READOUT AND RELAY: TEMPORAL FUNCTIONAL GEOMETRY IN AUTOREGRESSIVE LANGUAGE MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Autoregressive transformers use each hidden state both to predict the next token and to support later predictions. We ask how these two roles coexist within a single representation. We identify a compact current-output-sensitive subspace, the READOUT, and decompose future credit into a component that reuses READOUT directions and a current-insensitive RELAY outside this subspace. Three patterns emerge. First, future credit substantially reuses the READOUT; within these reused directions, current and future credit can align or oppose, giving rise to synergy and conflict. Second, the RELAY is temporally selective: the component associated with a given future loss mainly affects the corresponding future prediction, with much smaller effects at other positions. For predictions farther in the future, a smaller fraction of future credit lies in the READOUT, while a larger fraction is carried by the RELAY. Third, across post-training stages, the balance between READOUT reuse and temporally selective RELAY changes little, while current-output sensitivity becomes concentrated in fewer directions. This suggests that post-training sharpens which directions matter for the current prediction without substantially reorganizing how hidden states support present and future predictions. Building on these findings, we introduce Temporal Proximal Credit (TPC), a lightweight training rule that reduces credit from later positions along directions important for the current prediction, without changing inference. TPC improves supervised fine-tuning across two model families, and a matched control suggests that both moderating future credit and directing that moderation toward current-sensitive directions contribute to the gains.

1 INTRODUCTION

A hidden state in an autoregressive transformer serves two roles at once. It must support next token prediction while also producing information that later positions reuse through causal attention. Prior work (Pal et al., 2023; Samragh et al., 2025) has shown that hidden states carry information about tokens several positions ahead, and Wu et al. (2024) asked whether such information reflects computation performed in advance or features that are already useful for the present prediction. More recently, State-Prediction Separation (SPS) (Monea et al., 2026) argues that using the same computation stream for prediction and state preparation can be suboptimal, and shows that separating the two roles architecturally can improve language modeling. These results leave open how present and future computation are organized inside the hidden states of a standard autoregressive transformer, where the two roles are not explicitly separated.

Answering this requires more than detecting future information in a hidden state. A direction that matters for a later prediction may also influence the current output, while another may remain invisible to it. Moreover, two directions in similar parts of the representation may act at very different points in the future. We therefore study the *functional geometry* of hidden states, defined by how directions affect present and future predictions rather than by what can be decoded from them.

Hypothesis: In a standard autoregressive transformer, present and future computation are neither cleanly separated nor uniformly shared. Instead, future computation is distributed across current output sensitive and current insensitive directions, which play distinct temporal roles.

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.

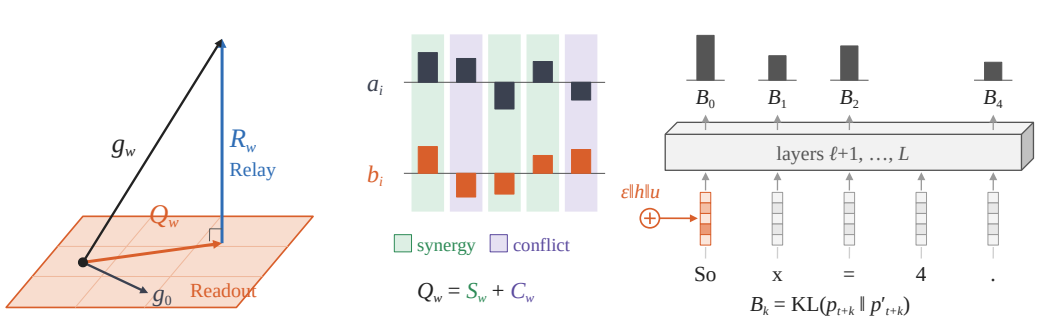

Figure 1: Temporal functional decomposition. **Left:** the READOUT is the span to which the current next token distribution is most sensitive. A future gradient g_w splits into shared credit Q_w inside it and RELAY credit R_w outside it. **Middle:** Q_w splits mode by mode into SYNERGY S_w and CONFLICT C_w according to the sign of $a_i b_i$, its interaction with current credit g_0. **Right:** the temporal bill perturbs $h_{\ell,t}$ along a component, reruns the blocks above layer ℓ on the fixed token trajectory, and records the KL at each later position.

We test this hypothesis with a token conditioned decomposition of the hidden state (Figure 1). The READOUT is the compact subspace to which the current next token distribution is sensitive; future credit splits into a shared part inside it and a RELAY outside it, which leaves the retained current logits unchanged to first order; the shared part splits into modes where current and future objectives agree (SYNERGY) or oppose (CONFLICT); and hidden state interventions on a fixed token trajectory measure when each part acts.

Through our analysis, we find that future computation substantially reuses directions involved in the current output, so present and future roles are not cleanly separated. What lies outside the READOUT is organized by horizon: each RELAY component acts mainly at the position whose loss produced it and far less at others, and longer range influence becomes increasingly RELAY dominated as sharing declines; donor controls show that both the sharing and this separation are position specific. Shared computation also has structured positive and negative interactions. Across post training stages, this organization changes little while current output sensitivity becomes more concentrated, suggesting that post training changes the functional *price* of existing computation more than its organization (Li et al., 2026; Shen et al., 2026; Qi et al., 2026).

