $\lambda_1 \geq \cdots \geq \lambda_T \geq 0$ be the eigenvalues of the unnormalized kernel $K = F^\top F$ and $r_\star = \operatorname{rank}(F)$. Under greedy selection and signed least-squares reconstruction in exact arithmetic, for $0 \leq m < r_\star$ set $N_m = T - m$ and $L_i = (T-i)^{-1} \sum_{j=i+1}^{T} \lambda_j$ for $0 \leq i < m$. With empty products equal to one, define $\mathcal{Q}_m(K) = \min_{1 \leq n \leq m+1} \min_{0 \leq k < n} \left[\prod_{j=1}^{n} \lambda_j / \left(N_m \prod_{i=0}^{k-1} L_i \right) \right]^{1/(n-k)}$. Then $\mathcal{E}_m \leq (N_m/T)^2 \mathcal{Q}_m(K)$, and at most $r_\star$ selected corrections recover the mean exactly.*

Proof sketch. Let $\delta_i = \max_t \|(I - P_i)c_t\|_2^2$ and $s = T^2 \mathcal{E}_m / N_m^2$. Selected corrections have zero residual. Combining this fact with a variational trace bound gives $\delta_i \geq \max\{s, L_i\}$ for $i < m$. Project K onto the first $n-1$ selected coordinate vectors and the normalized indicator of the unselected positions. The resulting determinant lies between $N_m s^{n-k} \prod_{i=0}^{k-1} L_i$ and $\prod_{j=1}^{n} \lambda_j$. Taking roots and minimizing gives the bound; Appendix D.2.4 gives the full proof. □

Spectral-decay token budgets, sharpness results, and numerical-error certificates appear in Appendices D.2.5–D.2.7.

5 Experiments

We evaluate reasoning quality and cost (**RQ1**, Section 5.2), the roles of selection and reconstruction (**RQ2**, Section 5.3), and token choices and favorable correction geometries (**RQ3**, Section 5.4).

5.1 Experimental setup

Benchmarks and metrics. We report mean@16 accuracy and pass@8 on MATH-500, AIME 2024/2025, AMC 2023, Minerva Math, and OlympiadBench (Jin et al., 2026; Li et al., 2026). Appendix B.1 details decoding and fidelity metrics.

Comparison protocol. Baselines include dense OPD, token selection, gradient matching, and recent distillation methods (Jin et al., 2026; Kim & Baek, 2026; Zhang et al., 2026). We match initialization, prompts, trainable parameters, and evaluation within each model pair, separating fixed-token-budget comparisons from end-to-end comparisons that preserve each method's objective and sampling mechanism (Appendix B.2).

Backbones and implementation. We train Qwen3-1.7B, Qwen3-4B, and Qwen2.5-1.5B-Instruct students (Yang et al., 2025; Qwen et al., 2025) using rank-32 LoRA (Hu et al., 2022), eight NVIDIA A100 80GB GPUs, three seeds, and $\rho = 0.25$ unless varied. RQ1 covers all three students; controlled studies use Qwen3-8B $\rightarrow$ Qwen3-4B. Appendix B.3 details pairings, hyperparameters, numerical settings, statistics, and cost measurement; Appendix C.3 contains the full controlled results.

5.2 RQ1: Reasoning quality and training cost

Accuracy at matched compute. At matched end-to-end GPU-hour caps (Table 1), CoRD has the highest mean among trained students in all six columns, exceeding the strongest baseline by 1.3, 1.2, and 1.3 percentage points in macro mean@16 and 2.3, 1.0, and 1.4 points in pass@8 on Qwen3-1.7B, Qwen3-4B, and Qwen2.5-1.5B, respectively (aggregation: Appendix C.2).

Quality at a fixed supervision budget. At $\rho = 0.25$ and 1,000 updates (Table 2), CoRD improves macro mean@16 over the strongest competing selector by 1.7, 2.0, and 2.4 points, respectively. On Qwen3-4B, 23.87% of loss weights are nonzero; GPU-hours fall from 412.6 to 397.3 (3.7%) and peak memory from 72.8 to 71.4 GB/GPU relative to Dense-RKL. Selection overhead and continued scoring of all candidates limit runtime savings (Table 14).

Takeaway (RQ1). CoRD improves reasoning quality under both compute and supervision constraints, with modest measured end-to-end savings.

Table 1: **Main results under matched compute budgets.** M16/P8 are six-benchmark macro mean@16/pass@8 (%) under matched GPU-hour caps within each model pair. Trained-student scores are mean ± SD over three training seeds; reference checkpoints omit SD. Bold/underline mark best/second-best trained-student means, including ties. Appendix C.2 gives settings and per-benchmark results.

	Qwen3-1.7B		Qwen3-4B		Qwen2.5-1.5B	
Method	M16 ↑	P8 ↑	M16 ↑	P8 ↑	M16 ↑	P8 ↑
Teacher	71.4	86.4	71.4	86.4	49.5	67.6
Initial student	46.2	64.9	59.5	77.2	29.4	46.6
SeqKD (Kim & Rush, 2016)	47.6±0.8	65.8±1.3	60.2±1.0	77.9±1.3	30.7±0.8	47.8±1.5
Teacher-response KD (Hinton et al., 2015)	48.3±0.8	66.1±0.8	60.8±0.8	77.7±0.2	31.3±0.8	49.0±2.4
Dense-RKL (Agarwal et al., 2024)	52.1±1.0	70.7±0.4	63.4±0.9	79.4±1.0	34.9±0.9	52.1±0.1
GKD (FKL) (Agarwal et al., 2024)	51.4±0.9	69.8±0.4	62.8±0.9	79.6±0.9	34.3±1.0	52.4±1.7
GKD (JSD) (Agarwal et al., 2024)	51.6±0.8	71.1±1.2	63.1±0.7	80.3±1.0	34.7±1.1	52.7±1.5
MiniLLM (Gu et al., 2024)	50.3±0.8	69.2±1.3	62.0±0.9	79.7±1.6	33.4±1.0	51.4±1.7
SKD (Xu et al., 2025)	50.9±0.9	70.2±0.6	62.3±1.1	79.7±1.3	33.9±1.0	51.5±1.9
DistiLLM (Ko et al., 2024)	50.6±1.0	69.1±2.3	62.2±0.9	79.3±1.0	33.7±0.8	51.9±1.2
DistiLLM-2 (Ko et al., 2025)	51.8±1.0	70.5±0.9	63.2±0.9	79.5±2.3	35.0±0.8	52.9±1.2
Sampled OPD (Jin et al., 2026)	51.5±0.8	70.4±1.8	63.0±1.0	80.1±1.6	34.5±1.0	52.0±2.1
EOPD (Jin et al., 2026)	53.0±0.9	71.2±1.1	64.4±0.9	80.4±1.1	36.1±1.0	55.4±1.5
TSD-KD (Kim & Baek, 2026)	52.3±1.0	71.3±0.6	63.7±0.9	80.1±1.8	36.5±1.0	55.6±2.7
Prefix-OPD (Zhang et al., 2026)	51.1±0.8	70.3±0.9	62.6±0.8	79.1±0.6	34.2±0.9	53.2±1.8
TIP (Xu et al., 2026)	53.5±1.1	71.7±1.8	65.1±0.9	80.9±1.0	36.6±0.9	54.6±0.7
TA-OPD (Wang et al., 2026)	53.8±0.8	72.2±2.4	65.5±1.1	82.3±0.7	36.9±0.8	55.0±1.4
FiRe-OPD (Li et al., 2026)	54.2±0.9	72.5±1.7	66.0±1.0	82.2±1.3	37.2±0.9	55.6±2.1
CoRD (ours)	**55.5±1.0**	**74.8±0.3**	**67.2±0.8**	**83.3±0.9**	**38.5±1.1**	**57.0±1.7**

5.3 RQ2: Component roles and update fidelity

Selection and reconstruction both contribute. With the same objective and 1,000 updates (Table 9), replacing uniform selection with CoRD raises macro mean@16 from 61.1 to 64.6 under uniform weights. Signed least-squares (LS) reconstruction improves uniform selection to 63.5 and CoRD selection to 66.9, the highest mean among these controls.

From proxy corrections to trainable gradients. Figure 4 compares the output-head proxy and actual LoRA gradients on shared trajectories from 256 held-out prompts per seed, at initialization and Dense-RKL update 1,000. Panel (a) reports relative mean-correction error e_μ and unexplained correction-energy fraction η_{cov}; panels (b,c) report cosine similarity c_g and relative error e_g against dense trajectory and length-weighted batch gradients before clipping (protocol: Appendix B.5.2).

CoRD selection attains the joint-lowest η_{cov} at both states; CoRD/uniform and CoRD/LS share the selected span and thus tie. CoRD/LS has the lowest trajectory- and batch-gradient errors, with batch cosine similarities of 0.984 and 0.981. The signed mean-pursuit comparison shows why low correction error alone is insufficient (Appendix C.3.2).

From gradients to optimizer steps. From cloned AdamW states, CoRD/LS achieves parameter-increment cosine similarities of 0.988 and 0.985 to the dense step at initialization and update 1,000, respectively. This alignment persists after gradient clipping and AdamW preconditioning, supporting local update fidelity beyond the raw-gradient comparison. Signed reconstruction has no failed numerical checks in 768 attempts per state (Table 13).

Takeaway (RQ2). Coverage complements mean reconstruction; gradient and optimizer agreement supports local fidelity beyond the output-head proxy.

Table 2: **Token-selection accuracy and training cost.** Selectors share Dense-RKL's objective, LoRA modules, data, and 1,000 updates. Accuracy is six-benchmark macro mean@16 (%; mean ± SD across training seeds). ρ caps supervision; Qwen3-4B measurements report nonzero loss-weight fraction $f_{\neq 0}$, end-to-end GPU-hours, and peak allocated memory per GPU. Bold/underline mark best/second-best in arrowed columns; † denotes proxy-token adaptations. Appendix B.2 details selectors and costs.

		Macro mean@16 ↑			Qwen3-4B: measured cost		
Method	ρ	Qwen3-1.7B	Qwen3-4B	Qwen2.5-1.5B	$f_{\neq 0}$ (%)	GPU-h ↓	GB/GPU ↓
Dense-RKL (Agarwal et al., 2024)	1.00	52.1±1.0	63.4±0.9	34.9±0.9	100.00	412.6	72.8
Uniform	0.25	50.2±0.7	61.1±1.1	32.8±0.8	24.95	383.4	69.7
Evenly spaced	0.25	50.5±1.0	61.5±0.6	33.1±1.2	24.96	<u>382.9</u>	<u>69.6</u>
Prefix-mask	0.25	48.9±0.8	59.8±1.3	31.7±0.7	24.95	**382.7**	**69.5**
Entropy top-m	0.25	51.2±1.1	62.3±0.7	33.7±0.9	24.96	386.8	70.0
Entropy sampling	0.25	50.8±0.6	62.0±1.0	33.4±1.2	24.95	388.1	70.1
Reverse-KL top-m	0.25	51.6±0.9	62.7±0.6	34.0±0.7	24.96	385.9	69.9
Correction-norm top-m	0.25	51.9±1.2	63.0±0.8	34.2±1.1	24.95	391.5	70.5
TIP-budget (Xu et al., 2026)	0.25	52.3±0.7	63.5±1.1	34.5±0.8	24.96	394.8	70.7
TA-OPD-budget (Wang et al., 2026)	0.25	52.6±1.0	63.9±0.7	34.8±1.2	24.95	399.6	70.9
CRAIG† (Mirzasoleiman et al., 2020)	0.25	52.0±0.8	63.3±1.3	34.3±0.9	24.95	438.2	73.4
GRAD-MATCH† (Killamsetty et al., 2021)	0.25	52.8±1.1	64.1±0.8	35.0±0.6	22.94	452.7	73.8
Signed mean pursuit	0.25	<u>53.5±0.9</u>	<u>64.9±0.7</u>	<u>35.7±1.0</u>	23.61	447.1	73.6
CoRD (ours)	0.25	**55.2±0.7**	**66.9±0.9**	**38.1±1.0**	23.87	397.3	71.4

Relative fidelity	**(a) Correction space**				**(b) Trajectory gradient**				**(c) Batch gradient**			
Lower Higher	e_μ ↓		η_{cov} ↓		c_g^{traj} ↑		e_g^{traj} ↓		c_g^{batch} ↑		e_g^{batch} ↓	
Method / weights	Init.	1k	Init.	1k	Init.	1k	Init.	1k	Init.	1k	Init.	1k
Uniform / uniform	0.463	0.428	0.331	0.323	0.821	0.801	0.594	0.621	0.879	0.861	0.482	0.522
Uniform / LS	0.204	0.198	0.331	0.323	0.912	0.899	0.410	0.408	0.951	0.938	0.319	0.352
Norm top-m / uniform	0.402	0.365	0.258	0.229	0.867	0.846	0.512	0.548	0.918	0.902	0.402	0.445
GRAD-MATCH†	0.151	0.136	0.304	0.282	0.923	0.911	0.398	0.430	0.959	0.948	0.287	0.321
Signed mean pursuit	0.083	0.069	0.391	0.367	0.931	0.918	0.377	0.405	0.957	0.951	0.296	0.316
CoRD / uniform	0.312	0.281	0.236	0.217	0.909	0.897	0.427	0.462	0.950	0.940	0.324	0.347
CoRD / LS	**0.158**	**0.144**	**0.236**	**0.217**	**0.961**	**0.954**	**0.373**	**0.350**	**0.984**	**0.981**	**0.181**	**0.199**

Figure 4: **CoRD combines correction fidelity with accurate LoRA gradients.** Qwen3-4B, $\rho = 0.25$. Cells show means across three seeds at initialization (Init.) and Dense-RKL update 1,000 (1k). Color is min–max scaled within each metric over all methods and both states; deeper purple means better fidelity. All values are dimensionless. Bold identifies CoRD/LS. Table 11 gives the means and sample SDs; † denotes a proxy-token adaptation.

5.4 RQ3: Selection behavior and favorable correction geometry

5.4.1 Which tokens enter the basis?

Figure 5 analyzes 120,426 tokens from 240 held-out trajectories at Dense-RKL update 1,000, using within-trajectory reverse-KL percentiles p_t^{RKL} and recorded residual-novelty percentiles p_t^{nov} (protocol: Appendix B.6).

CoRD selects 96.69% of high-novelty, moderate-disagreement tokens ($p_t^{\text{nov}} \geq 0.80$, $0.40 \leq p_t^{\text{RKL}} < 0.65$), but none of the 5,462 high-disagreement, low-novelty tokens ($p_t^{\text{RKL}} \geq 0.80$, $p_t^{\text{nov}} < 0.25$). This pattern reflects residual-based selection: disagreement matters alongside novelty relative to already selected directions.

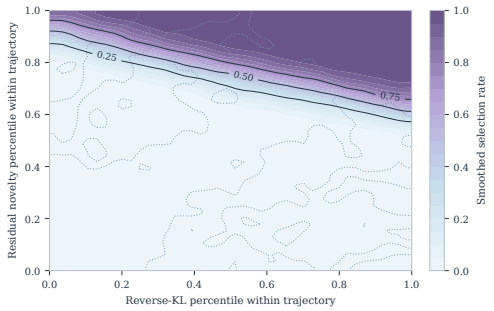

Figure 5: **Token selection reflects both novelty and disagreement.** Across 120,426 tokens from 240 trajectories, color and solid contours show smoothed selection rates; dotted contours show token density. Selection concentrates at high novelty; the same selection rate is reached at lower novelty as disagreement increases. Both axes use within-trajectory percentiles. Appendix B.6 gives estimation details.

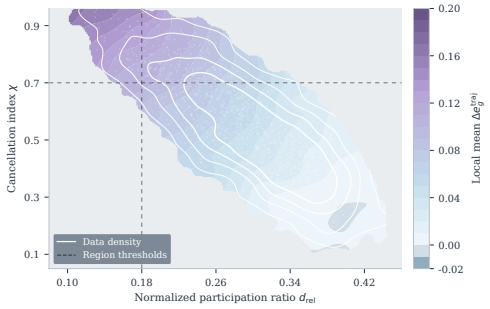

Figure 6: **CoRD gains are largest in compressible, high-cancellation regimes.** Across 1,536 trajectory-state records, color shows the local mean reduction in LoRA trajectory-gradient error relative to Uniform/LS. White contours show density; gray masks indicate insufficient support. Dashed lines mark the low-d_{rel}, high-χ region summarized in the text. Table 12 reports unsmoothed regional means and medians; Appendix B.5.3 gives estimation details.

5.4.2 Which response geometries favor selection?

Setup. We reuse the 1,536 trajectory-state records from Section 5.3. For corrections c_t (Eq. (2)), smaller normalized participation ratio d_{rel} (Eq. (11)) indicates greater directional compressibility. Cancellation is measured by

$$\chi = 1 - \frac{\| \sum_{t=1}^{T} c_t \|_2}{\sum_{t=1}^{T} \|c_t\|_2 + 10^{-12}}. \tag{8}$$

Larger χ indicates stronger cancellation. Holding signed LS reconstruction and $\rho = 0.25$ fixed, we measure the selection gain as

$$\Delta e_g^{\mathrm{traj}} = e_g^{\mathrm{traj}}(\mathrm{Uniform/LS}) - e_g^{\mathrm{traj}}(\mathrm{CoRD/LS}), \tag{9}$$

where e_g^{traj} is relative LoRA gradient error against the dense trajectory gradient before clipping; positive gains favor CoRD/LS.

Gains depend on the joint geometry. Across both states and all seeds (Figure 6), unsmoothed gains at $d_{\mathrm{rel}} \leq 0.18$ and $\chi \geq 0.70$ have mean 0.1358 and median 0.1362 ($n = 74$; IQR 0.1183–0.1530). For $d_{\mathrm{rel}} \geq 0.30$ and $\chi \leq 0.35$, the corresponding mean and median are 0.0021 and 0.0007 ($n = 218$). This contrast holds at each state (Table 12); all 135 negative-gain records are retained. Compressibility and cancellation co-vary, so the landscape describes their joint association with gains (Appendix B.5.3).

Takeaway (RQ3). Selection favors novel corrections and yields larger gradient-fidelity gains in compressible responses with strong cancellation.

6 Conclusion

CoRD preserves a full response's collective influence under sparse on-policy supervision by selecting correction directions and reconstructing the mean correction, with spectral guarantees in the output-head proxy space. Experiments across three student backbones and six mathematics benchmarks show improved reasoning accuracy, while controlled analyses show closer agreement with dense LoRA gradients. These findings support evaluating sparse supervision by collective update fidelity alongside individual-token informativeness. An open question is how correction geometry can guide the allocation of a shared supervision budget across responses as training progresses.

AI USE STATEMENT

We used generative AI tools to assist with organizing and reviewing related literature, checking mathematical proofs, refining mathematical notation and exposition, improving the presentation of experimental results, and drafting and editing the manuscript. The authors are responsible for evaluating AI-generated suggestions and take full responsibility for the accuracy, originality, and integrity of the final paper, including its mathematical claims, experimental results, references, and conclusions.