This view suggests a training heuristic: if future credit is multiplexed with current prediction, treating all shared credit uniformly can disturb directions on which the current computation depends, while the coexistence of SYNERGY and CONFLICT gives no basis for removing credit by its sign. We therefore introduce Temporal Proximal Credit (TPC), which shrinks future credit according to current sensitivity during supervised fine tuning and leaves inference unchanged. Moderating this credit already helps, and placing it along current sensitive directions adds a small but consistent gain.

Our contributions are:

- We introduce a token-conditioned functional decomposition of hidden states that relates future credit to current-output geometry, separating READOUT reuse from a current-insensitive RELAY and characterizing SYNERGY and CONFLICT within the reused directions.
- We characterize how future credit is organized across time: it substantially reuses the READOUT, while the RELAY carries more credit for more distant predictions. This organization remains largely stable across post-training as current-output sensitivity becomes more concentrated.
- Temporal Proximal Credit, a lightweight sensitivity aware training rule motivated by this geometry that improves supervised fine tuning on two model families without changing inference.

2 PRESENT AND FUTURE CREDIT IN TRANSFORMERS

Let $h_{\ell,t} \in \mathbb{R}^d$ be the hidden state at layer ℓ and position t of an autoregressive transformer, and let L_{t+w} be the next token loss at position $t + w$. Through causal attention, $h_{\ell,t}$ receives credit from

the current loss ($w = 0$) and from every later loss ($w > 0$). We characterize the directions of $h_{\ell,t}$ by two functional questions: whether they change the current prediction (§2.1), and at which later position they act (§2.3). Figure 1 summarizes the construction.

2.1 THE READOUT: WHERE THE CURRENT OUTPUT IS SENSITIVE

Let $z_t \in \mathbb{R}^V$ be the logits at position t, S_t the top K tokens of the next token distribution, and $\tilde{p}_t$ the probabilities renormalized over S_t. We take the probability weighted Jacobian of the retained logits and its thin SVD,

$$B_t = \mathrm{Diag}\left(\sqrt{\tilde{p}_t}\right) \frac{\partial z_t[S_t]}{\partial h_{\ell,t}} \in \mathbb{R}^{K\times d}, \qquad B_t = V_t \Sigma_t U_t^\top, \qquad P_t = U_t U_t^\top, \tag{1}$$

where $U_t \in \mathbb{R}^{d\times r}$ holds the leading r right singular vectors, and call $\mathrm{span}(U_t)$ the READOUT. It contains the directions along which a perturbation of $h_{\ell,t}$ most changes the current prediction. $B_t^\top B_t$ pulls the output Fisher geometry back to $h_{\ell,t}$, as in FishBack (Wang et al., 2026), which uses this geometry for steering; we use it only as a measurement. Unlike FishBack, or an unembedding subspace (Hu et al., 2026) or an activation PCA basis, the READOUT is conditioned on the token and its context. We use $r = 64$ and the uncentered factor, which keeps the common mode logit direction needed to reconstruct the full output. Further details are in Appendix B.

2.2 DECOMPOSING FUTURE CREDIT

For a horizon $w > 0$, let $g_0 = \nabla_{h_{\ell,t}} L_t$ and $g_w = \nabla_{h_{\ell,t}} L_{t+w}$ be the current and future loss gradients at the same hidden state, with READOUT coordinates $a = U_t^\top g_0$ and $b = U_t^\top g_w$. We split the future gradient by whether it lies inside the READOUT, and split the shared part mode by mode by the sign of its interaction with current credit:

$$g_w = \underbrace{(I - P_t)\, g_w}_{R_w\,:\,\textsc{Relay}} + \underbrace{P_t\, g_w}_{Q_w\,:\,\text{shared}}, \tag{2}$$

$$Q_w = \underbrace{U_t\big(b \odot \mathbf{1}[a \odot b > 0]\big)}_{S_w\,:\,\textsc{Synergy}} + \underbrace{U_t\big(b \odot \mathbf{1}[a \odot b < 0]\big)}_{C_w\,:\,\textsc{Conflict}}. \tag{3}$$

Along mode i, $a_i b_i > 0$ means that descending the future loss also descends the current loss to first order, so SYNERGY and CONFLICT are local labels; they do not say whether a mode is useful. We summarize geometric sharing by $\Omega_w = \|Q_w\|_2 / \|g_w\|_2$, and Appendix D gives the energy of each component. The decomposition is silent about function. R_w is orthogonal to the retained READOUT by construction, so to first order it leaves the retained current logits unchanged; whether it is inert, broadly disruptive, or organized by horizon is an empirical question.

2.3 TEMPORAL BILLS: WHEN A DIRECTION ACTS

Geometry locates a direction but does not say when it acts. To measure timing, we perturb the hidden state along a unit direction u, rerun the transformer suffix with the original token trajectory held fixed, and record the change of each later next token distribution:

$$h'_{\ell,t} = h_{\ell,t} + \varepsilon \|h_{\ell,t}\|_2\, u, \qquad \mathcal{B}_k(u) = D_{\mathrm{KL}}\big(p_{t+k} \,\|\, p'_{t+k}\big), \qquad \mathcal{B}(u) = \big[\mathcal{B}_0, \mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_4\big](u). \tag{4}$$

We call $\mathcal{B}(u)$ the *temporal bill* of u and evaluate it for the normalized components $Q_w/\|Q_w\|_2$ and $R_w/\|R_w\|_2$ and for random controls, reporting it both as shares over offsets and in absolute KL. Holding the trajectory fixed removes autoregressive cascades, so the bill measures the direct latent effect of the perturbation. Further details are in Appendix E.

3 HOW PRESENT AND FUTURE COMPUTATION COEXIST

We study Qwen3-4B (Yang et al., 2025) and the OLMo-2 1B post training lineage (Base, SFT, DPO, Instruct; Team et al., 2025). Temporal analyses use correct chains of thought from VersaPRM (Zeng et al., 2025), whose reasoning traces cover many disciplines such as law and biology in addition to mathematics, so the measured geometry is not tied to a single domain. We analyze hidden states at a middle layer, where future influence is both large and separable from the current output (Appendix C); inputs, anchors, and numerics are described in Appendix A.

Table 1: Output interface validation on Qwen3-4B. Retained KL, top 1 agreement, and top 20 overlap reconstruct the output from the subspace; residual KL reconstructs it from the complement. HARP uses its native semantic rank.

Basis	Rank	Retained KL ↓	Top 1 ↑	Top 20 ↑	Residual KL ↑
Top K READOUT (ours)	64	**0.0095**	**0.993**	**0.970**	**15.906**
HARP (Hu et al., 2026)	2552	0.2301	0.653	0.758	8.537
Activation PCA	64	2.7875	0.315	0.352	2.825

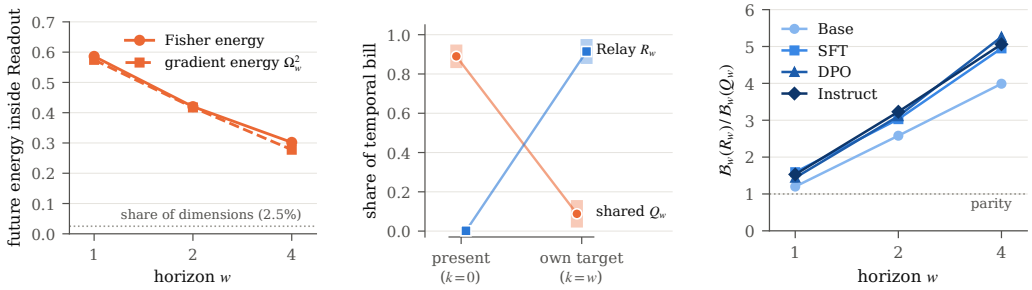

Figure 2: Where future credit lies does not determine when it acts. **Left:** fraction of future energy inside the READOUT. **Middle:** share of the temporal bill at the present and at the target horizon across OLMo-2 checkpoints and horizons. **Right:** ratio of RELAY to shared effect at the target horizon.

3.1 A SMALL READOUT FOR THE PRESENT

We first validate the READOUT at the output interface, where the map from hidden state to logits is exact. On Qwen3-4B, the rank 64 READOUT spans 2.5% of the 2560 dimensions yet reproduces the current distribution almost exactly, while its complement retains none of the original top 1 predictions (Table 1). A rank matched activation PCA basis barely separates retained from residual behavior, so the READOUT captures the local output function rather than high variance directions. A detailed rank analysis is in Appendix B.

3.2 SHARED CREDIT AND A HORIZON SEPARATED RELAY

Future credit is not confined to a separate current insensitive space. At every horizon, the READOUT holds 30 to 59% of the future Fisher energy while spanning only 2.5% of the dimensions, and this share decreases as the horizon grows (Figure 2, left). A donor control shows that the sharing is specific to each position rather than a property of the READOUT alone (Appendix F).

Sharing a subspace, however, does not mean acting at the same time (Figure 2, middle). Part of this follows from the construction. Nearly all of the effect of Q_w lands on the present prediction, as for any direction inside the READOUT, which spends 89 to 93% of its normalized bill at t even when random. R_w leaves the current logits unchanged to first order and is the steepest change of L_{t+w} within the complement, so it peaks at its own target in all 12 conditions (Appendix E). What the construction does not fix is where else R_w acts. On a representative checkpoint, R_4 spends 87% of its bill at $t+4$ and less than 7% at each of $t+1$ and $t+2$ (Table 7), and the RELAY bill centroids stay within 0.05 of 1.07, 1.96, and 3.24 across checkpoints (Appendix G), so different later positions read largely distinct directions of the complement. A donor future gradient shows no such horizon and peaks at $t+1$ for every w, while the shared component keeps a target effect 27 to 53 times that of its donor (Appendix F).