REFERENCES

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos Garea, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/5be69a584901a26c521c2b51e40a4c20-Abstract-Conference.html.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, et al. Evaluating Large Language Models Trained on Code, 2021. URL https://arxiv.org/abs/2107.03374.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, et al. Training Verifiers to Solve Math Word Problems, 2021. URL https://arxiv.org/abs/2110.14168.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. MiniLLM: Knowledge distillation of large language models. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/8ac015d409635f196f9e3e9dcfb9a94e-Abstract-Conference.html.

Helmut Harbrecht, Michael Peters, and Reinhold Schneider. On the low-rank approximation by the pivoted Cholesky decomposition. *Applied Numerical Mathematics*, 62(4):428–440, 2012. doi: 10.1016/j.apnum.2011.10.001. URL https://edoc.unibas.ch/entities/publication/d4eba30a-9b48-45ca-a4cb-915bfe3a83ec.

Chaoqun He, Renjie Luo, Yuzhuo Bai, Shengding Hu, Zhen Leng Thai, Junhao Shen, et al. OlympiadBench: A Challenging Benchmark for Promoting AGI with Olympiad-Level Bilingual Multimodal Scientific Problems, 2024. URL https://arxiv.org/abs/2402.14008.

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, 2021. URL https://arxiv.org/abs/2103.03874.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. doi: 10.48550/arXiv.1503.02531. URL https://arxiv.org/abs/1503.02531.

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 *International Conference on Learning Representations*, 2022. URL https://arxiv.org/abs/2106.09685.

Woogyeol Jin, Taywon Min, Yongjin Yang, Dennis Wei, Yi Zhou, Swanand Ravindra Kadhe, Nathalie Baracaldo, and Kimin Lee. Entropy-aware on-policy distillation of language models. In *International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/64855.

Krishnateja Killamsetty, Durga Sivasubramanian, Ganesh Ramakrishnan, Abir De, and Rishabh Iyer. GRAD-MATCH: Gradient matching based data subset selection for efficient deep model training. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5464–5474. PMLR, 2021. URL https://proceedings.mlr.press/v139/killamsetty21a.html.

Minsang Kim and Seung Jun Baek. Explain in your own words: Improving reasoning via token-selective dual knowledge distillation. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/a89331e2ecbbd9d9618bc94717a91155-Abstract-Conference.html.

Yoon Kim and Alexander M. Rush. Sequence-level knowledge distillation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1317–1327. Association for Computational Linguistics, 2016. doi: 10.18653/v1/D16-1139. URL https://aclanthology.org/D16-1139/.

Jongwoo Ko, Sungnyun Kim, Tianyi Chen, and Se-Young Yun. DistiLLM: Towards streamlined distillation for large language models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 24872–24895. PMLR, 2024. URL https://proceedings.mlr.press/v235/ko24c.html.

Jongwoo Ko, Tianyi Chen, Sungnyun Kim, Tianyu Ding, Luming Liang, Ilya Zharkov, and Se-Young Yun. DistiLLM-2: A contrastive approach boosts the distillation of LLMs. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 31044–31062. PMLR, 2025. URL https://proceedings.mlr.press/v267/ko25a.html.

Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, et al. Solving Quantitative Reasoning Problems with Language Models, 2022. URL https://arxiv.org/abs/2206.14858.

Yuying Li, Leqi Zheng, Yongzi Yu, Wenrui Zhou, Xuchang Zhong, Xing Hu, Jing Jin, Hangjie Yuan, and Tao Feng. Filter, Then Reweight: Rethinking Optimization Granularity in On-Policy Distillation, 2026. URL https://arxiv.org/abs/2606.02684.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1711.05101.

Baharan Mirzasoleiman, Jeff Bilmes, and Jure Leskovec. Coresets for data-efficient training of machine learning models. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6950–6960. PMLR, 2020. URL https://proceedings.mlr.press/v119/mirzasoleiman20a.html.

Qwen, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, et al. Qwen2.5 Technical Report, 2025. URL https://arxiv.org/abs/2412.15115.

Zhennan Shen, Yanshu Li, Qingyu Yin, Chak Tou Leong, Zhilin Wang, Yanxu Chen, Rongduo Han, Sunbowen Lee, and Yi R. Fung. On the geometry of on-policy distillation, 2026. URL https://arxiv.org/abs/2606.07082.

Huanxin Sheng, Zhiling Ye, Haonan Wang, Jian Wang, Jinjie Gu, and Jian Kang. 1% of tokens can be enough: On gradient estimation in on-policy distillation, 2026. URL https://arxiv.org/abs/2609.24432.

Yuanyi Wang, Su Lu, Yanggan Gu, Pengkai Wang, Yifan Yang, Zhaoyi Yan, Congkai Xie, Jianmin Wu, and Hongxia Yang. Not All Disagreement Is Learnable: Token Teachability in On-Policy Distillation, 2026. URL https://arxiv.org/abs/2605.26844.

Wenda Xu, Rujun Han, Zifeng Wang, Long Le, Dhruv Madeka, Lei Li, William Yang Wang, Rishabh Agarwal, Chen-Yu Lee, and Tomas Pfister. Speculative knowledge distillation: Bridging the teacher-student gap through interleaved sampling. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/a2747a3844ca1e4667fbff3f558eb39b-Abstract-Conference.html.

Yuanda Xu, Hejian Sang, Zhengze Zhou, Ran He, Zhipeng Wang, and Alborz Geramifard. TIP: Token Importance in On-Policy Distillation, 2026. URL https://arxiv.org/abs/2604.14084.

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 preprint arXiv:2505.09388*, 2025. doi: 10.48550/arXiv.2505.09388. URL https://arxiv.org/abs/2505.09388.

Guo Yu, Wenlin Liu, Yulan Hu, Hao-Xuan Ma, Jun-Peng Jiang, and Han-Jia Ye. Dense supervision, sparse updates: On the sparsity and geometry of on-policy distillation, 2026. URL https://arxiv.org/abs/2606.13657.

Dongxu Zhang, Zhichao Yang, Sepehr Janghorbani, Jun Han, Andrew Ressler II, Qian Qian, Gregory D Lyng, Sanjit Singh Batra, and Robert E. Tillman. Fast and effective on-policy distillation from reasoning prefixes. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 25553–25569, San Diego, California, United States, July 2026. Association for Computational Linguistics. doi: 10.18653/v1/2026.findings-acl.1276. URL https://aclanthology.org/2026.findings-acl.1276/.

Appendix Contents

Supplementary material for CoRD Page

A Technical Details . . . 14
A.1 Scoring and parameterization . . . 14
A.2 Basis construction and reconstruction . . . 14
A.3 One on-policy update . . . 14
A.4 Numerical checks . . . 14
A.5 Computational cost . . . 14

B Experimental Details . . . 16
B.1 Data, evaluation, and metrics . . . 16
B.2 Comparison groups and fairness . . . 17
B.3 Models, optimization, and reporting . . . 18
B.4 Preliminary-study protocol . . . 20
B.5 Controlled-study protocols . . . 21
B.6 Token-selection diagnostic protocol . . . 22
B.7 Spectral-budget diagnostic protocol . . . 23

C Additional Experimental Results . . . 25
C.1 Preliminary-study results . . . 25
C.2 Detailed main-result tables . . . 25
C.3 Controlled analyses . . . 29
C.4 Token-selection results . . . 32
C.5 Spectral compressibility and token budgets . . . 33

D Proofs . . . 34
D.1 Derivations and geometric properties . . . 34
D.2 Correction-space compressibility . . . 35

A TECHNICAL DETAILS

A.1 SCORING AND PARAMETERIZATION

The teacher is frozen. The student's attention and MLP projections contain LoRA adapters; the embedding matrix and output head remain fixed. Dropout is disabled for generation and scoring. Each prompt produces one response using temperature 0.6 and nucleus probability 0.95, with a maximum of 2,048 response tokens. A generated EOS ends the response and remains a valid candidate. Sampling transforms do not affect the temperature-one distributions used in the loss.

Teacher and student score identical tokenized prefixes. A differentiable student scoring pass retains the computation graph for all token losses. The vectors used for selection are detached copies of the logit derivatives and the actual inputs to the output head from this pass. Selection is followed directly by backward propagation through the retained graph, with no additional student scoring pass. The sampled tokens and teacher scores remain detached.

Model forward computation uses BF16; log-softmax, token losses, and logit derivatives use FP32. Log probability ratios are evaluated as differences of log-softmax values. Kernel inner products, factor updates, and triangular solves use FP64. Reconstruction coefficients are cast to FP32 before multiplying the token losses. Equations (7)–(25) describe exact arithmetic; the stated precisions specify their numerical implementation.

A.2 BASIS CONSTRUCTION AND RECONSTRUCTION

Algorithm 1 takes a kernel-column routine and its diagonal. Selected positions retain pivot order. The full-budget branch returns all positions before selection. When the kernel is zero, a uniform subset is sampled without replacement using the run's seeded random-number generator. This branch supplies a defined loss subset when head geometry provides no preference. The actual adapter gradient remains determined by the weighted losses.

A.3 ONE ON-POLICY UPDATE

Algorithm 2 applies the trajectory surrogates with the dense reference's token normalization. Empty responses contribute neither loss nor denominator. An empty batch leaves the optimizer state unchanged. The basis and weights are discarded after each update. Inference uses only the trained student.

A.4 NUMERICAL CHECKS

The rank threshold in Algorithm 1 uses FP64 machine precision ϵ_{mach}. Residuals in $[-\tau, 0)$ are rounded to zero; residuals below $-\tau$ fail the check. For $s = |B|$ and $\beta = A^\top \mathbf{1}_T / T$, the triangular solve must satisfy $\|Rw - \beta\|_\infty \le 100 s \epsilon_{\mathrm{mach}} (\|R\|_\infty \|w\|_\infty + \|\beta\|_\infty)$. The strictly lower triangle of R, which is zero in exact arithmetic, is discarded for the solve. Nonfinite scores, factors, coefficients, losses, and gradients are numerical failures. A failed check terminates training before the optimizer step and saves the offending batch for diagnosis.

Reconstruction uses the unregularized triangular solve. Coefficients are neither clipped nor normalized. The selected count, number of nonzero weights, $\|w\|_1$, and solve residual are recorded. The solve check measures backward error; coefficient accuracy also depends on the conditioning of R. Numerical rank stopping determines the implemented basis size at the stated tolerance; exact rank is a property of the underlying real-valued correction matrix.

A.5 COMPUTATIONAL COST

Caching all logit derivatives r_t requires $O(TV)$ storage, hidden states require $O(Td)$, and the Cholesky factor requires $O(Ts)$. Each selected column uses T vocabulary inner products and T hidden-state inner products, for $O(Ts(V + d))$ arithmetic across selection. Factor updates add $O(Ts^2)$. Forming $A^\top \mathbf{1}_T$ and solving for w cost $O(Ts + s^2)$. The implementation stores neither the Vd-dimensional correction vectors nor the full $T \times T$ kernel.

Algorithm 1 Correction basis and mean reconstruction

Require: Kernel-column access $\mathrm{Kcol}(i) = K_{:i}$, diagonal $k_t = K_{tt}$, budget m, seeded random-number generator
Ensure: Ordered indices B and detached coefficients w

$T \leftarrow \mathrm{length}(k)$
if $T = 0$ **then**
  return $(\varnothing, \varnothing)$
end if
if $m = T$ **then**
  return $((1, \ldots, T), \mathbf{1}_T/T)$
end if
Check that k is finite and nonnegative
if $\max_t k_t = 0$ **then**
  Sample m distinct positions uniformly; sort them to obtain B
  return $(B, \mathbf{1}_m/m)$
end if
$A \leftarrow [\,] \in \mathbb{R}^{T\times 0}$; $v \leftarrow k$; $B \leftarrow [\,]$
$\tau \leftarrow 100T\epsilon_{\mathrm{mach}} \max_t k_t$
for $j = 1, \ldots, m$ **do**
  $i \leftarrow$ smallest-index maximizer of v_i outside B
  if $v_i \le \tau$ **then**
    break
  end if
  $a \leftarrow (\mathrm{Kcol}(i) - AA^\top_{i,:})/\sqrt{v_i}$
  $A \leftarrow [A, a]$; append i to B; $v \leftarrow v - a \odot a$
  Check finite values and $\min_t v_t \ge -\tau$; set negative residuals to zero
end for
$R \leftarrow A^\top_{B,:}$; $\beta \leftarrow A^\top \mathbf{1}_T/T$
$w \leftarrow \mathrm{UpperTriangularSolve}(R, \beta)$
Check finite coefficients and the solve residual
return (B, w)

Algorithm 2 One CoRD training update

Require: Prompt batch X, student adapters θ, frozen teacher, fraction $\rho \in (0, 1]$, AdamW state
Ensure: Updated adapters and optimizer state

Generate one response per prompt without gradient
Score each response with both models; retain the student scoring graph
$J \leftarrow 0$; $N \leftarrow 0$
for each response with $T > 0$ valid positions **do**
  Compute ℓ_t by Equation (1)
  Form detached r_t, h_t and kernel access by Equations (2)–(3)
  $m \leftarrow \max\{1, \lfloor \rho T \rfloor\}$
  $(B, w) \leftarrow \mathrm{CorrectionBasis}(\mathrm{Kcol}, \mathrm{diag}\, K, m)$
  $J \leftarrow J + T \sum_{j=1}^{|B|} \mathrm{sg}(w_j)\ell_{b_j}$; $N \leftarrow N + T$
end for
if $N > 0$ **then**
  Clear gradients; backpropagate J/N; clip the adapter-gradient norm to 1
  Take one AdamW step
end if

These costs are additional to response generation and teacher/student scoring at every valid position. The student scoring graph is retained until the weighted backward pass. Through causal attention, a selected token loss also requires backward computation at earlier states. CoRD reduces the number of loss positions carrying reconstruction weights; teacher access and end-to-end computation are measured separately.

B EXPERIMENTAL DETAILS

B.1 DATA, EVALUATION, AND METRICS

Training data and held-out development sets. The primary track uses the official training splits of MATH and GSM8K (Hendrycks et al., 2021; Cobbe et al., 2021), using questions alone for on-policy distillation. We normalize Unicode and whitespace, deduplicate questions, and group detected near-duplicates before splitting. For each group, the SHA-256 hash of its canonical question, interpreted as an integer modulo 20, assigns bucket 0 to configuration development, bucket 1 to held-out mechanism evaluation, and buckets 2–19 to training. This yields approximately 90% training and two disjoint 5% development partitions. We sample uniformly over the pooled training questions. Known overlaps with benchmark questions are removed from training; public test answers never select hyperparameters, checkpoints, or token budgets. The data manifest records the resulting IDs, exclusion rules, and counts.

Benchmark suite. Table 3 lists the six mathematics benchmarks. The principal aggregate is their unweighted mean, with AIME 2024 and 2025 treated as separate tasks. Tables 6–8 give the component scores for the GPU-hour-limited main comparison. The controlled selection and supplementary studies use the same six-task macro metrics.

Table 3: The six-task mathematics evaluation suite. All tasks use the common zero-shot decoder and mean@16/pass@8 definitions below. Dataset revisions, task IDs, prompts, and scorer commits are fixed for each comparison.

Benchmark	Evaluation subset
MATH-500	500 problems from MATH test (Hendrycks et al., 2021)
AIME 2024	30 competition problems
AIME 2025	30 competition problems
AMC 2023	40 problems in the fixed evaluation list
Minerva Math	Fixed mathematics evaluation subset (Lewkowycz et al., 2022)
OlympiadBench	English, text-only mathematics (He et al., 2024)

Decoding and scoring. The principal mathematics panel uses zero-shot question prompts, temperature 1.0, top-$p = 1.0$, no top-k restriction, and at most 16,384 new tokens. Sixteen independent responses are sampled for every question using fixed evaluation seeds. Qwen3 uses its official chat template with thinking enabled for all compared methods; Qwen2.5-Instruct uses its own official template. Teachers and initial students are evaluated with the same settings within each pair. We use the same final-answer extraction and exact/symbolic correctness checker for every method. A malformed, missing, or incorrect final answer scores zero. Hitting the token cap is logged separately; an otherwise valid answer is scored by the same checker.

For N questions and c_i correct responses among $n = 16$, we define $\mathrm{mean@16} = N^{-1} \sum_i c_i/16$ and $\widehat{\mathrm{pass@8}} = N^{-1} \sum_i [1 - \binom{16-c_i}{8} / \binom{16}{8}]$, taking $\binom{a}{8} = 0$ when $a < 8$ (Chen et al., 2021). Mean@16 estimates single-sample correctness; pass@8 estimates the probability that at least one of eight samples is correct, independently of a deployment-time answer-selection procedure. No external tools, test-time teacher access, or answer-dependent resampling are used.

Geometric and optimization diagnostics. The supplementary diagnostics use only Qwen3-4B with the Qwen3-8B teacher. For each seed, 256 fixed held-out mechanism prompts are evaluated at initialization and at the final 1,000-update Dense-RKL state. The compared selectors receive identical cached trajectories, teacher/student distributions, and student parameters at each state. Figure 4 defines the correction and actual LoRA-gradient comparisons, and Table 13 defines the cloned-AdamW and weight-stability measurements. Both per-trajectory gradients and the implemented length-weighted batch gradients are compared with the dense reference. Accuracy is evaluated in separate, independently trained runs.

B.2 Comparison groups and fairness

Matched-budget token selection. The dense reference is on-policy GKD instantiated with full-vocabulary reverse KL and detached student trajectories (Agarwal et al., 2024); it is labelled Dense-RKL to distinguish it from other GKD divergences. CoRD and its controlled selection baselines share this loss, one response per prompt, one optimizer update per fresh batch, and $m = \max\{1, \lfloor \rho T \rfloor\}$ candidate positions per response. Prompt tokens, padding, and artificially appended EOS tokens are excluded; a sampled EOS is included. The per-trajectory loss is multiplied by T before division by the total valid-token count of the batch. At $\rho = 1$, CoRD directly uses every position with weight $1/T$.

Controls include uniform sampling without replacement, evenly spaced positions, the first m positions (Prefix-mask), top-m student entropy, entropy-proportional sampling without replacement, top-m reverse KL, and top-m correction norm $\sqrt{K_{tt}}$. Ties are resolved by position; the entropy sampler falls back to uniform when all scores vanish. These simple controls use uniform weights $1/m$. Prefix-mask scores a complete response; Prefix-OPD stops response generation early.

TIP (Xu et al., 2026) retains its published entropy/disagreement score and normalization. TA-OPD (Wang et al., 2026) retains its compatibility-weighted disagreement score, its robust normalization, and $K = 16$ for the union of teacher/student top-K scoring supports. For the controlled panel, both select m positions within each response, use uniform selected-loss weights, and share the batch normalization above. The labels TIP-budget and TA-OPD-budget in Table 2 identify these quota and loss-reduction adaptations. Their original selectors and loss reductions are evaluated in the end-to-end panel. The training KL and CoRD correction vectors use the full vocabulary, including when TA-OPD uses restricted support for scoring.