The division sharpens with temporal distance. At the target horizon, the absolute effects of the RELAY and the shared component are comparable at $w=1$, but the RELAY dominates by a factor of four to five at $w=4$, and the ratio rises monotonically with w in every checkpoint (Figure 2, right). To first order, the change of L_{t+w} along a unit direction is $\|R_w\|_2$ for the RELAY and $\|Q_w\|_2$ for the shared component, so this trend is the interventional counterpart of the falling Ω_w. Sharing sets how much of the target effect travels through the RELAY, and the RELAY keeps different horizons apart.

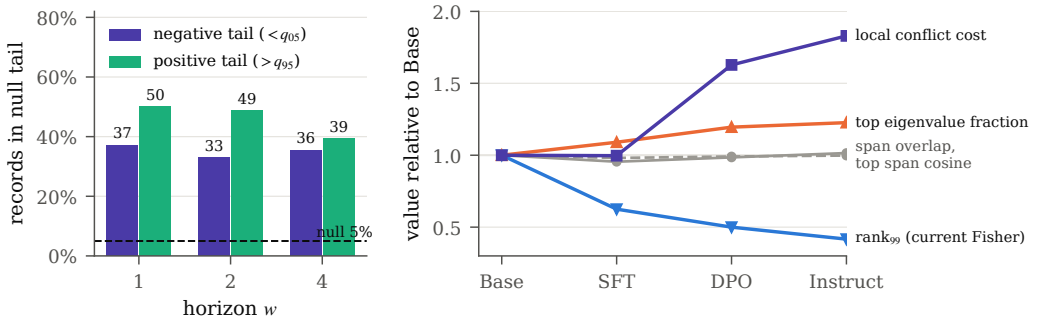

Figure 3: Coupling runs in both directions, and post training re-prices rather than re-wires. **Left:** share of true pairs in each tail of the permutation null of $\cos(g_0, g_w)$ on Qwen3-4B; the dashed line is the 5% null expectation. **Right:** OLMo-2 metrics relative to Base. Organization (gray) stays flat while current sensitivity concentrates and conflict grows costlier.

3.3 Synergy and Conflict

Inside the Readout, shared credit can pull with or against current credit. Because mixed signs arise by chance in any anisotropic high dimensional space, we first ask whether true current and future pairs interact more strongly than unrelated ones. We compare $\chi_w = \cos(g_0, g_w)$ with a horizon matched permutation null that pairs the same g_0 with the future gradient of a different anchor; the null keeps the marginal geometry of both gradients, breaks their temporal pairing, and by construction places 5% of records in each tail. True pairs fall in both tails six to ten times more often (Figure 3, left): about a third of records lie in the negative tail and 40 to 50% in the positive tail at every horizon. Current and future gradients are thus strongly coupled, in both directions rather than toward agreement or opposition. Within the shared component, the energy split between Synergy and Conflict is close to balanced and matches a donor control (Appendix D), so the coupling lies in magnitude rather than sign, which is why the rule of Section 4 does not act on sign.

3.4 Post training re-prices, not re-wires

Finally, we follow the OLMo-2 lineage from Base through SFT, DPO, and Instruct at a fixed middle layer with identical inputs. The organization of the geometry barely moves. Span overlap and top span cosine stay nearly constant, and the temporal bill centroids shift by less than 0.1 across stages (Figure 3 right, gray). What changes is how concentrated, and how costly, current sensitivity becomes. At a typical position, the number of eigendirections needed to capture 99% of the current output's Fisher energy falls from 24 to 10, and the share held by the largest eigendirection rises from 0.54 to 0.67. Part of this concentration is expected, since post training lowers the entropy of the next token distribution (Kirk et al., 2024; Achiam et al., 2023) and thereby concentrates the probability weighting in Eq. 1; all checkpoints are compared on a shared anchor population (Appendix A). On records where current and future credit conflict, a small step that lowers the future loss raises the current loss almost twice as much per unit of future gain, from 0.32 to 0.58. Post training thus changes the functional *price* of existing directions more than their arrangement, and the temporal organization of §3.2 is already present in the base model, consistent with evidence that fine tuning largely reuses and adjusts existing mechanisms rather than building new ones (Jain et al., 2024; Prakash et al., 2024). Full values are in Appendix G.

4 Temporal Proximal Credit

The analysis motivates a heuristic for assigning credit to a middle layer hidden state, and TPC implements an inexpensive surrogate of it. Credit that later positions send back splits into a shared part inside the Readout, which acts mostly on the present prediction, and a Relay outside it, which carries the effect to its target horizon (§3.2); this suggests charging the shared part by how strongly the current computation relies on each direction while letting the Relay pass. Because the shared part agrees and conflicts with current credit in comparable measure (§3.3), we do not let the charge depend on sign, and because post training changes the price of directions rather than the directions themselves (§3.4), we let it act on magnitude and leave the forward computation intact.

Computing the decomposition exactly during training would take a backward pass per horizon and a Jacobian per token, so TPC makes two approximations: it separates future credit at its first read site in one backward pass (§4.1) and replaces the per-token READOUT with a fixed data-averaged basis (§4.2). It then shrinks the separated credit with a closed form proximal step (Parikh & Boyd, 2014) (Figure 4).