CRAIG and GRAD-MATCH (Mirzasoleiman et al., 2020; Killamsetty et al., 2021) are adapted from example selection to token selection using the same detached correction geometry and budget. CRAIG uses distances $d_{ij} = \sqrt{\max\{0, K_{ii} + K_{jj} - 2K_{ij}\}}$ and weights representatives by assigned-token counts divided by T. GRAD-MATCH uses orthogonal matching pursuit to fit the full-trajectory mean with nonnegative coefficients. Its ridge coefficient is $\lambda = \lambda_{\text{rel}} \operatorname{tr}(K)/T$, with $\lambda_{\text{rel}} \in \{0, 10^{-6}, 10^{-4}\}$ selected on development data. Both are proxy-token adaptations, marked with † in Table 2. A separate signed mean-matching pursuit control removes the nonnegativity constraint, isolating mean-directed selection from CoRD's residual-directed selection. The Qwen3-4B factorial study crosses uniform versus CoRD selection with uniform versus signed least-squares reconstruction weights (Table 9). Uniform weights are $1/|B|$ on the actual selected set; singular uniform subsets use a rank-revealing pseudoinverse with the same numerical tolerance convention. This separates selection from reconstruction while preserving the common batch normalization.

Comparisons preserving each method's mechanism. We rerun the following methods in the common model/data/LoRA setting, preserving their defining objectives, sampling mechanisms, and loss reductions.

- GKD uses fully on-policy data with forward KL and JSD($\beta = 0.9$) as separate dense references (Agarwal et al., 2024). MiniLLM retains its policy-gradient estimator and teacher-mixed sampling (Gu et al., 2024); SKD retains interleaved teacher/student sampling (Xu et al., 2025). DistiLLM and DistiLLM-2 retain their respective objectives and data/reuse policies, including DistiLLM-2's teacher/student contrastive construction (Ko et al., 2024; 2025).
- EOPD retains its sampled, clipped reverse-KL update and teacher-entropy-dependent forward-KL term, with the published defaults $\tau = 0.8$, $\alpha = 1$, and teacher top-16 support (Jin et al., 2026). Its matching sampled-OPD reference is included; Dense-RKL is a different estimator.
- TSD-KD retains teacher ranking of the student's top-10 candidate *tokens*, direct JSD($\beta = 0.9$), and entropy regularization. Defaults are indirect-loss weight 0.1 and selection ratios 0.1 for the indirect and entropy terms (Kim & Baek, 2026).
- Prefix-OPD retains early stopping of response generation and its sampled sequence-sum objective (Zhang et al., 2026). Development questions alone choose its prefix cap; the main comparison uses the last completed checkpoint within the common GPU-hour cap.

- FiRe-OPD retains trajectory filtering, soft token weights, and its policy-gradient objective, with default filtering percentile 20% and $\alpha = \beta = 1$ (Li et al., 2026). Discarded trajectories still incur generation and scoring costs. All tokens in retained trajectories can contribute to an update; its filtering percentile and CoRD's token-level fraction ρ therefore measure different quantities.
- SeqKD trains on frozen-teacher responses to the same training prompts (Kim & Rush, 2016); teacher-response token-level KD is an additional off-policy control (Hinton et al., 2015). Their response-generation and teacher-scoring costs are included. Initial students and teachers serve as reference checkpoints.

What is matched and what is charged. The measured costs in Table 2 include all teacher/student scoring, token selection, and optimizer costs. Within each model/data track, all methods start from the same student checkpoint and adapter initialization, use the same teacher, prompt pool, tokenizer mapping, trainable modules, prompt templates, and evaluator. No method receives extra labelled solutions, an external verifier, a specialist teacher, or a method-specific SFT warm start in the primary comparison. Method-specific rollout multiplicity, teacher proposals, replay age, and updates per rollout are logged and included in cost accounting. Online runs sample from their own evolving students; sharing cached trajectories is restricted to fixed-state diagnostics.

Table 2 fixes the controlled selectors at 1,000 optimizer updates; Table 1 fixes an end-to-end GPU-hour cap within each model pair. Each cap is determined from a development-only Dense-RKL timing run of 1,000 updates before examining test scores. On the eight-GPU allocation, GPU-hours equal eight times elapsed job hours; offline teacher-data construction is added. The run records identify the actual achieved budget, last completed checkpoint, and checkpoint overhead; validation/testing cost is accounted for separately. The supplementary quality–cost study varies only CoRD's supervision fraction on Qwen3-4B (Table 10). Even with matched ρ, prompts, or update counts, computation depends on each method's execution costs.

Relation to published protocols. The original methods use different experimental settings: TIP and TA-OPD use DAPO and different teacher configurations (Xu et al., 2026; Wang et al., 2026); EOPD and Prefix-OPD use Base students (Jin et al., 2026; Zhang et al., 2026); TSD-KD uses UltraInteract and few-shot mathematics evaluation (Kim & Baek, 2026); DistiLLM-2 uses specialist mathematics models (Ko et al., 2025); and FiRe-OPD uses different teachers and training data (Li et al., 2026). Our comparison reruns these methods under the shared checkpoint, training, budget, and evaluation protocol in Appendix B, with variation measured across training seeds.

Interpreting the matched-budget panel. Table 2 uses final 1,000-update checkpoints, independently of the GPU-hour-limited checkpoints in Table 1. Its $f_{\neq 0}$ is 100 times the total number of valid response positions with nonzero loss coefficients divided by the total number of valid candidate positions across training. It can be below 100ρ because of integer quotas, numerical rank stopping, or zero reconstruction coefficients. GPU-hours include the entire training job; peak allocated memory is the maximum across all eight GPUs over the prescribed measurement interval. The paired quality and cost entries come from the same run and are measured independently of the configured ρ.

B.3 Models, optimization, and reporting

Model checkpoints and training modes. The main student–teacher pairs are Qwen3-1.7B–Qwen3-8B, Qwen3-4B–Qwen3-8B, and Qwen2.5-1.5B-Instruct–Qwen2.5-14B-Instruct (Yang et al., 2025; Qwen et al., 2025). We use the released post-trained Qwen3 checkpoints with thinking mode fixed across methods. The Qwen2.5 pair matches TSD-KD's model sizes; the two Qwen3 sizes cover the family most common in the recent token-selection studies. We verify and record checkpoint revisions, tokenizer hashes, chat templates, and the one-to-one mapping of scored vocabulary IDs. Token-level comparisons require compatible tokenizers. Qwen3-0.6B–Qwen3-4B is used only for the preliminary diagnostic study.

Configuration selection and checkpoint policy. Table 4 summarizes the controlled configuration. Global batch size is maintained by gradient accumulation with one response per training microbatch; the parallel layout and accumulation count are recorded. Training questions exceeding the prompt cap are excluded before forming the shared pool. All evaluation questions are retained regardless of

Table 4: Hyperparameters for CoRD and the controlled Dense-RKL token-selection panel. Method-specific mechanisms in the other comparison panel are preserved.

Item	Configuration
Hardware	$8\times$ NVIDIA A100, 80GB per GPU
Trainable parameters	LoRA on attention q, k, v, o and MLP gate/up/down projections; frozen base weights, embeddings, and output head
LoRA	Rank 32, scaling $\alpha = 64$, dropout 0
Optimizer	AdamW; $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$, weight decay 0; gradient-norm clipping at 1
Learning rate / schedule	2×10^{-5} default; 10^{-5} alternative; linear warmup for 3% of updates, then cosine decay to zero
Batch / update budget	Global batch 16 prompts; one response per prompt; one update per fresh batch; 1,000 updates
Training generation	Temperature 0.6, top-$p = 0.95$, no top-k restriction; prompt cap 2,048; response cap 2,048
CoRD budget	$\rho = 0.25$ default; Qwen3-4B-only sweep $\{0.10, 0.25, 0.50, 1.00\}$; numerical rank stopping may select fewer than m positions
Arithmetic	BF16 model computation; FP32 log-softmax, KL, and logit derivatives; FP64 kernel arithmetic, factorization, and solve
Training seeds	17, 29, 43; identical seed list for every method

length. Configuration selection uses only held-out configuration-development questions, scored by pooled mean@8 accuracy under the mathematics decoder. The held-out mechanism split is reserved for diagnostics.

Controlled training and supplementary ablations use their final 1,000-update checkpoints. Cost-limited main runs use the last completed checkpoint within the fixed GPU-hour cap. Before test scoring, hyperparameters are frozen together with the baseline configurations, code commits, package/driver versions, data/model revisions, numerical settings, seeds, and trainable-parameter counts.

CoRD execution and cost measurement. We use the exact full-vocabulary reverse KL at distillation temperature one and the actual output-head input vectors. The sampled trajectory, teacher outputs, correction geometry, selected indices, and weights are detached. A single differentiable student scoring pass supplies the graph reused for selection and the weighted backward pass. Coefficients are signed, unregularized, and neither clipped nor normalized. Numerical rank stopping, zero-kernel fallback, and failure handling follow Appendix A.4; failed runs are reported with the saved offending batches.

End-to-end timing includes response generation, teacher scoring, student scoring, kernel/selection/solve, backward propagation, optimizer update, and synchronization. The Qwen3-4B profiling comparison in Table 14 uses 100 consecutive updates after 20 warm-up updates. Stage times are non-overlapping elapsed job times, with optimizer and synchronization overhead assigned to the final stage; their sum is the measured update time. Total GPU-hours include warm-up and job overhead. Peak allocated and reserved GPU memory are the maxima across all eight GPUs during the measurement interval. Generated tokens per second and teacher-scored positions are measured from the same runs. Dense graph storage, scoring every candidate, and backpropagation through earlier causal states contribute costs beyond the selected loss positions. FLOPs, teacher access, memory, and wall-clock time therefore depend on the full execution procedure.

Aggregation and uncertainty. Accuracy is reported as mean and sample standard deviation over training seeds 17, 29, and 43, computing each seed's six-task macro before aggregation. The supplementary studies retain the same three-seed convention; their diagnostic and profiling aggregations are specified in Appendix B.5. The training run is the unit of replication. Tables summarize measured performance and variation across seeds. Seed-level scores, truncation counts, numerical failures, and job logs support the reported aggregates.

Reading and aggregating entries. Scores are percentages, reported as mean $\pm$ sample standard deviation over training seeds 17, 29, and 43. M16 denotes mean@16, while P8 denotes the combinatorial pass@8 estimate from the same 16 responses; both are computed directly from response

correctness. For each seed, the macro score is the unweighted mean of the six benchmark scores. Its standard deviation is computed across these seed-level macro scores. Reference checkpoints have no training-seed deviation. CoRD's shading identifies our method; bold marks the highest mean among trained students in each column. Macro scores are computed before rounding; averaging the displayed benchmark means can therefore differ by 0.1 percentage point from the displayed macro score.

B.4 Preliminary-study protocol

B.4.1 Trajectories and correction geometry

We randomly sample 256 prompts from the configuration-development split. A Qwen3-0.6B student generates one response per prompt with sampling temperature 0.6, top-$p = 0.95$, and a maximum of 512 response tokens. The frozen Qwen3-4B teacher and the student score the same generated prefixes. The recorded response lengths range from 80 to 512 tokens. The 512-token cap belongs to this diagnostic study; the training configuration in Appendix A.1 uses a separate 2,048-token cap. No CoRD selection, reweighting, or parameter update is performed in the preliminary study.

For a response with T valid positions, let q_t and p_t be the student and teacher distributions and h_t the student output-head input. The per-token reverse-KL loss and head correction are

$$\ell_t = D_{\mathrm{KL}}(q_t \| p_t), \qquad r_t = q_t \odot [\log(q_t/p_t) - \ell_t \mathbf{1}], \qquad c_t = \mathrm{vec}(r_t h_t^\top). \tag{10}$$

The logarithm, division, and product act elementwise. With the sampled prefix and h_t held fixed, c_t is the vectorized gradient with respect to an independent output-head matrix and serves as a proxy for adapter training. Appendix D.1.1 derives this identity. Correction normalization is used only for the directional diagnostics below, while the mean-distortion calculation retains the original vectors and their magnitudes.

B.4.2 Spectral metrics and visualization

For nonzero corrections, define $u_t = c_t/\|c_t\|_2$ and the cosine Gram matrix $\bar{K}_{ij} = u_i^\top u_j$. Let $\lambda_1 \geq \cdots \geq \lambda_T \geq 0$ denote its full spectrum, and set $\pi_j = \lambda_j / \sum_i \lambda_i$. The normalized participation ratio is

$$d_{\mathrm{rel}} = \frac{1}{T} \frac{(\sum_j \lambda_j)^2}{\sum_j \lambda_j^2} = \frac{1}{T \sum_j \pi_j^2}. \tag{11}$$

It equals one for mutually orthogonal unit corrections and $1/T$ for a rank-one spectrum. Small values indicate low effective directional dimension, without requiring all corrections to be parallel or to have the same sign. The other spectral summaries are

$$E(k) = \sum_{j=1}^{k} \pi_j, \qquad k_{80} = \min\{k : E(k) \geq 0.8\}, \qquad r_{80} = \frac{k_{80}}{T}, \qquad E_{20} = E(\lceil 0.2T \rceil). \tag{12}$$

Here "energy" refers to the spectrum of unit-normalized corrections. All summaries use the full spectrum.

Figure 2(a) displays $E(k)$ against k/T. Each trajectory occupies one equally spaced row, ordered by increasing T; vertical ticks label the response lengths of the corresponding rows. For display only, the cumulative curves, including $E(0) = 0$, are linearly interpolated onto 96 equally spaced coordinates in $[0, 1]$. The overlaid curve uses the exact discrete r_{80} for each trajectory. All reported statistics use the discrete spectra.

B.4.3 Uniform de-duplication diagnostic

The diagnostic uses average-linkage angular clustering of u_t, with a cosine-similarity threshold of 0.92. Average linkage averages cross-cluster pairwise dissimilarities; a medoid is a cluster member minimizing its total within-cluster dissimilarity. For clusters $S_1, \ldots, S_C$ and their medoid indices $m_1, \ldots, m_C$, define

$$\mu = \frac{1}{T} \sum_{t=1}^{T} c_t, \qquad \mu_{\mathrm{dedup}} = \frac{1}{C} \sum_{k=1}^{C} c_{m_k}, \qquad \Delta_{\mathrm{mean}} = \frac{\|\mu_{\mathrm{dedup}} - \mu\|_2}{\|\mu\|_2 + \epsilon}, \quad \epsilon > 0. \tag{13}$$

The positive stabilizer prevents division by zero; the metric measures relative change in the head-correction mean. Uniform weighting assigns every cluster weight $1/C$, regardless of the number of original tokens it contains.

This probe changes both cluster weights and within-cluster representations. To distinguish these effects algebraically, let $n_k = |S_k|$ and $\bar{c}_k = n_k^{-1} \sum_{t \in S_k} c_t$. Then

$$\mu_{\text{dedup}} - \mu = \underbrace{\sum_{k=1}^{C} \left(\frac{1}{C} - \frac{n_k}{T} \right) c_{m_k}}_{\text{discarding cluster multiplicities}} + \underbrace{\sum_{k=1}^{C} \frac{n_k}{T} (c_{m_k} - \bar{c}_k)}_{\text{replacing cluster means by medoids}} . \tag{14}$$

The second term can be nonzero even for parallel vectors with different magnitudes. The reported distortion combines both effects under uniform medoid weighting; CoRD's fitted reconstruction weights are evaluated separately.

B.4.4 Aggregation and visualization

Figure 2(b) groups trajectories by the five length intervals in Table 5 and by d_{rel} edges 0, 0.12, 0.20, 0.28, 0.36, 0.44, and 0.56. Participation-ratio bins include the left edge and exclude the right edge, except that the final bin includes both endpoints. Each populated cell reports the median Δ_{mean} and its trajectory count n; gray cells indicate empty bins. All 256 trajectories are included. Length-group and subgroup summaries pool individual trajectories with equal weight.

B.5 Controlled-study protocols

These studies use Qwen3-8B $\rightarrow$ Qwen3-4B, the same MATH/GSM8K training pool, and seeds 17, 29, and 43. They isolate selection, reconstruction, supervision budget, and measured execution cost. Accuracy uses the same six mathematics benchmarks and decoder as the main comparison.

B.5.1 Selection, reconstruction, and budget controls

The factorial comparison in Table 9 fixes $\rho = 0.25$ and 1,000 updates. Each configuration is trained on its own evolving on-policy trajectories. Uniform loss weights are $1/|B|$ on the actual selected set; signed least-squares (LS) weights reconstruct the full mean correction. For a rank-deficient uniform subset, the minimum-norm pseudoinverse uses the numerical tolerance in Appendix A.4. CoRD's uniform-weight control uses the same rank-stopped selection rule as CoRD/LS, changing only the weights; their selected bases coincide when evaluated on an identical trajectory and state.

B.5.2 Correction and trainable-gradient fidelity

For each seed, the fixed-state study uses 256 mechanism prompts at initialization and the final Dense-RKL checkpoint. Within each state, all methods share trajectories and scoring outputs. Prompts are grouped in a fixed order into batches of 16; batch gradients use the implemented token-length weighting. The compared methods include the factorial controls, a correction-norm selector, and both nonnegative and signed mean-matching controls. All selected methods use $\rho = 0.25$.

Define $e(a, \widehat{a}) = \|\widehat{a} - a\|_2 / (\|a\|_2 + 10^{-12})$. In Figure 4, $e_\mu = e(\mu, \widehat{\mu})$, and $\eta_{\text{cov}} = \sum_t \|(I - P_B)c_t\|_2^2 / \sum_t \|c_t\|_2^2$. Panels (b,c) compare actual trainable LoRA gradients with their dense counterparts, before clipping. Each seed is summarized by the unweighted mean over its 256 trajectories or 16 batches; the heatmap shows means across seeds, while Table 11 retains all means and sample SDs. For each metric, the color scale pools both states and all methods and maps the observed minimum and maximum to its endpoints, reversing the scale for lower-is-better errors. Color therefore encodes relative fidelity on a metric-specific scale. Undefined cosines and zero-energy denominators are excluded only from the affected average and their counts are retained with the result records.

B.5.3 Correction geometry and landscape estimation

The analysis in Section 5.4.2 uses all 1,536 non-empty trajectory-state records (256 for each of seeds 17, 29, and 43 at each reference state). Uniform/LS and CoRD/LS are paired within each record

at $\rho = 0.25$. The e_μ, η_{cov}, and trajectory-gradient-error entries for these two methods in Figure 4 and Table 11 summarize these records by first averaging within a seed and then taking the mean and sample SD across the three seeds. Since η_{cov} depends only on the selected span, each selector's uniform-weight control shares its η_{cov} values.