4.1 SEPARATING FUTURE PATH CREDIT IN ONE BACKWARD PASS

Later positions read a hidden state only through the keys and values of the attention blocks above it (Wu et al., 2024). At the normalized input x_t of the attention block directly above the analyzed layer, we feed the key and value projections a second copy of x_t that is identical in the forward pass but forms its own branch of the autograd graph. One ordinary backward pass then returns the gradient of each branch separately,

$$a_t = \nabla^{\mathrm{cur}}_{x_t} L, \qquad b_t = \nabla^{\mathrm{fut}}_{x_t} L, \qquad g_t^{\mathrm{SFT}} = a_t + b_t. \tag{5}$$

We call b_t the *future path credit*, the credit that later positions send back through the keys and values, and a_t the *current path credit*, which arrives through the rest of the block. Adding the two gives back the ordinary SFT gradient, so the split alters training only when a rule acts on it. Because the split of Eq. 2 is linear in the future gradient, projecting b_t, which sums the key and value credit of all later losses, yields the sum of their shared components, and its complement the sum of their RELAY components, so a single projection applies the decomposition to every horizon at once (Table 12). Credit that reaches $h_{\ell,t}$ through the keys and values of higher blocks returns along the residual stream and is left untouched, so TPC acts on the first hop of future credit. The key and value path also carries the position's attention to itself, but this self credit is small. The current path is less pure, as most of a_t comes from later losses through the attention output of position t; we use it only to locate directions (§4.2; Appendix H.3).

4.2 A SENSITIVITY METRIC FROM CURRENT PATH CREDIT

The READOUT of Section 2.1 identifies, for a single token, the directions to which the current prediction is sensitive. During training we use a data averaged counterpart computed at the site of the rule, the uncentered second moment of current path credit,

$$M = \mathbb{E}\big[a_t a_t^\top\big] \approx U\,\mathrm{Diag}(\mu_1, \ldots, \mu_r)\,U^\top, \qquad \hat{\mu}_i = \frac{\mu_i}{\frac{1}{r}\sum_{j=1}^{r}\mu_j}. \tag{6}$$

Whatever loss it originates from, a_t enters through the query of position t, so M is a Gauss Newton type metric over the directions in which x_t steers what position t attends to, weighted by how much the loss depends on that choice. What position t reads feeds its own prediction first, and 82 to 90% of the energy of a_t lies in the site READOUT (Appendix H.3), so we treat the leading directions of M as the data averaged counterpart of the READOUT at the rule site and $\hat{\mu}_i$ as how strongly the computation at position t relies on each. We estimate M once before fine tuning, with $r = 64$, the rank at which the READOUT already preserves the current output, and keep it fixed (Appendix H.2). On held out chains, this basis captures 16 to 23 times more of the current output sensitivity at the rule site than a random basis of the same rank, and 6 to 9 times more than activation PCA (Appendix H.3).

4.3 PROXIMAL PRICING OF FUTURE PATH CREDIT

Given future path credit b_t, TPC returns the credit closest to b_t once movement along sensitive directions carries a price,

$$d_t^\star = \arg\min_d \Big[\tfrac{1}{2}\|d - b_t\|_2^2 + \tfrac{\lambda}{2}\, d^\top U\,\mathrm{Diag}(\hat{\mu})\,U^\top d\Big] \tag{7}$$

$$= (I - UU^\top)\,b_t + U\,\mathrm{Diag}(s)\,U^\top b_t, \qquad s_i = \frac{1}{1+\lambda\hat{\mu}_i}. \tag{8}$$

Credit outside the metric passes unchanged, and each retained mode shrinks according to how strongly the current computation relies on it. Because $s_i \in (0, 1]$, every retained mode keeps its sign, $u_i^\top d_t^\star = s_i\, u_i^\top b_t$, and so does its interaction with current path credit, also after the trust region of Section 4.4 (Appendix I). Unlike gradient surgery (Yu et al., 2020; Liu et al., 2021), which projects conflicting components away, TPC never removes, adds, or reverses credit; it only limits how far future path credit moves the hidden state along directions the current computation relies on.

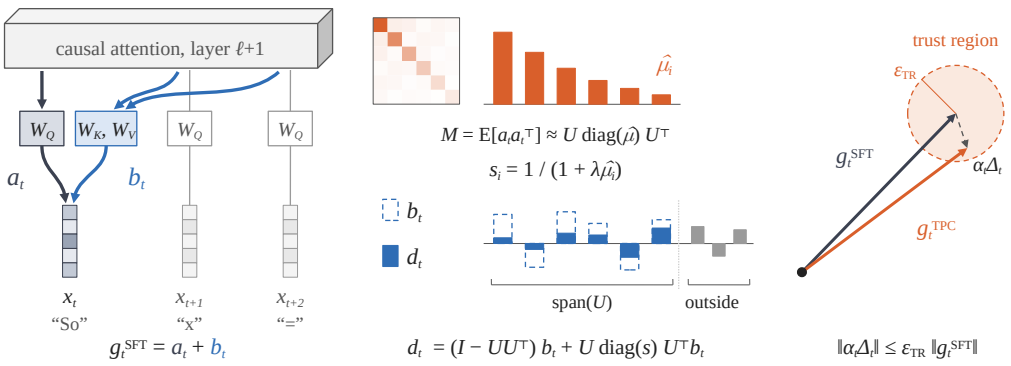

Figure 4: TPC re-prices the credit that later positions send back to a hidden state. **Left:** one ordinary backward pass separates current path credit a_t from future path credit b_t, which arrives through the key and value projections. **Middle:** a fixed low rank metric of current path credit shrinks each mode of b_t by s_i, most where $\hat{\mu}_i$ is large, while keeping its sign; credit outside the metric passes unchanged. **Right:** the update stays within a trust region around the SFT gradient.