Local averaging and support. We bin (d_{rel}, χ) on a 240×240 grid over $[0.08, 0.46] \times [0.05, 0.96]$. Counts and sums of $\Delta e_g^{\mathrm{traj}}$ are smoothed separately using separable Gaussian kernels with bandwidth 0.18 times the pooled standard deviation of each coordinate (at least one grid cell), truncated at approximately four bandwidths. Their ratio estimates the local mean gain. Each trajectory-state record has equal weight in this surface. We mask grid cells outside the empirical convex hull, below an effective local count of eight, or below 1.5% of the peak smoothed density. Effective count is $(\sum_i w_i)^2 / \sum_i w_i^2$ for the local kernel weights. White contours enclose approximately 50%, 70%, 85%, and 95% of the smoothed mass, with segments outside supported cells omitted. The color scale retains zero and negative gains.

B.5.4 Optimizer transfer and numerical stability

Each batch starts from an independent clone of the same complete AdamW state. After the common gradient-norm clipping and one optimizer step, $c_{\Delta\theta}$ and $e_{\Delta\theta}$ compare the selected-loss parameter increment with the dense increment. $\Delta\mathcal{L}_{\mathrm{full}}$ is the change in dense, length-weighted reverse KL on that same fixed batch, evaluated before and after the step; negative values indicate a local decrease. This uses the same states and batches as Figure 4. The one-step probe uses a fixed learning rate of 2×10^{-5} for every clone at both states, preserving the saved moments and step counters while holding the scheduler fixed. This fixed positive rate makes local update comparisons possible at both states, including the training-schedule endpoint.

For CoRD's signed reconstruction, define $M_- = \sum_j \max\{-w_j, 0\}$ and report $100 M_- / (\|w\|_1 + 10^{-12})$. Conditioning is measured by $\log_{10} \kappa_2(R)$ for the selected triangular factor. The stability panel reports maxima of $\|w\|_1$ and conditioning over all diagnostic trajectories, the seed-aggregated mean negative-weight fraction, and the total failed numerical checks divided by attempted reconstructions. Zero-kernel fallbacks and undefined gradient comparisons are counted separately in the associated records.

B.5.5 Execution-cost measurement

Table 14 profiles six representative controlled methods at $\rho = 0.25$, with Dense-RKL at $\rho = 1$. It uses their independent 1,000-update training runs and the timing interval in Appendix B.3. Each stage is averaged over the 100 measured updates within a seed before computing mean $\pm$ sample SD across seeds. The resource panel also summarizes three seed-level measurements. Allocated and reserved memory are per-run maxima across GPUs; throughput is generated tokens divided by elapsed update time over the profiling interval. Teacher-scored positions are counted over the full training run, and GPU-hours include the full job.

B.6 Token-selection diagnostic protocol

Protocol and quantities. The diagnostic in Section 5.4.1 uses Qwen3-8B $\rightarrow$ Qwen3-4B at Dense-RKL update 1,000, with the held-out mechanism split and scoring pipeline of Section 5.3. The token-level records comprise 240 complete trajectories of 264–755 tokens each, totaling 120,426 response tokens. At $\rho = 0.25$, every trajectory selects exactly $\lfloor 0.25T \rfloor$ tokens, yielding 30,017 selections (24.93%); the deviation from 25% is due to integer rounding. We use the recorded reverse-KL and residual-novelty percentiles and the indicator $s_t = \mathbf{1}\{t \in B\}$ of membership in the final basis. Each percentile column spans the within-trajectory rank grid $\{0, 1/(T-1), \ldots, 1\}$.

Conditional selection rates. Let $\mathcal{B}_{ab}$ be a cell in a uniform 42×42 grid over $[0,1]^2$. We accumulate token counts and selected-token counts as

$$N_{ab} = \sum_t \mathbf{1}\{(p_t^{\mathrm{RKL}}, p_t^{\mathrm{nov}}) \in \mathcal{B}_{ab}\}, \qquad S_{ab} = \sum_t s_t \mathbf{1}\{(p_t^{\mathrm{RKL}}, p_t^{\mathrm{nov}}) \in \mathcal{B}_{ab}\}. \tag{15}$$

For occupied cells, the unsmoothed conditional selection rate is $\widehat{P}_{ab} = S_{ab}/N_{ab}$. The displayed field uses

$$\widetilde{P}_{ab} = \frac{(G_\sigma * S)_{ab}}{(G_\sigma * N)_{ab}}, \qquad \sigma = 1 \text{ grid cell}, \tag{16}$$

with Gaussian kernels truncated at four standard deviations and reflecting boundaries. Smoothing counts before taking their ratio weights the estimate by observed token support. Cells with smoothed count below five are masked; all cells meet this threshold in the present diagnostic. Every token has equal weight, so longer trajectories contribute more observations. Both binning and regional statistics use the recorded percentiles directly, without coordinate jitter. Contours of the normalized smoothed counts indicate approximate 50%, 80%, and 95% density-mass regions; solid contours indicate selection rates of 0.25, 0.50, and 0.75.

B.7 Spectral-budget diagnostic protocol

B.7.1 Fixed-state reconstruction scan

We use the Qwen3-8B $\rightarrow$ Qwen3-4B fixed-state protocol from Section 5.3: 256 held-out prompts per seed, three seeds, and initialization and Dense-RKL update 1,000, giving 1,536 trajectory-state instances. For each instance, all budgets share the response, teacher/student distributions, hidden states, and kernel. We vary $m = \max\{1, \lfloor \rho T \rfloor\}$ over $\mathcal{R} = \{0.02, 0.04, \ldots, 0.60\} \cup \{0.25, 0.75, 1.00\}$, rerunning only selection and reconstruction, with no parameter update. The relative mean-correction error is

$$e_\mu(m) = \frac{\|\mu - P_m \mu\|_2}{\|\mu\|_2 + 10^{-12}}, \qquad b_\mu(m) = \frac{T-m}{T} \frac{\sqrt{\mathcal{Q}_m(K)}}{\|\mu\|_2 + 10^{-12}}, \tag{17}$$

where $\mathcal{Q}_m(K)$ is defined in Theorem 1. Before rank exhaustion, $e_\mu(m) \le b_\mu(m)$ under that theorem's exact-arithmetic assumptions. At $\epsilon = 0.10$, we compare the continuously interpolated spectral budget $\widehat{\rho}^{\mathrm{spec}}_{0.10}$ with the first evaluated budget $\widehat{\rho}^{\mathrm{emp}}_{0.10} = \min\{\rho \in \mathcal{R} : e_\mu(m(\rho)) \le 0.10\}$. Appendix B.7.2 gives the full certificate definition and aggregation procedure.

B.7.2 Spectral certificate and budget definitions

For each fixed trajectory, let $F = [c_1, \ldots, c_T]$, $K = F^\top F$, $\mu = T^{-1} \sum_t c_t$, and $r_\star = \mathrm{rank}(F)$. The eigenvalues of the unnormalized kernel are $\lambda_1 \ge \cdots \ge \lambda_T \ge 0$. For $0 \le m < r_\star$, set $N_m = T - m$ and

$$L_i = \frac{1}{T-i} \sum_{j=i+1}^{T} \lambda_j, \qquad \mathcal{Q}_m(K) = \min_{\substack{1 \le n \le m+1 \\ 0 \le k < n}} \left(\frac{\prod_{j=1}^{n} \lambda_j}{N_m \prod_{i=0}^{k-1} L_i} \right)^{1/(n-k)}, \tag{18}$$

with empty products equal to one. Theorem 1 gives

$$b_\mu(m) = \frac{N_m}{T} \frac{\sqrt{\mathcal{Q}_m(K)}}{\|\mu\|_2 + 10^{-12}}, \qquad e_\mu(m) \le b_\mu(m). \tag{19}$$

At rank exhaustion, exact reconstruction gives $e_\mu(m) = 0$; we extend $b_\mu(m) = 0$ for $r_\star \le m \le T$, including the zero-kernel case. At the full-budget endpoint, where $N_m = 0$, this extension supplies the bound in place of Eq. (18). For target error ϵ, define

$$\rho^{\mathrm{spec}}_\epsilon = \min_{m \in \{0,\ldots,T\}} \left\{ \frac{m}{T} : b_\mu(m) \le \epsilon \right\}, \qquad \rho^{\mathrm{emp}}_\epsilon = \min_{m \in \{0,\ldots,T\}} \left\{ \frac{m}{T} : e_\mu(m) \le \epsilon \right\}. \tag{20}$$

Under the theorem's assumptions, $\rho^{\mathrm{emp}}_\epsilon \le \rho^{\mathrm{spec}}_\epsilon$. This is a sufficient certificate for output-head mean reconstruction.

The experiment uses a continuously interpolated spectral budget, denoted $\widehat{\rho}^{\mathrm{spec}}_{0.10}$, and a threshold on the 33-point nominal scan grid $\mathcal{R}$ in Appendix B.7.1:

$$\widehat{\rho}^{\mathrm{emp}}_{0.10} = \min_{\rho \in \mathcal{R}} \{\rho : e_\mu(\max\{1, \lfloor \rho T \rfloor\}) \le 0.10\}. \tag{21}$$

There are 50,688 budget observations, with all 33 budgets present for each of the 1,536 instances. Every instance reaches the threshold at an evaluated budget. Statistics use the original recorded budgets, without the small coordinate offsets used to display the points.

B.7.3 LANDSCAPE CONSTRUCTION AND AGGREGATION

The horizontal coordinate is $\widehat{\rho}^{\mathrm{spec}}_{0.10}$; the vertical coordinate is the nominal scan budget ρ. We use 29 equal-width spectral-budget bins on $[0.08, 0.62]$ and retain the median e_μ in each budget–bin cell only when at least eight instances support that cell. The displayed field covers scanned budgets through 0.60; the 0.75 and 1.00 observations remain in the threshold calculation. Bilinear interpolation to a 260×260 grid and Gaussian smoothing with standard deviation 1.8 grid cells produce the background. Smoothing uses normalized finite-value weights and retains the pre-smoothing support mask. Color is logarithmic, with darker purple indicating smaller error. Contours identify $e_\mu \in \{0.05, 0.10, 0.20\}$.

The solid curve is the bin-median first empirical threshold, shown only for bins with at least 12 instances. Points display all 1,536 threshold pairs with offsets of at most 0.001 on each axis. All reported correlations, gaps, ratios, and errors use unsmoothed values and unshifted coordinates. The pooled statistics give each instance equal weight. Each state contains 768 instances, with 256 from each of the three seeds; the mean errors at $\rho = 0.25$ therefore also equal the mean of the three seed means.

C Additional Experimental Results

C.1 Preliminary-study results

The protocol and metric definitions are given in Appendix B.4. Across trajectories, the median d_{rel} is 0.27, with 25th and 75th percentiles 0.20 and 0.34. The mean E_{20} is 77.9%, with sample standard deviation 10.3% (denominator $256 - 1$).

Table 5: Trajectory-level diagnostic summaries. d_{rel} and Δ_{mean} are medians; r_{80} and k_{80} are means. The first displayed interval starts at 64, although the shortest recorded response has 80 tokens.

Response length T	n	d_{rel}	r_{80} (%)	k_{80}	Δ_{mean}
64–128	30	0.391	32.6	34.6	0.147
129–192	58	0.339	27.7	44.1	0.220
193–256	60	0.274	22.5	51.2	0.247
257–384	64	0.221	18.1	57.1	0.317
385–512	44	0.167	13.4	60.6	0.377

The group means of r_{80} decrease with length, whereas those of the absolute rank k_{80} increase from 34.6 to 60.6. Thus, longer responses require a smaller *fraction* of eigen-directions to explain 80% of their normalized spectral energy even as the absolute number of directions grows. Among the 44 responses of 385–512 tokens, 37 have $d_{\mathrm{rel}} < 0.20$; their median Δ_{mean} is 0.38.

The plotting assets contain complete trajectory spectra and per-trajectory mean-distortion values, from which the script computes the displayed summaries. Rerunning the clustering diagnostic additionally requires raw correction vectors, cluster assignments, and the numerical stabilizer. The observed length trends describe output-head geometry in this sample; adapter-gradient fidelity and training outcomes are evaluated in the main experiments.

C.2 Detailed main-result tables

Tables 6–8 decompose Table 1 into its six mathematics benchmarks. Aggregation and checkpoint conventions are specified in Appendices B.2 and B.3. The Qwen3-1.7B and Qwen3-4B students use Qwen3-8B teachers; Qwen2.5-1.5B-Instruct uses Qwen2.5-14B-Instruct. CoRD uses $\rho = 0.25$. Within each model pair, every trained method uses the same MATH/GSM8K prompt pool, rank-32 LoRA modules, and end-to-end GPU-hour cap. The cap is fixed from the development-only Dense-RKL timing run; each method contributes its last completed checkpoint before that cap.

Performance by benchmark. CoRD has the highest trained-student mean in all 36 benchmark–metric–backbone comparisons. On AIME 2024 and AIME 2025, its pass@8 gains over the strongest baseline are 3.1/3.5, 1.7/1.8, and 1.9/2.0 percentage points for Qwen3-1.7B, Qwen3-4B, and Qwen2.5-1.5B, respectively. The corresponding MATH-500 gains are 0.7, 0.3, and 0.4 points.

Quality and cost at matched update counts. Relative to Dense-RKL, CoRD improves mean@16 by 3.1, 3.5, and 3.2 percentage points on Qwen3-1.7B, Qwen3-4B, and Qwen2.5-1.5B, respectively. On Qwen3-4B, its GPU-hour and peak-memory reductions are 3.7% and 1.9%. Compared with Prefix-mask, CoRD uses 14.6 additional GPU-hours and 1.9 additional GB/GPU for 7.1 points higher mean@16. These comparisons separate prediction quality, supervision sparsity, and measured training cost.

Table 6: **Qwen3-8B → Qwen3-1.7B.** Per-benchmark reasoning performance under the common GPU-hour cap. Panel (a) reports mean@16; panel (b) reports pass@8. Scores are percentages (mean ± training-seed standard deviation); reference checkpoints have no training-seed deviation. Bold marks the highest mean among trained students in each column.

(a) mean@16 ↑

Method	MATH-500	AIME 2024	AIME 2025	AMC 2023	Minerva	Olympiad
Teacher	93.8	58.5	52.3	90.8	61.9	70.8
Initial student	78.4	24.0	20.0	69.7	39.2	45.8
SeqKD	80.1±0.4	25.4±1.4	21.3±1.3	71.0±0.5	40.8±0.7	47.2±0.7
Teacher-response KD	80.7±0.5	26.1±1.5	22.0±1.4	71.7±0.6	41.5±0.7	47.8±0.8
Dense-RKL	83.6±0.6	30.1±1.6	25.7±1.6	75.8±0.7	44.9±0.8	52.6±1.0
GKD (FKL)	82.8±0.6	29.4±1.6	24.9±1.5	75.1±0.7	44.1±0.9	51.8±0.8
GKD (JSD)	83.1±0.5	29.7±1.4	25.1±1.3	75.4±0.7	44.5±0.7	52.0±0.8
MiniLLM	81.9±0.5	28.2±1.5	23.8±1.4	73.9±0.7	43.3±0.8	50.6±0.8
SKD	82.4±0.6	28.8±1.5	24.4±1.4	74.6±0.6	43.8±0.8	51.2±0.8
DistiLLM	82.2±0.6	28.6±1.5	24.0±1.6	74.2±0.7	43.6±0.9	50.9±0.9
DistiLLM-2	83.3±0.6	30.0±1.9	25.3±1.7	75.2±0.7	44.5±0.8	52.2±1.0
Sampled OPD	83.0±0.5	29.6±1.3	25.2±1.2	75.3±0.6	44.2±0.7	51.9±0.8
EOPD	84.2±0.6	31.7±1.3	27.2±1.5	76.5±0.7	45.2±0.8	53.1±0.8
TSD-KD	83.5±0.6	30.6±1.7	26.1±1.6	75.8±0.6	45.0±0.8	52.7±0.9
Prefix-OPD	82.6±0.4	29.0±1.3	24.7±1.3	74.8±0.6	44.0±0.7	51.5±0.8
TIP	84.5±0.7	32.3±1.9	27.8±1.7	76.9±0.8	45.7±0.8	53.8±1.0
TA-OPD	84.7±0.5	32.7±1.5	28.1±1.2	77.2±0.6	46.1±0.7	54.0±0.7
FiRe-OPD	85.0±0.6	33.0±1.5	28.6±1.3	77.5±0.6	46.4±0.8	54.4±0.9
CoRD (ours)	**85.8±0.5**	**34.4±1.7**	**30.2±1.5**	**79.0±0.8**	**47.6±0.7**	**56.0±1.0**

(b) pass@8 ↑

Method	MATH-500	AIME 2024	AIME 2025	AMC 2023	Minerva	Olympiad
Teacher	98.3	79.3	72.9	97.4	82.7	88.0
Initial student	92.5	42.1	35.2	88.4	62.5	68.9
SeqKD	93.0±0.5	43.4±2.2	36.4±2.4	88.7±0.6	63.6±1.1	69.6±1.2
Teacher-response KD	93.1±0.1	43.9±1.7	36.9±2.1	88.8±0.4	63.9±0.8	69.7±0.8
Dense-RKL	94.8±0.2	50.5±1.6	43.3±0.9	91.6±0.3	68.6±0.5	75.2±0.5
GKD (FKL)	94.4±0.3	49.3±1.4	42.0±0.8	91.2±0.5	67.6±0.9	74.3±0.9
GKD (JSD)	95.0±0.4	51.1±1.1	43.7±2.4	91.9±0.7	69.3±1.6	75.7±1.4
MiniLLM	94.2±0.3	48.2±2.6	40.9±2.5	90.8±0.5	67.2±1.3	73.6±1.0
SKD	94.6±0.5	49.6±1.0	42.3±0.7	91.4±0.6	68.3±1.3	74.7±1.1
DistiLLM	94.2±0.8	48.4±3.7	40.8±3.7	90.7±1.1	67.1±2.5	73.5±2.4
DistiLLM-2	94.8±0.3	50.6±2.4	43.0±1.9	91.4±0.1	68.4±0.9	75.0±0.7
Sampled OPD	94.7±0.7	50.2±2.7	43.0±3.0	91.5±0.8	68.2±2.1	74.9±2.0
EOPD	94.9±0.2	52.1±2.0	45.0±2.2	91.7±0.5	68.5±1.2	75.2±0.9
TSD-KD	94.9±0.1	51.6±1.4	44.4±1.9	91.8±0.4	69.2±0.3	75.7±0.3
Prefix-OPD	94.7±0.2	49.8±1.4	42.7±1.8	91.5±0.6	68.4±0.7	74.9±0.9
TIP	95.0±0.5	52.8±3.2	45.7±3.2	91.9±0.7	68.9±1.5	75.8±1.6
TA-OPD	95.2±0.8	53.6±3.9	46.5±3.7	92.2±1.1	69.6±2.5	76.2±2.4
FiRe-OPD	95.3±0.6	54.0±2.7	47.1±2.9	92.3±0.8	69.9±1.7	76.5±1.5
CoRD (ours)	**96.0±0.3**	**57.1±0.4**	**50.6±0.7**	**93.5±0.4**	**72.3±0.9**	**79.0±0.7**

Table 7: **Qwen3-8B → Qwen3-4B.** Per-benchmark reasoning performance under the common GPU-hour cap. Panel (a) reports mean@16; panel (b) reports pass@8. Scores are percentages (mean ± training-seed standard deviation); reference checkpoints have no training-seed deviation. Bold marks the highest mean among trained students in each column.