4.4 Trust region and a matched control

Attenuating one path can reduce its cancellation with the other, so we bound the change relative to SFT for every token,

$$g_t^{\mathrm{TPC}} = g_t^{\mathrm{SFT}} + \alpha_t \Delta_t, \qquad \Delta_t = d_t^\star - b_t, \qquad \alpha_t = \min\left(1, \frac{\epsilon_{\mathrm{TR}}\,\|g_t^{\mathrm{SFT}}\|_2}{\|\Delta_t\|_2}\right). \tag{9}$$

To separate where credit is changed from the fact that it is changed at all, the Uniform control applies a correction of the same norm to every token but spreads it along b_t itself, $b_t^{\mathrm{uni}} = (1 - q_t)\,b_t$ with $q_t = \|\Delta_t\|_2 / \|b_t\|_2$, under the same trust region. The two rules change future path credit by the same amount and differ only in direction.

5 Experiments

5.1 Experimental Setup

Models We fine tune Qwen3-4B-Base, a pretrained model without post training, and Llama-3.2-3B-Instruct (Grattafiori et al., 2024), an instruction tuned model, to test TPC across model families and training stages.

Data Training uses 20k mathematical reasoning traces from Mixture of Thoughts (Fein-Ashley et al., 2025) whose responses fit within the training context. We evaluate in domain on MATH500 (Lightman et al., 2024), Minerva (Lewkowycz et al., 2022), AIME 2025 and 2026 (Balunovic et al., 2026), and BeyondAIME (Seed et al., 2025), and beyond the training domain on MMLU-Pro (Wang et al., 2024), BBH (Suzgun et al., 2023), MuSR (Sprague et al., 2024), and ZebraLogic (Lin et al., 2025).

Baselines We compare TPC with Vanilla SFT and with Dynamic Fine Tuning (DFT) (Wu et al., 2026), which rescales the loss of each token by its predicted probability. Like TPC, DFT changes only how supervised fine tuning assigns credit, but it reweights whole tokens, whereas TPC reshapes credit along directions of the hidden state. We also report the Uniform control of Section 4.4. All methods share the data, the data order, and every optimization setting; Appendices J and K give the full training and evaluation details.

5.2 Main results

In domain Averaged over the three competition sets (Table 2), TPC has the best Avg@k on both models and the best Pass@k on Llama-3.2-3B-Instruct; these sets are small, individual entries are noisy, and Llama-3.2-3B-Instruct stays near the floor on all three, so we treat them as supporting

Table 2: Sampled accuracy (%) on competition math (P: Pass@k, A: Avg@k).

	AIME25				AIME26				BeyondAIME			
Method	P@8	A@8	P@16	A@16	P@8	A@8	P@16	A@16	P@8	A@8	P@16	A@16
Qwen3-4B-Base												
Base	16.67	6.25	23.33	7.08	16.67	7.08	20.00	6.67	14.00	4.50	16.00	4.06
Vanilla SFT	26.67	18.33	33.33	**18.33**	**36.67**	12.92	**40.00**	13.33	15.00	5.00	23.00	5.88
DFT	26.67	11.67	30.00	12.29	20.00	9.17	23.33	9.38	13.00	6.00	18.00	**6.19**
Uniform	26.67	17.08	33.33	17.92	20.00	12.08	26.67	12.71	**20.00**	**6.25**	**27.00**	6.06
TPC	**33.33**	**19.17**	**36.67**	17.92	26.67	**14.58**	33.33	**14.79**	17.00	6.12	23.00	5.88
Llama-3.2-3B-Instruct												
Base	3.33	0.42	6.67	0.83	6.67	0.83	6.67	0.62	**6.00**	**1.00**	10.00	0.88
Vanilla SFT	3.33	0.42	**23.33**	1.46	6.67	0.83	10.00	0.62	4.00	0.50	5.00	0.44
DFT	3.33	0.42	6.67	0.62	**13.33**	**2.50**	**16.67**	**1.88**	5.00	**1.00**	6.00	0.69
Uniform	10.00	2.50	16.67	**2.08**	3.33	0.42	3.33	0.62	5.00	**1.00**	**11.00**	**1.12**
TPC	**16.67**	**2.92**	20.00	**2.08**	10.00	1.67	**16.67**	1.25	**6.00**	**1.00**	8.00	0.81

Table 3: Greedy accuracy (%) on in domain and out of domain benchmarks.