(a) mean@16 ↑

Method	MATH-500	AIME 2024	AIME 2025	AMC 2023	Minerva	Olympiad
Teacher	93.8	58.5	52.3	90.8	61.9	70.8
Initial student	88.1	39.6	35.8	83.4	50.6	59.3
SeqKD	88.8±0.6	40.4±1.9	36.5±1.7	84.0±0.8	51.4±0.8	60.0±1.0
Teacher-response KD	89.2±0.5	41.2±1.5	37.1±1.2	84.5±0.5	52.0±0.7	60.5±0.8
Dense-RKL	90.8±0.6	44.7±1.5	40.5±1.5	86.6±0.6	54.7±0.7	63.3±0.8
GKD (FKL)	90.3±0.5	43.8±1.4	39.8±1.4	86.0±0.7	54.1±0.8	62.6±0.8
GKD (JSD)	90.6±0.5	44.3±1.2	40.1±1.2	86.3±0.6	54.4±0.6	62.9±0.6
MiniLLM	89.9±0.6	42.8±1.5	38.9±1.4	85.4±0.7	53.3±0.9	61.7±0.9
SKD	90.1±0.7	43.2±1.8	39.2±1.7	85.7±0.9	53.7±0.8	62.0±1.0
DistiLLM	90.0±0.6	43.0±1.4	39.0±1.4	85.5±0.7	53.5±0.9	61.9±0.9
DistiLLM-2	90.7±0.5	44.4±1.7	40.3±1.3	86.4±0.7	54.5±0.7	63.0±0.8
Sampled OPD	90.5±0.7	44.0±1.7	40.0±1.8	86.2±0.8	54.2±0.9	62.8±1.0
EOPD	91.3±0.6	46.2±1.8	41.7±1.4	87.2±0.7	55.6±0.8	64.4±0.9
TSD-KD	90.9±0.5	45.0±1.5	40.8±1.4	86.7±0.8	55.0±0.8	63.7±0.9
Prefix-OPD	90.2±0.5	43.5±1.4	39.6±1.4	85.9±0.6	53.9±0.7	62.3±0.7
TIP	91.6±0.5	47.3±1.3	42.6±1.3	87.7±0.7	56.3±0.7	65.2±0.8
TA-OPD	91.7±0.6	48.0±1.7	43.2±1.8	88.0±0.8	56.7±0.9	65.6±0.9
FiRe-OPD	92.0±0.5	48.6±1.6	43.8±1.6	88.3±0.7	57.1±0.8	66.0±0.9
CoRD (ours)	**92.7±0.5**	**50.2±1.5**	**45.3±1.4**	**89.1±0.6**	**58.4±0.7**	**67.4±0.8**

(b) pass@8 ↑

Method	MATH-500	AIME 2024	AIME 2025	AMC 2023	Minerva	Olympiad
Teacher	98.3	79.3	72.9	97.4	82.7	88.0
Initial student	96.5	61.6	56.1	94.9	73.8	80.4
SeqKD	96.7±0.3	62.7±2.4	57.0±2.5	95.1±0.5	74.6±1.1	81.0±1.1
Teacher-response KD	96.7±0.2	62.7±0.3	56.8±0.4	95.1±0.1	74.4±0.6	80.7±0.3
Dense-RKL	97.1±0.3	65.6±1.8	59.7±1.9	95.7±0.3	76.0±0.9	82.2±0.8
GKD (FKL)	97.1±0.2	65.6±1.8	60.1±0.8	95.7±0.5	76.4±1.3	82.4±1.1
GKD (JSD)	97.4±0.2	66.9±1.8	61.2±2.0	96.0±0.3	77.3±0.8	83.2±1.1
MiniLLM	97.2±0.4	65.7±2.4	60.2±3.0	95.8±0.5	76.6±1.6	82.6±1.7
SKD	97.2±0.4	65.8±2.1	60.2±2.9	95.8±0.5	76.7±1.1	82.6±1.0
DistiLLM	97.1±0.3	65.2±0.7	59.6±1.8	95.6±0.4	76.2±1.5	82.2±1.3
DistiLLM-2	97.1±0.6	65.7±3.8	60.0±3.9	95.7±0.8	76.2±2.4	82.2±2.3
Sampled OPD	97.3±0.5	66.5±3.0	61.0±2.4	95.9±0.5	77.0±1.9	83.0±1.6
EOPD	97.3±0.3	67.4±2.4	61.4±2.0	96.0±0.4	77.0±0.8	83.1±0.9
TSD-KD	97.3±0.4	66.6±2.8	60.8±3.5	95.9±0.5	76.8±2.1	82.9±1.7
Prefix-OPD	97.0±0.1	65.0±1.3	59.4±1.3	95.6±0.2	75.9±0.6	81.9±0.5
TIP	97.4±0.2	68.3±1.6	62.2±1.6	96.1±0.3	77.5±1.2	83.6±1.0
TA-OPD	97.7±0.2	70.6±0.6	64.6±1.1	96.6±0.4	79.2±1.1	85.1±1.2
FiRe-OPD	97.7±0.4	70.6±2.1	64.6±2.6	96.6±0.5	79.0±1.2	84.9±1.4
CoRD (ours)	**98.0±0.2**	**72.3±1.6**	**66.4±1.1**	**96.9±0.3**	**80.2±1.2**	**85.9±1.4**

Table 8: **Qwen2.5-14B-Instruct → Qwen2.5-1.5B-Instruct.** Per-benchmark reasoning performance under the common GPU-hour cap. Panel (a) reports mean@16; panel (b) reports pass@8. Scores are percentages (mean ± training-seed standard deviation); reference checkpoints have no training-seed deviation. Bold marks the highest mean among trained students in each column.

(a) mean@16 ↑

Method	MATH-500	AIME 2024	AIME 2025	AMC 2023	Minerva	Olympiad
Teacher	82.4	23.7	19.8	71.9	46.7	52.6
Initial student	56.8	7.1	5.8	45.3	28.7	32.8
SeqKD	58.9±0.5	8.0±1.3	6.6±1.2	47.1±0.6	29.8±0.7	34.0±0.7
Teacher-response KD	59.6±0.5	8.5±1.3	7.0±1.3	47.8±0.5	30.5±0.6	34.6±0.7
Dense-RKL	64.7±0.5	11.4±1.6	9.3±1.6	52.4±0.8	33.6±0.8	38.1±1.0
GKD (FKL)	63.9±0.6	10.9±1.5	8.9±1.5	51.6±0.7	33.0±0.8	37.4±0.8
GKD (JSD)	64.4±0.6	11.2±1.6	9.1±1.6	52.0±0.8	33.4±0.9	37.8±1.1
MiniLLM	62.8±0.6	10.1±1.6	8.2±1.7	50.7±0.8	32.3±0.9	36.4±0.9
SKD	63.4±0.7	10.5±1.7	8.5±1.5	51.1±0.7	32.7±0.8	36.9±1.0
DistiLLM	63.2±0.5	10.3±1.3	8.4±1.4	50.9±0.5	32.5±0.7	36.7±0.8
DistiLLM-2	64.9±0.5	11.5±1.3	9.4±1.4	52.3±0.6	33.6±0.7	38.2±0.7
Sampled OPD	64.2±0.6	11.0±1.7	9.0±1.5	51.8±0.8	33.2±0.8	37.6±0.9
EOPD	66.0±0.5	12.6±1.8	10.5±1.4	53.8±0.7	34.5±0.7	39.3±0.9
TSD-KD	66.5±0.6	12.9±1.6	10.8±1.6	54.2±0.7	34.9±0.8	39.8±1.0
Prefix-OPD	63.8±0.6	10.7±1.6	8.8±1.4	51.4±0.7	33.0±0.8	37.2±0.8
TIP	66.3±0.6	13.1±1.7	10.9±1.6	54.1±0.8	35.0±0.7	39.9±0.9
TA-OPD	66.8±0.5	13.4±1.4	11.2±1.4	54.5±0.7	35.3±0.7	40.3±0.9
FiRe-OPD	67.1±0.5	13.7±1.3	11.5±1.4	54.8±0.6	35.6±0.7	40.6±0.8
CoRD (ours)	**68.3±0.6**	**14.8±1.8**	**12.6±1.8**	**56.2±0.9**	**36.8±0.8**	**42.0±1.0**

(b) pass@8 ↑

Method	MATH-500	AIME 2024	AIME 2025	AMC 2023	Minerva	Olympiad
Teacher	94.3	41.6	34.8	89.5	70.2	75.1
Initial student	79.5	14.1	11.3	70.0	49.8	54.6
SeqKD	80.8±0.6	15.6±2.7	12.6±2.3	71.4±0.9	50.8±1.4	55.6±1.1
Teacher-response KD	81.7±1.7	16.9±2.9	13.6±2.8	72.5±2.2	52.3±2.4	56.9±2.9
Dense-RKL	84.3±0.7	21.3±2.0	17.0±2.1	75.5±0.8	55.0±0.8	59.7±0.7
GKD (FKL)	84.7±0.8	21.2±3.0	17.0±2.9	75.9±1.1	55.4±1.5	60.2±1.6
GKD (JSD)	84.9±0.4	21.6±2.9	17.2±2.9	76.1±0.6	55.8±1.2	60.5±1.2
MiniLLM	84.0±1.1	19.7±2.6	15.7±3.1	75.2±1.6	54.6±1.3	59.1±2.3
SKD	84.0±1.5	20.2±3.0	16.0±2.5	75.1±1.8	54.6±2.0	59.1±2.9
DistiLLM	84.4±0.5	20.3±2.5	16.2±2.5	75.6±0.5	55.1±1.2	59.7±1.0
DistiLLM-2	85.1±0.4	22.0±2.3	17.7±2.5	76.2±0.6	55.8±1.0	60.7±0.9
Sampled OPD	84.3±1.1	20.9±3.3	16.8±3.1	75.4±1.7	54.9±2.1	59.6±1.9
EOPD	86.7±0.6	24.8±3.1	20.4±2.4	78.6±0.8	58.3±0.8	63.4±1.7
TSD-KD	86.8±1.4	25.2±3.7	20.8±3.7	78.6±2.2	58.5±3.0	63.6±2.9
Prefix-OPD	85.4±1.2	21.5±2.5	17.4±2.7	76.8±1.7	56.7±2.5	61.2±2.2
TIP	85.8±0.3	24.6±2.4	20.1±2.4	77.5±0.1	57.3±0.3	62.4±0.4
TA-OPD	86.1±0.9	25.1±2.2	20.6±2.7	77.8±1.2	57.6±1.5	62.8±0.9
FiRe-OPD	86.4±1.4	25.9±3.1	21.3±2.7	78.2±1.5	58.2±2.3	63.4±2.2
CoRD (ours)	**87.2±0.7**	**27.8±3.0**	**23.3±2.8**	**79.6±1.0**	**59.5±1.3**	**64.7±1.6**

C.3 CONTROLLED ANALYSES

The shared settings and diagnostic protocols are given in Appendix B.5. Bold and underline mark the best and second-best distinct means, respectively, in columns with arrows within each panel; ties are marked equally. Shading identifies CoRD, and † denotes a proxy-token adaptation.

C.3.1 SELECTION, RECONSTRUCTION, AND SUPERVISION BUDGET

Table 9 reports the factorial comparison at $\rho = 0.25$ and 1,000 updates.

Both components contribute to accuracy. With uniform weights, CoRD selection raises mean@16 from 61.1 to 64.6; adding signed reconstruction raises it further to 66.9. LS also improves uniform selection to 63.5. The combined method has the highest mean@16 and pass@8, with a modest cost increase over uniform selection and weights: 397.3 versus 383.4 GPU-hours. These comparisons support the use of both selection and reconstruction.

Table 9: **Selection–reconstruction factorial study.** Qwen3-4B, $\rho = 0.25$, 1,000 updates. M16/P8 are six-task macro mean@16/pass@8 (%). Entries are mean ± sample SD over three training seeds. Bold/underline mark the best/second-best distinct means in each column; ties are marked equally.

Selection	Loss weights	M16 ↑	P8 ↑	GPU-h ↓
Uniform	Uniform	61.1±1.1	78.0±1.3	**383.4±3.7**
Uniform	Signed LS	63.5±0.8	79.9±1.0	<u>391.6±4.1</u>
CoRD	Uniform	<u>64.6±0.9</u>	<u>80.9±1.1</u>	395.9±3.9
CoRD	Signed LS (ours)	**66.9±0.9**	**82.9±0.8**	397.3±4.0

Table 10 varies only CoRD's configured supervision fraction at the same update count. At $\rho = 1$, the full-budget branch uses every valid position with weight $1/T$. Both the Dense-RKL reference and the full-budget CoRD endpoint are shown explicitly.

Table 10: **Quality and cost across supervision budgets.** Qwen3-4B, 1,000 updates. Accuracy and cost entries summarize the same three runs as mean ± sample SD. $f_{\neq 0}$ is the percentage of valid positions with nonzero loss weights over the complete run. Allocated memory is the per-run maximum across GPUs in the profiling interval, in GB/GPU.

Method	ρ	M16 ↑	P8 ↑	$f_{\neq 0}$ (%)	GPU-h ↓	Alloc. GB ↓
Dense-RKL	1.00	63.4±0.9	79.4±1.0	100.00±0.00	412.6±4.2	72.8±0.3
CoRD	0.10	65.5±0.8	81.6±1.1	9.63±0.10	**389.5±3.8**	**70.5±0.2**
CoRD	0.25	**66.9±0.9**	**82.9±0.8**	23.87±0.29	<u>397.3±4.0</u>	<u>71.4±0.3</u>
CoRD	0.50	<u>66.2±0.8</u>	<u>82.0±1.0</u>	48.21±0.41	405.6±4.1	72.1±0.3
CoRD	1.00	63.4±0.9	79.4±1.0	100.00±0.00	412.6±4.2	72.8±0.3

C.3.2 CORRECTION AND TRAINABLE-GRADIENT FIDELITY

Why correction fit alone is insufficient. At Dense-RKL update 1,000, signed mean pursuit achieves relative mean-correction error 0.069, but leaves more unexplained energy than CoRD selection (0.367 versus 0.217) and a larger batch-gradient error than CoRD/LS (0.316 versus 0.199). Thus, a small mean error can coexist with uncovered correction directions and poorer LoRA-gradient fidelity. These fixed-state comparisons support complementary checks of mean fit, directional coverage, and actual trainable gradients.

Table 11 gives the complete fixed-state results; Appendix B.5.2 defines the metrics and aggregation.

Table 11: **From correction geometry to LoRA gradients.** Qwen3-4B, $\rho = 0.25$; mean $\pm$ sample SD of seed-level averages over three seeds. All quantities are dimensionless. c_g is cosine similarity and e_g is relative gradient error; superscripts denote per-trajectory and length-weighted batch gradients. Uniform/LS and CoRD/uniform complete the factorial controls. Shared selected sets give identical η_{cov} for the two weighting rules. Bold/underline mark the best/second-best distinct means within each panel; † denotes a proxy-token adaptation.

Method / weights	$e_\mu \downarrow$	$\eta_{\text{cov}} \downarrow$	$c_g^{\text{traj}} \uparrow$	$e_g^{\text{traj}} \downarrow$	$c_g^{\text{batch}} \uparrow$	$e_g^{\text{batch}} \downarrow$
(a) Initialization						
Uniform / uniform	0.463±0.021	0.331±0.001	0.821±0.012	0.594±0.019	0.879±0.015	0.482±0.021
Uniform / LS	0.204±0.001	0.331±0.001	0.912±0.010	0.410±0.002	0.951±0.008	0.319±0.015
Norm top-m / uniform	0.402±0.018	0.258±0.014	0.867±0.013	0.512±0.018	0.918±0.011	0.402±0.016
GRAD-MATCH†	0.151±0.013	0.304±0.017	0.923±0.009	0.398±0.015	0.959±0.007	0.287±0.012
Signed mean pursuit	**0.083±0.010**	0.391±0.021	0.931±0.011	0.377±0.016	0.957±0.009	0.296±0.014
CoRD / uniform	0.312±0.017	**0.236±0.001**	0.909±0.010	0.427±0.014	0.950±0.009	0.324±0.013
CoRD / LS (ours)	0.158±0.002	**0.236±0.001**	**0.961±0.006**	**0.373±0.005**	**0.984±0.004**	**0.181±0.009**
(b) Dense-RKL update 1,000						
Uniform / uniform	0.428±0.024	0.323±0.002	0.801±0.014	0.621±0.021	0.861±0.017	0.522±0.022
Uniform / LS	0.198±0.003	0.323±0.002	0.899±0.011	0.408±0.001	0.938±0.010	0.352±0.016
Norm top-m / uniform	0.365±0.020	0.229±0.016	0.846±0.015	0.548±0.020	0.902±0.013	0.445±0.018
GRAD-MATCH†	0.136±0.011	0.282±0.018	0.911±0.010	0.430±0.016	0.948±0.008	0.321±0.013
Signed mean pursuit	**0.069±0.009**	0.367±0.022	0.918±0.010	0.405±0.015	0.951±0.008	0.316±0.013
CoRD / uniform	0.281±0.016	**0.217±0.003**	0.897±0.012	0.462±0.016	0.940±0.009	0.347±0.014
CoRD / LS (ours)	0.144±0.004	**0.217±0.003**	**0.954±0.007**	**0.350±0.005**	**0.981±0.004**	**0.199±0.010**

C.3.3 Correction geometry and conditional gains

Appendix B.5.3 details the paired records and landscape estimator.

Regional summaries. Table 12 reports arithmetic means and medians computed directly from equally weighted records, without smoothing. For each region, the pooled row uses all matching records across states and seeds: its mean is weighted by state-specific record counts, and its median is computed over the pooled records. In the low-dimension, high-cancellation region, the gain ranges from 0.0931 to 0.1891, with mean 0.1358, median 0.1362, and IQR 0.1183–0.1530. Across all records, 135/1,536 (8.8%) have negative gains.

Table 12: **CoRD gain by geometry and reference state.** High-cancellation region: $d_{\text{rel}} \leq 0.18$, $\chi \geq 0.70$. Weak-cancellation region: $d_{\text{rel}} \geq 0.30$, $\chi \leq 0.35$. Counts are trajectory-state records; gains are dimensionless relative-error reductions. Mean and median gains are reported separately from the raw records, without the local smoothing used in Figure 6.

	High cancellation, low d_{rel}			Weak cancellation, high d_{rel}		
State	n	Mean	Median	n	Mean	Median
Initialization	26	0.1274	0.1303	166	0.0019	0.0008
Update 1,000	48	0.1404	0.1379	52	0.0025	0.0002
Pooled	74	0.1358	0.1362	218	0.0021	0.0007

The descriptors are strongly correlated (Pearson $r = -0.814$). There are no records with $d_{\text{rel}} \leq 0.18$ and $\chi \leq 0.35$, and only seven with $d_{\text{rel}} \geq 0.30$ and $\chi \geq 0.70$. The landscape therefore characterizes gains under the observed joint geometry. It evaluates selection with signed reconstruction held fixed; evidence for the benefit of reconstruction itself comes from Table 9.