	In Domain		Out of Domain				
Method	MATH500	Minerva	MMLU-Pro	BBH Ded.	MuSR	ZebraLogic	Avg.
Qwen3-4B-Base							
Base	71.60	40.81	54.43	42.80	52.65	37.50	49.97
Vanilla SFT	**82.40**	50.37	58.37	83.60	53.17	41.20	61.51
DFT	78.60	54.04	57.71	77.87	51.06	40.00	59.88
Uniform	82.00	52.94	**59.14**	**84.27**	53.44	41.40	62.20
TPC	82.20	**55.14**	58.71	**84.27**	**53.57**	**42.30**	**62.70**
Llama-3.2-3B-Instruct							
Base	50.20	24.26	45.00	53.60	47.62	**34.00**	42.45
Vanilla SFT	53.80	22.79	42.57	61.47	47.35	25.00	42.16
DFT	52.40	26.47	41.57	59.07	48.41	25.00	42.15
Uniform	54.40	**29.78**	43.71	61.47	48.81	25.50	43.95
TPC	**55.20**	28.68	**45.43**	**61.60**	**52.51**	26.00	**44.90**

evidence only. TPC improves Minerva Math by 4.8 and 5.9 points over Vanilla SFT and is within 0.2 points of the best MATH500 score (Table 3).

Out of domain On Qwen3-4B-Base, TPC improves on Vanilla SFT on all four benchmarks. On Llama-3.2-3B-Instruct, Vanilla SFT and DFT fall below the base model, mainly on MMLU-Pro and ZebraLogic, whereas TPC raises the average from 42.5 to 44.9. TPC has the best greedy average on both models.

5.3 Does the direction of the change matter?

A gain could come from reducing future path credit in any direction, which Uniform (§4.4) tests by changing the same amount along b_t. Uniform already improves on Vanilla SFT, from 61.51 to 62.20 on Qwen3-4B-Base and from 42.16 to 43.95 on Llama-3.2-3B-Instruct, suggesting that moderating future path credit helps by itself. TPC adds a further gain on both models (62.70 and 44.90) and exceeds Uniform on 9 of the 12 greedy benchmark and model pairs, tying on one. At the rule site only 9 to 11% of the energy of b_t lies in the site READOUT, so Uniform takes about 90% of its correction from outside the READOUT, the part that by Section 3.2 carries the effect at the target, whereas the TPC correction concentrates in the READOUT four to seven times more (Appendix H.3). The correction is small in norm, changing a median 5 to 7% of b_t at a single site (Appendix H.2), but per unit norm it moves the current prediction 8.7 and 17.8 times more than a change along b_t. Moderation thus appears to account for most of the gain over SFT; placing it along current sensitive directions adds a further 0.5 and 0.95 points on average, a single-seed margin we read through its consistency across models and benchmarks, not any single score.

5.4 Does the metric remain valid after fine tuning?

We repeat the rule site measurement of Section 4.2 on the fine tuned models with the basis estimated before training. After Vanilla SFT, Uniform, or TPC, it still captures 37% of the site READOUT on Qwen3-4B-Base and 48 to 49% on Llama-3.2-3B-Instruct (against 40% and 50% before training), and the TPC correction stays 6 to 16 times more output sensitive than a change along b_t (Appendix H.3). The geometry thus survives fine tuning, which justifies estimating it once, and since the small shifts also appear under Vanilla SFT, we attribute the gain of TPC to how it assigns credit rather than to a change in this geometry.

6 Related Work

Present and future roles of a hidden state Hidden states encode tokens several positions ahead (Pal et al., 2023) and longer range properties such as a rhyme, the next paragraph, or response length (Lindsey et al., 2025; Pochinkov et al., 2025; Dong et al., 2025). Whether this information is used rather than merely decodable is less settled: latent planning emerges with scale (Hanna & Ameisen, 2026), varies in location across models (Maar et al., 2026), and need not be used where encoded (Ma & Rui, 2026). With myopic training, which blocks gradients from later losses to earlier positions, Wu et al. (2024) find that *breadcrumbs* useful now and later dominate *pre caching* for later use in language modeling, though pre caching grows with scale. SPS (Monea et al., 2026) argues that one stream serving both roles causes tension and adds a prediction stream, and multi token and future summary objectives (Gloeckle et al., 2024; Mahajan et al., 2026) train the future role explicitly. In a standard transformer, where nothing separates the two roles, we find substantial sharing with the current output alongside a RELAY that keeps horizons apart.

Output geometry of intermediate representations FishBack (Wang et al., 2026) pulls the softmax Fisher metric back to intermediate activations as an input dependent metric for steering, and finds its effective dimension to be a small fraction of the hidden space, in line with the compact READOUT we observe. The Jacobian lens (Gurnee et al., 2026) linearizes the map from activations to outputs to read and write the concepts an activation carries, and HARP (Hu et al., 2026) builds static semantic and reasoning subspaces from the unembedding matrix. We use related machinery as a measurement instrument rather than as a contribution: our READOUT is a truncated, uncentered factor over the top K next tokens that reconstructs the current distribution from a rank 64 span, paired with interventions that ask when, rather than whether, a direction changes the output.