C.3.4 Optimizer transfer and numerical stability

Table 13 reports the cloned-AdamW comparison and signed-weight stability under the protocol in Appendix B.5.4.

Table 13: **Cloned optimizer steps and signed-weight stability.** Panels (a,b): mean ± sample SD of seed-level batch averages. $\Delta\mathcal{L}_{\text{full}}$ is reported in units of 10^{-3}. Dense-RKL is excluded from ranking and supplies the local-loss reference; its increment similarity to itself is an identity, marked by a dash. Panel (c): pooled maxima and failure counts, with the negative-weight fraction aggregated across seeds.

Method / weights	$c_{\Delta\theta}\uparrow$	$e_{\Delta\theta}\downarrow$	$\Delta\mathcal{L}_{\text{full}}$ $(10^{-3})\downarrow$
(a) Initialization			
Dense-RKL	—	—	-4.86 ± 0.21
Uniform / uniform	0.901 ± 0.010	0.449 ± 0.016	-3.52 ± 0.24
Uniform / LS	0.958 ± 0.007	0.291 ± 0.012	-4.24 ± 0.19
Norm top-m / uniform	0.929 ± 0.009	0.375 ± 0.014	-3.86 ± 0.22
GRAD-MATCH†	$\underline{0.968\pm0.006}$	$\underline{0.255\pm0.010}$	$\underline{-4.35\pm0.18}$
Signed mean pursuit	0.966 ± 0.006	0.263 ± 0.011	-4.31 ± 0.20
CoRD / uniform	0.954 ± 0.007	0.305 ± 0.012	-4.17 ± 0.17
CoRD / LS (ours)	$\mathbf{0.988\pm0.003}$	$\mathbf{0.158\pm0.008}$	$\mathbf{-4.64\pm0.16}$
(b) Dense-RKL update 1,000			
Dense-RKL	—	—	-2.44 ± 0.17
Uniform / uniform	0.881 ± 0.012	0.481 ± 0.018	-1.68 ± 0.18
Uniform / LS	0.949 ± 0.008	0.326 ± 0.013	-2.08 ± 0.16
Norm top-m / uniform	0.917 ± 0.010	0.401 ± 0.014	-1.89 ± 0.17
GRAD-MATCH†	$\underline{0.959\pm0.007}$	$\underline{0.289\pm0.011}$	$\underline{-2.14\pm0.14}$
Signed mean pursuit	0.958 ± 0.007	0.292 ± 0.012	-2.12 ± 0.16
CoRD / uniform	0.947 ± 0.008	0.331 ± 0.013	-2.06 ± 0.15
CoRD / LS (ours)	$\mathbf{0.985\pm0.004}$	$\mathbf{0.175\pm0.009}$	$\mathbf{-2.31\pm0.13}$

(c) CoRD reconstruction stability

Reference state	$\max\|w\|_1$	Negative mass (%)	$\max\log_{10}\kappa_2(R)$	Failures / attempts
Initialization	2.84	13.7 ± 0.8	2.36	$0/768$
Update 1,000	3.27	15.2 ± 1.1	2.61	$0/768$

C.3.5 Measured execution cost

Cost tradeoff. Relative to Dense-RKL, CoRD reduces backward time from 44.5 to 32.3 seconds per update, while selection and reconstruction take 6.3 seconds. Generation and teacher/student scoring remain similar because all candidate positions are still scored. The resulting end-to-end savings are therefore modest despite the sparse loss weights; the loss-weight fraction and runtime measure different aspects of sparsity.

Table 14 reports stage times and resources; Appendix B.5.5 specifies the measurement and aggregation.

Table 14: **End-to-end cost decomposition.** Qwen3-4B on eight A100 80GB GPUs. Panel (a) gives elapsed job seconds per update. The selection column includes scoring statistics, kernel construction, selection, and the weight solve, counting their costs beyond model scoring; the optimizer column includes synchronization and other update overhead. Panel (b) reports complete-run GPU-hours and teacher-scored positions (millions), plus interval throughput and peak memory (GB/GPU). Entries are mean ± sample SD over three seeds.

(a) Profiling interval: seconds per update

Method	Generate	Teacher	Student	Select	Backward	Optim./sync	Total ↓
Dense-RKL	71.8±1.1	31.7±0.5	28.4±0.5	0.2±0.0	44.5±0.7	5.1±0.2	181.7±1.9
Uniform	71.2±1.0	31.5±0.5	28.2±0.4	0.3±0.0	32.9±0.6	4.8±0.2	**168.9±1.7**
TIP-budget	71.9±1.2	31.8±0.5	28.5±0.5	2.4±0.1	35.1±0.6	4.9±0.2	174.6±1.8
TA-OPD-budget	72.1±1.1	31.9±0.6	28.6±0.4	3.6±0.2	35.3±0.7	5.0±0.2	176.5±1.9
Signed mean pursuit	72.6±1.3	32.2±0.6	28.9±0.5	22.4±1.0	37.1±0.8	5.1±0.2	198.3±2.5
CoRD	72.0±1.2	31.8±0.5	28.5±0.5	6.3±0.3	32.3±0.6	4.9±0.2	175.8±1.8

(b) Training resources and access

Method	GPU-h ↓	Alloc. GB ↓	Reserved GB ↓	Generated tok/s ↑	Teacher positions (M)
Dense-RKL	412.6±4.2	72.8±0.3	76.5±0.4	63.0±1.0	11.45±0.12
Uniform	**383.4±3.7**	**69.7±0.2**	**73.4±0.3**	**66.2±1.2**	11.18±0.10
TIP-budget	394.8±4.1	70.7±0.3	74.6±0.3	64.8±1.1	11.31±0.11
TA-OPD-budget	399.6±4.3	70.9±0.3	74.8±0.4	64.4±1.0	11.36±0.13
Signed mean pursuit	447.1±5.0	73.6±0.4	77.5±0.5	57.6±1.2	11.42±0.12
CoRD	397.3±4.0	71.4±0.3	75.3±0.4	64.3±1.1	11.30±0.11

C.4 TOKEN-SELECTION RESULTS

The diagnostic protocol is specified in Appendix B.6.

Regional summaries. Table 15 uses raw counts without smoothing. Intervals are left-closed and right-open, except that an upper endpoint of one is included. Rates pool tokens across the diagnostic trajectories.

Table 15: **Observed token-selection rates by disagreement and novelty.** Percentile intervals refer to ranks within each trajectory. Counts are tokens from the update-1,000 diagnostic; percentages use the displayed selected and total counts directly.

p_t^{RKL}	p_t^{nov}	Tokens	Selected	Rate (%)
$[0.40, 0.65)$	$[0.80, 1]$	6,067	5,866	96.69
$[0.80, 1]$	$[0, 0.25)$	5,462	0	0.00
$[0, 0.25)$	$[0.80, 1]$	4,778	2,767	57.91
$[0.80, 1]$	$[0.80, 1]$	5,843	5,842	99.98

Within the high-novelty region ($p_t^{\mathrm{nov}} \geq 0.80$), selection rates rise from 57.91% at low disagreement ($p_t^{\mathrm{RKL}} < 0.25$) to 96.69% at moderate disagreement ($0.40 \leq p_t^{\mathrm{RKL}} < 0.65$) and 99.98% at high disagreement ($p_t^{\mathrm{RKL}} \geq 0.80$). This variation is consistent with the downward-sloping selection-rate contours: selection rates vary with disagreement even at high novelty. The zero count in the high-disagreement, low-novelty region is an empirical observation in this sample.

Reading the landscape. The landscape describes how selection varies with recorded novelty and disagreement. Density contours summarize token support, while selection-rate contours summarize selection frequency. The records contain final descriptors; checking them against a greedy replay would additionally require per-pivot residual histories. Rates are descriptive aggregates of tokens within trajectories. Reconstruction and gradient fidelity are evaluated separately in Appendix C.3.

C.5 SPECTRAL COMPRESSIBILITY AND TOKEN BUDGETS

We separate two budget questions: how the retained fraction affects trained-model quality and how many tokens suffice to reconstruct a fixed response's mean correction. Theorem 1 informs the latter through the correction-kernel spectrum.

Training quality peaks at an intermediate budget. At 1,000 updates, Qwen3-4B macro mean@16 is 65.5, 66.9, 66.2, and 63.4 for $\rho = 0.10, 0.25, 0.50, 1.00$, respectively (Table 10). The full-budget branch recovers Dense-RKL, while $\rho = 0.25$ has the highest mean among these tested settings. We next assess reconstruction fidelity with each response and model state held fixed.

The fixed-state scan and spectral certificate are defined in Appendix B.7.

Spectral information tracks reconstruction difficulty. Figure 7 shows a geometry-dependent budget transition: instances with smaller interpolated spectral budgets typically reach low error at smaller empirical budgets. Across all 1,536 instances, the two thresholds have Spearman correlation 0.695. The median spectral-minus-empirical budget gap is 0.114, and the median spectral-to-empirical ratio is 1.668. The empirical grid threshold is at or below the interpolated spectral threshold in 1,535 instances (99.93%). At $\rho = 0.25$, mean e_μ is 0.0692 at initialization and 0.0592 at update 1,000.

These results support the spectral certificate as a generally conservative indicator of the budget needed for output-head mean reconstruction. The certificate combines eigenvalues with $\|\mu\|_2$; reconstruction error also depends on the mean's alignment with individual eigenspaces.

At initialization, the median interpolated spectral and empirical grid budgets are 0.3163 and 0.18, respectively. At update 1,000, they are 0.2816 and 0.16. The diagnostic measures reconstruction budgets using shared trajectories at each fixed state. Training quality across budgets is evaluated separately in Table 10.

Takeaway. A budget for preserving a dense update and a budget for improving reasoning solve different problems. Geometry can guide candidate retention levels, while trained-model quality remains the criterion for choosing among them.

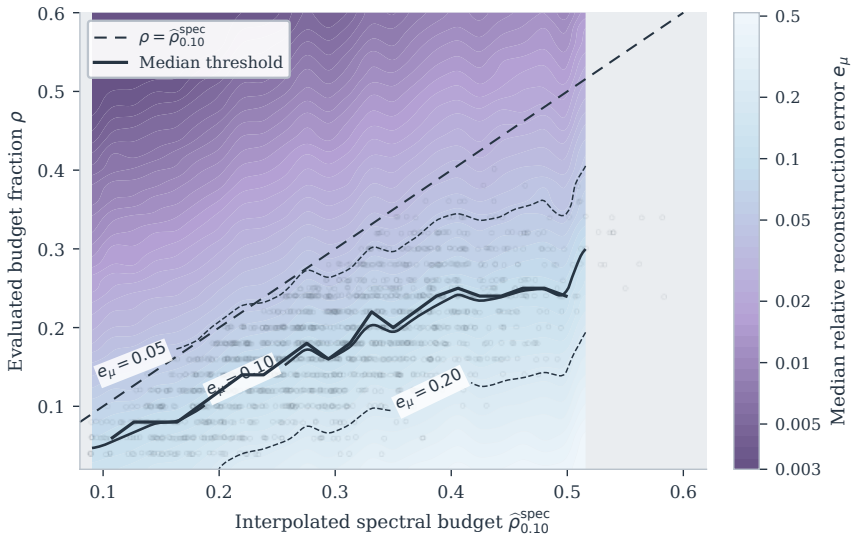

Figure 7: **Spectral budgets track the transition to low reconstruction error.** Color shows the smoothed bin-median e_μ at each scanned budget; darker purple means lower error. Contours mark $e_\mu = 0.05, 0.10, 0.20$. Points show the interpolated spectral and first empirical threshold budgets for 1,536 trajectory-state instances. The solid line is the binned median empirical threshold; the dashed diagonal marks equality of the two budgets. Gray regions lack sufficient support.

D PROOFS

D.1 DERIVATIONS AND GEOMETRIC PROPERTIES

D.1.1 LOGIT DERIVATIVES AND THE CORRECTION KERNEL

Fix one sampled prefix and suppress its position index. With $q = \mathrm{softmax}(z)$ and a fixed teacher distribution p, the softmax Jacobian is $\partial q/\partial z = \mathrm{diag}(q) - qq^\top$. Multiplying this Jacobian by $\nabla_q \ell = \log(q/p) + \mathbf{1}_V$ gives $r = q \odot [\log(q/p) - \ell \mathbf{1}_V]$, where the logarithm and division act elementwise and $\odot$ is elementwise multiplication. The constant terms cancel because $\sum_u q_u = 1$. For an independent linear output head, differentiating $z = Wh$ at fixed h gives $\nabla_W \ell = rh^\top$. In the adapter training setting, this outer product represents the direct head correction while W remains fixed.

The Frobenius identity $\langle r_i h_i^\top, r_j h_j^\top \rangle_F = \mathrm{tr}(h_i r_i^\top r_j h_j^\top) = (r_i^\top r_j)(h_i^\top h_j)$ proves Equation (3). Thus K is positive semidefinite even when individual entries are negative. All identities use the full-vocabulary distributions in Equation (1).

D.1.2 PIVOTED CHOLESKY AND PROJECTION COORDINATES

Initialize A with no columns and $v_t = K_{tt}$. At iteration j, choose b_j by the largest-residual rule in Section 4.2. Let $A_{<j} \in \mathbb{R}^{T\times(j-1)}$ denote the existing factor. One kernel-column evaluation gives

$$a_j = \frac{K_{:b_j} - A_{<j} A_{b_j,<j}^\top}{\sqrt{v_{b_j}}}, \qquad A \leftarrow [A_{<j}, a_j], \qquad v \leftarrow v - a_j \odot a_j, \tag{22}$$

Positive pivots select linearly independent columns in exact arithmetic.

To identify the factor, let $Q_{<j}$ be the orthonormal basis constructed from earlier selected corrections. The next direction is $Q_{:j} = (I - Q_{<j}Q_{<j}^\top)c_{b_j}/\sqrt{v_{b_j}}$. Inductively, $A_{<j} = F^\top Q_{<j}$, and Equation (22) yields $a_j = F^\top Q_{:j}$. Consequently, $A = F^\top Q$ and $v_t = K_{tt} - \sum_{j=1}^s A_{t,j}^2 = \|(I - QQ^\top)c_t\|_2^2$. Each new direction is orthogonal to earlier selected columns, so $R = Q^\top F_B = A_{B,:}^\top$ is upper triangular with positive diagonal. Neither Q nor F is constructed by the algorithm.

D.1.3 RECONSTRUCTION AND THE MEAN-ERROR BOUND

Since $F_B = QR$ and $Q^\top Q = I$, the least-squares objective decomposes as $\|F_B u - \mu\|_2^2 = \|Ru - Q^\top\mu\|_2^2 + \|(I - QQ^\top)\mu\|_2^2$. The second term does not depend on u, while $Q^\top\mu = A^\top \mathbf{1}_T/T$. Solving Equation (5) therefore gives $F_B w = QQ^\top\mu = P_B\mu$. For this orthogonal projection,

$$\begin{aligned} \|\mu - \widehat{\mu}\|_2^2 &= \left\| \frac{1}{T}\sum_{t=1}^T (I - P_B)c_t \right\|_2^2 \le \frac{1}{T}\sum_{t=1}^T \|(I - P_B)c_t\|_2^2 = \frac{1}{T}\sum_{t=1}^T v_t, \\ \mu^\top\widehat{\mu} &= \|\widehat{\mu}\|_2^2, \qquad \|\mu - \widehat{\mu}\|_2^2 = \|\mu\|_2^2 - \|\widehat{\mu}\|_2^2. \end{aligned} \tag{23}$$

The inequality is Jensen's inequality for the squared norm. Expanding the selected span cannot increase the mean projection error. When this span contains every column of F, the reconstruction equals the full mean, including the multiplicity of repeated columns.

At full budget, uniform weights recover the dense loss directly without factorization. In the zero-kernel case, every c_t is zero, so the reconstructed head correction is zero for the fallback subset as well. Neither branch requires an inverse of a singular Gram matrix.

D.1.4 COVERAGE AND MEAN MATCHING

Let $S = F^\top(I - P_B)F$ be the residual Gram matrix. Adding position i with $S_{ii} > 0$ reduces the total residual energy by $\|S_{:i}\|_2^2/S_{ii}$ and reduces the squared mean error by $(\sum_t S_{it})^2/(T^2 S_{ii})$. These quantities differ from the pivot score S_{ii}. The pivoted-Cholesky rule uses the diagonal, updating every candidate's score from the newly accessed column; reconstruction subsequently minimizes mean error over the selected span.

For example, consider $c_1 = (1, 0)^\top$, $c_2 = (-1, 0)^\top$, and $c_3 = (0, \delta)^\top$ with $0 < \delta < 1$. A budget of one selects c_1, whose optimal coefficient is zero because $\mu = (0, \delta/3)^\top$. Selecting c_3 recovers the mean exactly. This example exhibits the effect of cancellation on the selection objective. It also has an exact realization as output-head corrections: take $W = 0$, $r = (1, -1)^\top/\sqrt{2}$, $p = \mathrm{softmax}(-2r)$, and hidden states equal to the three displayed vectors. Then $q = (1/2, 1/2)^\top$, Equation (2) gives the common derivative r, and the head corrections have the same Gram matrix.

D.1.5 Single-trajectory head descent and batching

Consider an independently parameterized output head with hidden states held fixed. In this setting, $\mu = \mathrm{vec}(\nabla_W \mathcal{L}_{\mathrm{full}})$ and $\widehat{\mu} = \mathrm{vec}(\nabla_W \mathcal{L}_{\mathrm{CoRD}})$ at the point where B, w were computed. Let $\widehat{C} = \mathrm{unvec}(\widehat{\mu})$, and suppose $\mathcal{L}_{\mathrm{full}}$ has an L-Lipschitz gradient along the step. The smoothness inequality and Equation (23) yield

$$\mathcal{L}_{\mathrm{full}}(W - \alpha_{\mathrm{step}}\widehat{C}) \le \mathcal{L}_{\mathrm{full}}(W) - \alpha_{\mathrm{step}}\left(1 - \frac{L\alpha_{\mathrm{step}}}{2}\right)\|\widehat{\mu}\|_2^2. \tag{24}$$

Strict decrease follows for $\widehat{\mu} \neq 0$ and $0 < \alpha_{\mathrm{step}} < 2/L$. This result is restricted to a Euclidean head-gradient step on one fixed trajectory. The adapter update uses different parameter Jacobians and AdamW preconditioning.

The implemented batch gradient is a length-weighted sum of trajectory gradients. Different trajectories have different projectors, so summation can reverse the alignment of their reconstructed head gradients. For two length-three trajectories with corrections $[(10, 0), (-9, 0), (0, 3)]$ and $[(0, -10), (0, 9), (-3, 0)]$, selecting one position per trajectory gives reconstructed means $(1/3, 0)$ and $(0, -1/3)$. The dense batch mean is $(-1/3, 1/3)$, while the reconstructed batch mean is $(1/6, -1/6)$; their inner product is $-1/9$. Batch-level descent therefore requires alignment of the aggregated gradients.

D.1.6 Signed coefficients and adapter-gradient error

Signed coefficients arise even with exact reconstruction. For corrections $(2, 0)^\top$, $(0, 2)^\top$, and three copies of $(-1, 0)^\top$, the first two positions form the selected basis at budget two. Their mean is $(-1/5, 2/5)^\top$, requiring weights $(-1/10, 1/5)^\top$. Clipping this negative coefficient changes the reconstructed mean, and normalizing the weights changes its scale. The surrogate in Equation (6) is differentiated only at the current parameters; its coefficients are recomputed after every update.

Let $g_t = \nabla_\theta \ell_t \in \mathbb{R}^D$, where D is the number of trainable adapter parameters. With $J_t = \partial z_t/\partial\theta \in \mathbb{R}^{V\times D}$, the actual gradient is $g_t = J_t^\top r_t$. To express the effect of proxy mismatch, suppose a common map $M \in \mathbb{R}^{D\times Vd}$ satisfies $g_t = Mc_t + e_t$ with $\max_t \|e_t\|_2 \le \epsilon$. Define $g_{\mathrm{full}} = T^{-1}\sum_t g_t$ and $\widehat{g} = \sum_j w_j g_{b_j}$. Then

$$\|g_{\mathrm{full}} - \widehat{g}\|_2 \le \|M\|_2\sqrt{\frac{1}{T}\sum_{t=1}^{T} v_t} + \epsilon\big(1 + \|w\|_1\big). \tag{25}$$

Indeed, the difference equals $M(\mu - \widehat{\mu}) + T^{-1}\sum_t e_t - \sum_j w_j e_{b_j}$; the result follows from Equation (7) and the triangle inequality. Applying this bound requires separate control of $\|M\|_2$ and ϵ. It identifies coefficient amplification as a source of sensitivity to differences between head geometry and adapter gradients.

D.2 Correction-space compressibility

This subsection proves Theorem 1 and Corollary 6. The argument is deterministic and applies to the actual maximum-residual rule in Section 4.2. No independence, distributional, or incoherence assumption is imposed on the token corrections.

D.2.1 Setup and conventions

Fix $T \ge 1$ correction vectors and define

$$F = [c_1, \ldots, c_T] \in \mathbb{R}^{Vd\times T}, \qquad K = F^\top F, \qquad \mu = \frac{1}{T}F\mathbf{1}_T, \qquad r_\star = \mathrm{rank}(F). \tag{26}$$

Let $\lambda_1 \geq \cdots \geq \lambda_T \geq 0$ denote the eigenvalues of K, including zeros. Starting with $B_0 = \varnothing$ and $P_0 = 0$, define the ordered greedy sequence by

$$\begin{aligned} v_t^{(i)} &= \|(I - P_i)c_t\|_2^2, & \delta_i &= \max_{t \in [T]} v_t^{(i)}, \\ b_{i+1} &= \min \arg\max_{t \notin B_i} v_t^{(i)}, & B_{i+1} &= (B_i, b_{i+1}), \end{aligned} \tag{27}$$

where P_i is the orthogonal projector onto $\mathrm{span}(F_{B_i})$. A new position is selected only when $\delta_i > 0$. For $m < r_\star$, write

$$\begin{aligned} U_m &= [T] \setminus B_m, \quad N_m = T - m, & S_m &= F^\top (I - P_m) F, \\ \widehat{\mu}_m &= P_m \mu, & \mathcal{E}_m &= \|(I - P_m)\mu\|_2^2, \quad L_i = \frac{\sum_{j=i+1}^{T} \lambda_j}{T - i} \quad (0 \leq i < m). \end{aligned} \tag{28}$$

The vector $\mathbf{1}_U \in \mathbb{R}^T$ is the indicator of a set U; e_t is the t-th standard coordinate vector. Empty products equal one. For budget indices $m \geq r_\star$, retain the exact projector $P_m = P_{r_\star}$ and reconstruction, so $\mathcal{E}_m = 0$; selection ends at rank exhaustion. Norms of vectors are Euclidean and $\|\cdot\|_2$ for matrices is the spectral norm. All statements up to Appendix D.2.7 use exact arithmetic.

For reference, the envelope and its determinant-only specialization are

$$\begin{aligned} \mathcal{Q}_m(K) &= \min_{1 \leq n \leq m+1} \min_{0 \leq k < n} \left(\frac{\prod_{j=1}^{n} \lambda_j}{N_m \prod_{i=0}^{k-1} L_i} \right)^{1/(n-k)}, \\ \mathcal{D}_m(K) &= \min_{1 \leq n \leq m+1} \left(\frac{\prod_{j=1}^{n} \lambda_j}{N_m} \right)^{1/n}. \end{aligned} \tag{29}$$

Before rank exhaustion, all eigenvalues and tail averages appearing in denominators are positive. The $k = 0$ choices immediately imply $\mathcal{Q}_m(K) \leq \mathcal{D}_m(K)$.

D.2.2 Projection geometry and residual support

Lemma 2 (Projection and support-aware certificates)**.** *Every positive pivot adds one independent correction. Least-squares reconstruction on the selected columns equals $P_m\mu$, and, for $m < r_\star$,*

$$\begin{aligned} \mathcal{E}_m &\leq \frac{1}{T^2} \left(\sum_{t \in U_m} \sqrt{v_t^{(m)}} \right)^2 \leq \frac{N_m \operatorname{tr}(S_m)}{T^2} \leq \frac{N_m^2 \delta_m}{T^2}, \\ \mathcal{E}_m &\leq \frac{N_m \|S_m\|_2}{T^2}. \end{aligned} \tag{30}$$

The errors $\mathcal{E}_i$ and pivots δ_i are nonincreasing. After $r_\star$ positive pivots, every correction and their mean are recovered exactly.

Proof. If $\delta_i > 0$, the selected vector has a nonzero component orthogonal to the current span. It is therefore independent of the previous columns. For $m > 0$ positive pivots, write a thin QR factorization $F_{B_m} = QR$, where $Q^\top Q = I_m$ and R is invertible. Orthogonality gives, for every coefficient vector w,

$$\begin{aligned} \|F_{B_m} w - \mu\|_2^2 &= \|Q(Rw - Q^\top \mu) - (I - QQ^\top)\mu\|_2^2 \\ &= \|Rw - Q^\top \mu\|_2^2 + \|(I - QQ^\top)\mu\|_2^2 \\ &\geq \|(I - P_m)\mu\|_2^2. \end{aligned} \tag{31}$$

Equality holds uniquely at $Rw = Q^\top \mu$, proving the reconstruction identity. At $m = 0$, the same identity holds by convention.

Put $z_t = (I - P_m)c_t$. Then $z_t = 0$ for $t \in B_m$, and hence

$$\begin{aligned} T^2\mathcal{E}_m &= \left\| \sum_{t \in U_m} z_t \right\|_2^2 \\ &\le \left(\sum_{t \in U_m} \|z_t\|_2 \right)^2 \\ &\le \left(\sum_{t \in U_m} 1^2 \right) \left(\sum_{t \in U_m} \|z_t\|_2^2 \right) \\ &= N_m \sum_{t=1}^{T} v_t^{(m)} = N_m \operatorname{tr}(S_m) \\ &\le N_m \sum_{t \in U_m} \delta_m = N_m^2 \delta_m. \end{aligned} \tag{32}$$

The first two inequalities are the triangle and Cauchy–Schwarz inequalities. Alternatively, positive semidefiniteness gives

$$T^2\mathcal{E}_m = \mathbf{1}_{U_m}^\top S_m \mathbf{1}_{U_m} \le \|S_m\|_2 \|\mathbf{1}_{U_m}\|_2^2 = N_m \|S_m\|_2. \tag{33}$$

Nested selected spans imply $P_{i+1} - P_i \succeq 0$ and thus, for every vector x,

$$\|(I - P_{i+1})x\|_2^2 = \|(I - P_i)x\|_2^2 - \|(P_{i+1} - P_i)x\|_2^2 \le \|(I - P_i)x\|_2^2. \tag{34}$$

Use $x = c_t$ and then maximize to prove monotonicity of δ_i; use $x = \mu$ for $\mathcal{E}_i$. If $i < r_\star$, some column lies outside the selected span, so $\delta_i > 0$. At $i = r_\star$, the selected span is the column space of F, so all residuals vanish. If $r_\star = 0$, every column is zero from the outset. □

D.2.3 Spectral and determinant inequalities

Lemma 3 (Spectral lower bounds on greedy pivots)**.** *For every* $0 \le i < r_\star$,

$$\delta_i \ge \frac{\operatorname{tr}(S_i)}{T - i} \ge \frac{\sum_{j=i+1}^{T} \lambda_j}{T - i} > 0. \tag{35}$$

Proof. The i selected diagonal entries of S_i are zero, and each remaining entry is at most δ_i. This proves the first inequality. We prove the required variational trace inequality directly. Take a singular value decomposition of F with nonzero left singular vectors $u_1, \dots, u_{r_\star}$ and squared singular values $\lambda_1, \dots, \lambda_{r_\star}$. Complete the u_j to an orthonormal basis of $\mathbb{R}^{Vd}$, and set $\nu_j = \lambda_j$ for $j \le r_\star$ and $\nu_j = 0$ for the remaining basis vectors. For $a_j = \|P_i u_j\|_2^2$, we have $0 \le a_j \le 1$ and $\sum_j a_j = \operatorname{tr}(P_i) = i$. If $i \ge 1$, then

$$\begin{aligned} \operatorname{tr}(P_i F F^\top) &= \sum_j \nu_j a_j \\ &\le \sum_{j=1}^{i} \lambda_j a_j + \lambda_i \sum_{j>i} a_j \\ &= \sum_{j=1}^{i} \lambda_j a_j + \lambda_i \sum_{j=1}^{i} (1 - a_j) \\ &= \sum_{j=1}^{i} \lambda_j - \sum_{j=1}^{i} (\lambda_j - \lambda_i)(1 - a_j) \\ &\le \sum_{j=1}^{i} \lambda_j. \end{aligned} \tag{36}$$

For $i = 0$, both sides are zero. Therefore

$$\begin{aligned}(T-i)\delta_i &\geq \sum_{t \notin B_i} (S_i)_{tt} = \operatorname{tr}(S_i) \\ &= \operatorname{tr}(FF^\top) - \operatorname{tr}(P_i F F^\top) \\ &\geq \sum_{j=1}^{T} \lambda_j - \sum_{j=1}^{i} \lambda_j = \sum_{j=i+1}^{T} \lambda_j > 0.\end{aligned} \tag{37}$$

Division by $T - i > 0$ completes the proof. □

Lemma 4 (Determinants of orthogonal Gram compressions)**.** *Let $Z \in \mathbb{R}^{T\times n}$ satisfy $Z^\top Z = I_n$, and set $C = Z^\top K Z$. Its ordered eigenvalues $\gamma_1 \geq \cdots \geq \gamma_n$ satisfy $0 \leq \gamma_j \leq \lambda_j$. Consequently,*

$$0 \leq \det C = \prod_{j=1}^{n} \gamma_j \leq \prod_{j=1}^{n} \lambda_j. \tag{38}$$

If $Y = FZ = [y_1, \ldots, y_n]$ has independent columns and d_j is the squared distance of y_j from the span of its predecessors, then

$$\det(Y^\top Y) = \prod_{j=1}^{n} d_j. \tag{39}$$

Proof. Positive semidefiniteness follows from $K \succeq 0$. For the upper eigenvalue bound, let $E_j \subseteq \mathbb{R}^n$ be the span of the first j orthonormal eigenvectors of C, and let $H_j \subseteq \mathbb{R}^T$ be the span of eigenvectors of K corresponding to $\lambda_j, \ldots, \lambda_T$. Since Z is an isometry, ZE_j has dimension j, while H_j has dimension $T - j + 1$. Their intersection contains a nonzero vector $x = Za$, where $a \in E_j$. Expanding a and x in these eigenbases gives the complete Rayleigh-quotient comparison

$$\gamma_j \leq \frac{a^\top C a}{a^\top a} = \frac{x^\top K x}{x^\top x} \leq \lambda_j. \tag{40}$$

Multiplication over j proves Equation (38).

For the second assertion, Gram–Schmidt gives $Y = QR$ with $Q^\top Q = I_n$ and an upper triangular matrix R satisfying $R_{jj} = \sqrt{d_j} > 0$. Indeed, the j-th new unit vector is the normalized residual of y_j, so its inner product with y_j is $\sqrt{d_j}$ and its inner product with each preceding column is zero. It follows that

$$\det(Y^\top Y) = \det(R^\top Q^\top Q R) = \det(R^\top R) = \left(\prod_{j=1}^{n} R_{jj}\right)^2 = \prod_{j=1}^{n} d_j. \tag{41}$$

□

D.2.4 Proof of the spectral compression theorem

Proof of the upper bound in Theorem 1. Lemma 2 proves the least-squares identity and rank recovery. Fix $0 \leq m < r_\star$, abbreviate $N = N_m$ and $U = U_m$, and set

$$s = \frac{T^2 \mathcal{E}_m}{N^2}. \tag{42}$$

If $\mathcal{E}_m = 0$, the upper bound is immediate because its right-hand side is nonnegative. Hence assume $s > 0$. Lemmas 2 and 3 imply, for every $0 \leq i < m$,

$$\delta_i \geq \delta_m \geq s, \qquad \delta_i \geq L_i > 0. \tag{43}$$

Fix $n \in \{1, \ldots, m+1\}$ and form

$$\begin{aligned}&Z_n = [e_{b_1}, \ldots, e_{b_{n-1}}, \mathbf{1}_U/\sqrt{N}], \\ &Y_n = FZ_n = [c_{b_1}, \ldots, c_{b_{n-1}}, F\mathbf{1}_U/\sqrt{N}], \qquad C_n = Y_n^\top Y_n = Z_n^\top K Z_n.\end{aligned} \tag{44}$$

The first $n-1$ coordinate vectors have disjoint support from $\mathbf{1}_U$, and the latter has squared norm N. Thus $Z_n^\top Z_n = I_n$. The first $n-1$ squared Gram–Schmidt residuals of Y_n are exactly $\delta_0, \ldots, \delta_{n-2}$. Its last squared residual, denoted by d_n, satisfies

$$\begin{aligned} d_n &= \frac{1}{N}\|(I - P_{n-1})F\mathbf{1}_U\|_2^2 \\ &= \frac{1}{N}\|(I - P_m)F\mathbf{1}_U\|_2^2 + \frac{1}{N}\|(P_m - P_{n-1})F\mathbf{1}_U\|_2^2 \\ &\geq \frac{1}{N}\|(I - P_m)F\mathbf{1}_U\|_2^2 \\ &= \frac{1}{N}\|(I - P_m)F\mathbf{1}_T\|_2^2 = \frac{T^2}{N}\mathcal{E}_m = Ns > 0. \end{aligned} \tag{45}$$

The second line uses nested orthogonal projections. The penultimate equality uses $(I - P_m)F\mathbf{1}_{B_m} = 0$. For $n = 1$, the same calculation has $P_{n-1} = P_0 = 0$ and no preceding columns. All residuals are positive, so Y_n has independent columns.

For every $k \in \{0, \ldots, n-1\}$, apply Lemma 4, preserve k of the spectral pivot floors, and use s for the remaining pivots:

$$\begin{aligned} \prod_{j=1}^{n} \lambda_j &\geq \det C_n = d_n \prod_{i=0}^{n-2} \delta_i \\ &\geq Ns \prod_{i=0}^{n-2} \delta_i \\ &= Ns \left(\prod_{i=0}^{k-1} \delta_i\right)\left(\prod_{i=k}^{n-2} \delta_i\right) \\ &\geq Ns \left(\prod_{i=0}^{k-1} L_i\right) s^{n-1-k} = Ns^{n-k} \prod_{i=0}^{k-1} L_i. \end{aligned} \tag{46}$$

All divisions below are legitimate: $N \geq 1$, $n - k \geq 1$, and $i \leq k-1 \leq m-1 < r_\star$ implies $L_i > 0$. Taking the $(n-k)$-th root, and then minimizing, proves

$$\begin{aligned} s &\leq \left(\frac{\prod_{j=1}^{n} \lambda_j}{N \prod_{i=0}^{k-1} L_i}\right)^{1/(n-k)}, \\ \mathcal{E}_m &= \frac{N^2}{T^2} s \leq \frac{N^2}{T^2} \mathcal{Q}_m(K) \leq \frac{N^2}{T^2} \mathcal{D}_m(K). \end{aligned} \tag{47}$$

At $m = 0$, the only pair is $(n, k) = (1, 0)$; the result reduces to $\|\mu\|_2^2 \leq \lambda_1/T$. Additional constant-sharpness and two-token results are proved in Appendix D.2.6. □

The proof also gives a certificate that retains the observed pivots. In particular, before rank exhaustion,

$$\mathcal{E}_m \leq \frac{N_m}{T^2} \min_{1 \leq n \leq m+1} \frac{\prod_{j=1}^{n} \lambda_j}{\prod_{i=0}^{n-2} \delta_i}. \tag{48}$$

This follows from the first two lines of Equation (46); the zero-error case is immediate. One may take the minimum of this expression, the spectral envelope, and the support-aware certificates in Equation (30).

D.2.5 SPECTRAL DECAY AND CERTIFIED TOKEN BUDGETS

The following finite profiles retain both minimization indices and require only a spectral decay upper envelope. For $0 \le k < T$, let $(T)_k = \prod_{i=0}^{k-1}(T-i) = T!/(T-k)!$. For $0 \le m < T$, define

$$\begin{aligned} H_{\mathrm{pol}}(m) &= C_\lambda \left(\frac{N_m}{T}\right)^2 \min_{\substack{1\le n\le m+1\\ 0\le k<n}} \left[\frac{(T)_k}{N_m}\left(\frac{n!}{k!}\right)^{-p}\right]^{1/(n-k)}, \\ H_{\exp}(m) &= C_\lambda \left(\frac{N_m}{T}\right)^2 \min_{\substack{1\le n\le m+1\\ 0\le k<n}} \left(\frac{(T)_k}{N_m}\right)^{1/(n-k)} \exp\left[-\frac{\beta(n+k-1)}{2}\right]. \end{aligned} \tag{49}$$

Set $H_{\mathrm{pol}}(T) = H_{\exp}(T) = 0$.

Proposition 5 (Finite spectral-decay profiles)**.** *If $\lambda_j \le C_\lambda j^{-p}$ for $C_\lambda, p > 0$, then $\mathcal{E}_m \le H_{\mathrm{pol}}(m)$. If $\lambda_j \le C_\lambda \exp[-\beta(j-1)]$ for $C_\lambda, \beta > 0$, then $\mathcal{E}_m \le H_{\exp}(m)$. In each statement, $\mathcal{E}_m = 0$ after rank exhaustion. For either corresponding profile H, define*

$$m_H(\varepsilon) = \min\{m \in \{0,\dots,T\} : H(m) \le \varepsilon^2\}. \tag{50}$$

This is a sufficient budget, and it can be replaced by $\min\{r_\star, m_H(\varepsilon)\}$ when the rank is known. Both profiles satisfy $H(0) = C_\lambda/T$.

Proof. It suffices to consider $m < r_\star$, since the error is zero otherwise. For all used indices, $L_i \ge \lambda_{i+1}/(T-i) > 0$. Consequently, for every admissible pair (n,k),

$$\begin{aligned} \frac{\prod_{j=1}^n \lambda_j}{N_m \prod_{i=0}^{k-1} L_i} &\le \frac{(T)_k}{N_m} \frac{\prod_{j=1}^n \lambda_j}{\prod_{i=0}^{k-1} \lambda_{i+1}} \\ &= \frac{(T)_k}{N_m} \prod_{j=k+1}^{n} \lambda_j. \end{aligned} \tag{51}$$

Under polynomial decay, the remaining product satisfies

$$\prod_{j=k+1}^{n} \lambda_j \le \prod_{j=k+1}^{n} (C_\lambda j^{-p}) = C_\lambda^{n-k} \left(\prod_{j=k+1}^{n} j\right)^{-p} = C_\lambda^{n-k} \left(\frac{n!}{k!}\right)^{-p}. \tag{52}$$

Under exponential decay,

$$\begin{aligned} \prod_{j=k+1}^{n} \lambda_j &\le C_\lambda^{n-k} \exp\left[-\beta \sum_{j=k+1}^{n} (j-1)\right] \\ &= C_\lambda^{n-k} \exp\left[-\frac{\beta(n-k)(n+k-1)}{2}\right]. \end{aligned} \tag{53}$$

Substitute into Equation (51), take the positive $(n-k)$-th root, multiply by $(N_m/T)^2$, and minimize over all pairs in Equation (29). This proves both profiles. The budget set is nonempty because it contains T, and its chosen element satisfies $\mathcal{E}_m \le H(m) \le \varepsilon^2$. If the rank is smaller, exact recovery supplies the asserted cap. No monotonicity assumption on the displayed profiles is needed. At $m = 0$, only $(n,k) = (1,0)$ is admissible, giving $H(0) = C_\lambda/T$. □

Corollary 6 (Closed-form sufficient token budgets)**.** *Suppose $r_\star > 0$ and let $\varepsilon > 0$. If $\lambda_j \le C_\lambda j^{-p}$ for $C_\lambda, p > 0$, then $\mathcal{E}_m \le C_\lambda e^p/(m+1)^p$ for $m < r_\star$; a sufficient budget for $\|\mu - \widehat{\mu}_m\|_2 \le \varepsilon$ is*

$$m_{\mathrm{pol}} = \min\left\{r_\star, \max\left\{0, \left\lceil e\left(\frac{C_\lambda}{\varepsilon^2}\right)^{1/p}\right\rceil - 1\right\}\right\}. \tag{54}$$

If $\lambda_j \le C_\lambda \exp[-\beta(j-1)]$ for $C_\lambda, \beta > 0$, then $\mathcal{E}_m \le C_\lambda \exp(-\beta m/2)$ for $m < r_\star$, and it suffices to retain

$$m_{\exp} = \min\left\{r_\star, \max\left\{0, \left\lceil \frac{2}{\beta}\log\frac{C_\lambda}{\varepsilon^2}\right\rceil\right\}\right\} \tag{55}$$

tokens. The finite profiles in Proposition 5 can certify smaller sufficient budgets.

Proof of Corollary 6. Choose $n = m + 1$ and $k = 0$ in the profiles. In the polynomial case,

$$\mathcal{E}_m \le H_{\mathrm{pol}}(m) \le C_\lambda \left(\frac{N_m}{T}\right)^2 N_m^{-1/(m+1)} ((m+1)!)^{-p/(m+1)}. \tag{56}$$

For each integer $n \ge 1$, monotonicity of $\log x$ gives

$$\begin{aligned} \log(n!) &= \sum_{j=2}^{n} \log j \ge \sum_{j=2}^{n} \int_{j-1}^{j} \log x \, dx \\ &= \int_1^n \log x \, dx = n \log n - n + 1 \ge n \log n - n. \end{aligned} \tag{57}$$

The sums are empty when $n = 1$. Therefore

$$\begin{aligned} \mathcal{E}_m &\le C_\lambda \left(\frac{N_m}{T}\right)^2 N_m^{-1/(m+1)} ((m+1)!)^{-p/(m+1)} \\ &\le C_\lambda ((m+1)!)^{-p/(m+1)} \\ &\le C_\lambda \exp\left[-p \log(m+1) + p - \frac{p}{m+1}\right] \\ &\le \frac{C_\lambda e^p}{(m+1)^p}. \end{aligned} \tag{58}$$

Let $x = e(C_\lambda/\varepsilon^2)^{1/p}$ and $\widetilde{m} = \max\{0, \lceil x \rceil - 1\}$. If $x \le 1$, then $\widetilde{m} + 1 = 1 \ge x$; otherwise $\widetilde{m} + 1 = \lceil x \rceil \ge x$. Provided $\widetilde{m} < r_\star$, this proves

$$\mathcal{E}_{\widetilde{m}} \le \frac{C_\lambda e^p}{(\widetilde{m}+1)^p} \le \frac{C_\lambda e^p}{x^p} = \varepsilon^2. \tag{59}$$

If $\widetilde{m} \ge r_\star$, the rank budget gives zero error. This establishes m_{pol} in the corollary.

For exponential decay, the same $(n, k) = (m + 1, 0)$ choice gives

$$\begin{aligned} \mathcal{E}_m &\le H_{\exp}(m) \\ &\le C_\lambda \left(\frac{N_m}{T}\right)^2 N_m^{-1/(m+1)} \exp\left(-\frac{\beta m}{2}\right) \\ &\le C_\lambda \exp\left(-\frac{\beta m}{2}\right). \end{aligned} \tag{60}$$

Let $y = (2/\beta)\log(C_\lambda/\varepsilon^2)$ and $\widetilde{m} = \max\{0, \lceil y \rceil\}$. If $y \le 0$, then $C_\lambda \le \varepsilon^2$ and $\mathcal{E}_0 \le C_\lambda$ suffices. If $y > 0$ and $\widetilde{m} < r_\star$, then

$$\mathcal{E}_{\widetilde{m}} \le C_\lambda \exp(-\beta \widetilde{m}/2) \le C_\lambda \exp(-\beta y/2) = \varepsilon^2. \tag{61}$$

The rank cap handles the remaining case. If the implementation requires at least one retained token, increasing a zero certified budget to one preserves the guarantee by projection monotonicity; for a zero kernel every reconstruction is zero. □

The refined profiles sharpen the exponent of the closed-form corollary in the following regime.

Proposition 7 (A refined exponential consequence)**.** *Suppose $T \ge 2$, $m < r_\star$, and $\lambda_j \le C_\lambda \exp[-\beta(j-1)]$. Set $d_* = \sqrt{2(m+1)\log T/\beta}$. If $d_* \ge 1$ and $\lceil d_* \rceil \le m + 1$, then*

$$\mathcal{E}_m \le \frac{C_\lambda}{T} \exp\left[-\beta m + \sqrt{2\beta(m+1)\log T}\right]. \tag{62}$$

Proof. Set $n = m + 1$. For any integer $1 \le d \le n$, use $k = n - d$ in the exponential profile and $(T)_k \le T^k$. Since $N_m \ge 1$ and $N_m/T \le 1$,

$$\begin{aligned} \mathcal{E}_m &\le C_\lambda \left(\frac{N_m}{T}\right)^2 \left(\frac{(T)_{n-d}}{N_m}\right)^{1/d} \exp\left[-\frac{\beta(2n-d-1)}{2}\right] \\ &\le C_\lambda \exp\left[\frac{n-d}{d}\log T - \frac{\beta(2n-d-1)}{2}\right] \\ &= C_\lambda \exp\left[-\beta m - \frac{\beta}{2} - \log T + \frac{n \log T}{d} + \frac{\beta d}{2}\right]. \end{aligned} \tag{63}$$

Choose the admissible integer $d = \lceil d_* \rceil$. Since $d_* \le d < d_* + 1$, we have

$$\begin{aligned} \frac{n \log T}{d} + \frac{\beta d}{2} &\le \frac{n \log T}{d_*} + \frac{\beta(d_* + 1)}{2} \\ &= \sqrt{2\beta n \log T} + \frac{\beta}{2}. \end{aligned} \tag{64}$$

Substitution proves Equation (62). □

The scale C_λ throughout this subsection belongs to $K = F^\top F$. If a spectral envelope is specified for K/T with scale $C_{\lambda,0}$, then $\lambda_j(K) = T\lambda_j(K/T)$ and one must use $C_\lambda = TC_{\lambda,0}$. These corollaries give sufficient budgets with a scale that can vary across trajectories.

D.2.6 Sharpness and limits of spectral information

Proposition 8 (Unimprovable universal multiplicative constant)**.** *For every $T \ge 1$ and $0 \le m < T$, in a sufficiently large correction space there is a family with rank $m + 1$ such that $\mathcal{E}_m = (N_m/T)^2 \mathcal{Q}_m(K) > 0$. Consequently, no smaller universal multiplicative constant can replace one in Theorem 1.*

Proof. Choose orthonormal vectors $u_1, \dots, u_{m+1}$ and $a > 0$, and let

$$c_t = \begin{cases} au_t, & 1 \le t \le m, \\ au_{m+1}, & m < t \le T. \end{cases} \tag{65}$$

All unselected directions have squared residual a^2 during the first m iterations. The prescribed smallest-index tie rule therefore selects the m singleton columns. Writing $N = T - m$ gives

$$(I - P_m)\mu = \frac{Na}{T} u_{m+1}, \qquad \mathcal{E}_m = \frac{N^2 a^2}{T^2}, \qquad (\lambda_1, \dots, \lambda_T) = (Na^2, \underbrace{a^2, \dots, a^2}_{m \text{ entries}}, 0, \dots, 0). \tag{66}$$

For $1 \le n \le m + 1$, the leading product is Na^{2n}. When $m > 0$, $L_0 = a^2$, and for $1 \le i < m$,

$$0 < L_i = \frac{(m + 1 - i)a^2}{T - i} \le a^2. \tag{67}$$

Consequently, every candidate in $\mathcal{Q}_m(K)$ is at least a^2:

$$\left(\frac{\prod_{j=1}^n \lambda_j}{N \prod_{i=0}^{k-1} L_i} \right)^{1/(n-k)} \ge \left(\frac{Na^{2n}}{Na^{2k}} \right)^{1/(n-k)} = a^2. \tag{68}$$

Every $k = 0$ candidate equals a^2, proving $\mathcal{Q}_m(K) = a^2$ and the desired equality. For $m = 0$, there are no L_i and the same conclusion follows from $\lambda_1 = Ta^2$. The ambient dimension in this construction may depend on m. The equality family establishes sharpness of the universal multiplicative constant when dimensions and spectra are allowed to vary. □

Proposition 9 (Exact fixed-spectrum bound for two tokens)**.** *For $T = 2$, $m = 1$, and prescribed eigenvalues $\lambda_1 \ge \lambda_2 > 0$, the largest possible reconstruction error under the greedy selector is*

$$\sup_{\mathrm{spec}(K) = (\lambda_1, \lambda_2)} \mathcal{E}_1 = \frac{\lambda_1 \lambda_2}{2(\lambda_1 + \lambda_2)} = \frac{1}{4} \mathcal{Q}_1(K). \tag{69}$$

The determinant-only spectral certificate can exceed this value by an arbitrarily large factor.

Proof. Write $K = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$ and $d = \max\{a, c\}$. The first selected correction has squared norm d. The remaining residual has squared norm $\det(K)/d$, and it is the only residual contributing to the mean. Since $a + c = \lambda_1 + \lambda_2$, we obtain

$$\mathcal{E}_1 = \frac{\det K}{4d} = \frac{\lambda_1 \lambda_2}{4d} \le \frac{\lambda_1 \lambda_2}{2(a + c)} = \frac{\lambda_1 \lambda_2}{2(\lambda_1 + \lambda_2)}. \tag{70}$$

The positive definite matrix

$$K_* = \frac{1}{2}\begin{pmatrix} \lambda_1+\lambda_2 & \lambda_1-\lambda_2 \\ \lambda_1-\lambda_2 & \lambda_1+\lambda_2 \end{pmatrix} \tag{71}$$

has the prescribed eigenvalues and equal diagonal entries, so either allowed first selection attains equality. Any such positive semidefinite matrix has a Gram realization, for instance $F = K_*^{1/2}$. Here $N_1 = 1$ and $L_0 = (\lambda_1+\lambda_2)/2$, so

$$\begin{aligned} \mathcal{Q}_1(K) &= \min\left\{\lambda_1, \sqrt{\lambda_1\lambda_2}, \frac{2\lambda_1\lambda_2}{\lambda_1+\lambda_2}\right\} = \frac{2\lambda_1\lambda_2}{\lambda_1+\lambda_2}, \\ \mathcal{D}_1(K) &= \sqrt{\lambda_1\lambda_2}. \end{aligned} \tag{72}$$

The equality in the first line follows from $2\lambda_2 \le \lambda_1+\lambda_2$ and $2\sqrt{\lambda_1\lambda_2} \le \lambda_1+\lambda_2$. Thus the ratio of the two certificates is

$$\frac{\mathcal{D}_1(K)}{\mathcal{Q}_1(K)} = \frac{\lambda_1+\lambda_2}{2\sqrt{\lambda_1\lambda_2}} \longrightarrow \infty \quad \text{as } \lambda_1/\lambda_2 \longrightarrow \infty. \tag{73}$$

□

Proposition 10 (Cancellation and a spectral-tail limitation)**.** *Identical Gram spectra and identical greedy residual sequences need not imply identical mean-reconstruction errors. Moreover, no universal constant C makes $\mathcal{E}_m \le (C/T)\sum_{j>m}\lambda_j$ valid for every correction family under maximum-residual selection.*

Proof. For the first claim, use orthonormal u_1, u_2 and the two families

$$F_+ = [\sqrt{2}u_1, u_2, u_2], \qquad F_- = [\sqrt{2}u_1, u_2, -u_2]. \tag{74}$$

Both have spectrum $(2,2,0)$ and first select the unique largest-norm column $\sqrt{2}u_1$. Both residual maxima are $\delta_0 = 2$, $\delta_1 = 1$, and $\delta_2 = 0$. Nevertheless,

$$\mathcal{E}_1(F_+) = \left\|\frac{2u_2}{3}\right\|_2^2 = \frac{4}{9}, \qquad \mathcal{E}_1(F_-) = 0. \tag{75}$$

For the second claim, take $T = N+1$ with $N > 2$, let $c_1 = \sqrt{2}u_1$, and let all remaining N columns equal u_2. The first greedy choice is uniquely c_1, the nonzero eigenvalues are $(N, 2)$, and hence

$$\frac{\mathcal{E}_1}{T^{-1}\sum_{j>1}\lambda_j} = \frac{N^2/(N+1)^2}{2/(N+1)} = \frac{N^2}{2(N+1)} \ge \frac{N}{4} \longrightarrow \infty. \tag{76}$$

This rules out such a constant C. □

These constructions are compatible with the correction representation in Equation (2). To see this, let $V = 2$, choose $r = (1,-1)^\top/\sqrt{2}$, take a zero output head so that $q = (1/2, 1/2)^\top$, and choose $p = \mathrm{softmax}(-2r)$. With $c_0 = \log(e^{-2r_1} + e^{-2r_2}) - \log 2$, direct calculation yields

$$\log(q/p) = 2r + c_0\mathbf{1}_2, \qquad \ell = c_0, \qquad q \odot \big(\log(q/p) - \ell\mathbf{1}_2\big) = r. \tag{77}$$

Assigning the displayed example vectors as hidden states therefore gives $c_t = \mathrm{vec}(rh_t^\top)$ and $c_i^\top c_j = \|r\|_2^2 h_i^\top h_j = h_i^\top h_j$. The map is an isometry, so the examples preserve all Gram matrices and projection errors. This algebraic construction establishes realizability within the stated correction model; the distribution of such corrections in trained Transformers is an empirical question.

D.2.7 Stopping tolerances and inexact reconstruction

The spectral theorem assumes exact greedy selection. The following certificate allows any selected span, provided residual and solve errors are bounded in the true geometry. Applying it to the implementation in Appendix A.4 requires residual-accuracy and solve-error bounds in addition to the tolerance and backward-error check.

Proposition 11 (A conditional numerical certificate)**.** *Let B contain b independent corrections, with $0 \le b < T$, and let $F_B = QR$ be its exact thin QR factorization. Suppose approximate residual scores $\widetilde{v}_t$ and coefficients $\widetilde{w}$ satisfy*

$$\begin{aligned} \max_{t \notin B} \widetilde{v}_t \le \tau, \quad \max_{t \notin B} |\widetilde{v}_t - \|(I - QQ^\top)c_t\|_2^2| \le \epsilon_{\mathrm{res}}, \\ \|R\widetilde{w} - Q^\top \mu\|_2 \le \zeta, \qquad\qquad \tau, \epsilon_{\mathrm{res}}, \zeta \ge 0. \end{aligned} \tag{78}$$

Then

$$\|\mu - F_B \widetilde{w}\|_2^2 \le \left(\frac{T-b}{T}\right)^2 (\tau + \epsilon_{\mathrm{res}}) + \zeta^2. \tag{79}$$

For exact residuals and an exact solve, this reduces to $\|(I - QQ^\top)\mu\|_2^2 \le ((T-b)/T)^2 \tau$.

Proof. For every unselected position, the assumptions imply

$$\|(I - QQ^\top)c_t\|_2^2 \le \widetilde{v}_t + |\widetilde{v}_t - \|(I - QQ^\top)c_t\|_2^2| \le \tau + \epsilon_{\mathrm{res}}. \tag{80}$$

Selected positions have zero true residual. Applying the support argument in Equation (32), which does not require greedy selection, and then the orthogonal decomposition in Equation (31), gives

$$\begin{aligned} \|\mu - F_B \widetilde{w}\|_2^2 &= \|(I - QQ^\top)\mu\|_2^2 + \|R\widetilde{w} - Q^\top \mu\|_2^2 \\ &\le \frac{1}{T^2} \left(\sum_{t \notin B} \|(I - QQ^\top)c_t\|_2 \right)^2 + \zeta^2 \\ &\le \frac{1}{T^2} \left((T-b)\sqrt{\tau + \epsilon_{\mathrm{res}}} \right)^2 + \zeta^2 = \left(\frac{T-b}{T}\right)^2 (\tau + \epsilon_{\mathrm{res}}) + \zeta^2. \end{aligned} \tag{81}$$

The same argument covers $b = 0$ with empty matrices and zero coefficient-solve error. □