Credit assignment in fine tuning PCGrad (Yu et al., 2020) projects away conflicting task gradients, and gradient similarity can decide when an auxiliary loss helps (Du et al., 2018). DFT (Wu et al., 2026) reweights token losses by predicted probability, PSFT (Zhu et al., 2026b) and ASFT (Zhu et al., 2026a) bound token probabilities or the output distribution, and FisherSFT (Deb et al., 2025) selects examples by Fisher information. TPC instead shrinks the credit later positions send to one hidden state along current sensitive directions, bounds this gradient rather than the output, and keeps conflicting credit with its sign, making it a soft, selective relative of myopic training.

7 Conclusion

We asked how present and future computation coexist in the hidden states of a standard autoregressive transformer. Future credit largely reuses the directions the current output depends on, while what lies outside them separates later horizons and carries a growing share of the target effect as the horizon grows, and post training keeps this organization while changing mainly how costly the current sensitive directions become. Building on this geometry, TPC moderates the credit later positions send back along current sensitive directions and improves supervised fine tuning, with a matched control suggesting that both the moderation and its direction contribute. Our study also has limits: temporal bills measure direct latent effects with the token trajectory fixed, and our fine tuning results come from a single seed on models of at most 4B parameters. Whether the organization and the gains hold across seeds, at larger scale, or with the rule at several layers remains open.

AI USE STATEMENT

We used generative AI tools (GPT and Claude) for three purposes. First, we implemented the code for our methods and analyses together with these tools, and the first author tested all of this code. Second, we used them to search for related work, and the first author checked every cited work against its original publication. Third, we used them to improve the fluency of the English writing. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Section 2 and Appendices A, B, E, and F describe the models, data, anchor selection, numerics, and intervention protocol of the analysis. Section 4 and Appendix H specify TPC, its implementation, and its pseudocode (Algorithm 1), and Appendix I derives its closed form and shows that it preserves the sign of every retained mode. Appendices J and K list the training data, every hyperparameter, the evaluation benchmarks, and the decoding settings. All models and datasets used in this work are publicly available.

REFERENCES

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Mislav Balunovic, Jasper Dekoninck, Ivo Petrov, Nikola Jovanović, and Martin Vechev. Matharena: Evaluating llms on uncontaminated math competitions. *Advances in Neural Information Processing Systems*, 38, 2026.

Rohan Deb, Kiran Thekumparampil, Kousha Kalantari, Gaurush Hiranandani, Shoham Sabach, and Branislav Kveton. Fishersft: Data-efficient supervised fine-tuning of language models using information gain. *arXiv preprint arXiv:2505.14826*, 2025.

Zhichen Dong, Zhanhui Zhou, Zhixuan Liu, Chao Yang, and Chaochao Lu. Emergent response planning in llms. *arXiv preprint arXiv:2502.06258*, 2025.

Yunshu Du, Wojciech M Czarnecki, Siddhant M Jayakumar, Mehrdad Farajtabar, Razvan Pascanu, and Balaji Lakshminarayanan. Adapting auxiliary losses using gradient similarity. *arXiv preprint arXiv:1812.02224*, 2018.

Jacob Fein-Ashley, Dhruv Parikh, Rajgopal Kannan, and Viktor Prasanna. Mixture of thoughts: Learning to aggregate what experts think, not just what they say. *arXiv preprint arXiv:2509.21164*, 2025.

Fabian Gloeckle, Badr Youbi Idrissi, Baptiste Rozière, David Lopez-Paz, and Gabriel Synnaeve. Better & faster large language models via multi-token prediction. *arXiv preprint arXiv:2404.19737*, 2024.

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

Wes Gurnee, Nicholas Sofroniew, Adam Pearce, Mateusz Piotrowski, Isaac Kauvar, Runjin Chen, Anna Soligo, Paul Bogdan, Euan Ong, Rowan Wang, et al. Verbalizable representations form a global workspace in language models. *arXiv preprint arXiv:2607.15495*, 2026.

Michael Hanna and Emmanuel Ameisen. Latent planning emerges with scale. *arXiv preprint arXiv:2604.12493*, 2026.

Junjie Hu, Gang Tu, Cheng Shengyu, Jinxin Li, Jinting Wang, Rui Chen, Zhilong Zhou, and Dongbo Shan. Harp: Hallucination detection via reasoning subspace projection. In *International Conference on Learning Representations*, volume 2026, pp. 17783–17802, 2026.

Samyak Jain, Robert Kirk, Ekdeep Singh Lubana, Robert Dick, Hidenori Tanaka, Tim Rocktäschel, Edward Grefenstette, and David Krueger. Mechanistically analyzing the effects of fine-tuning on procedurally defined tasks. In *International Conference on Learning Representations*, volume 2024, pp. 9944–10015, 2024.

Robert Kirk, Ishita Mediratta, Christoforos Nalmpantis, Jelena Luketina, Eric Hambro, Edward Grefenstette, and Roberta Raileanu. Understanding the effects of rlhf on llm generalisation and diversity. In *International Conference on Learning Representations*, volume 2024, pp. 20620–20653, 2024.

Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, Ambrose Slone, Cem Anil, Imanol Schlag, Theo Gutman-Solo, et al. Solving quantitative reasoning problems with language models. *Advances in neural information processing systems*, 35:3843–3857, 2022.

Melody Li, Kumar Krishna Agrawal, Arna Ghosh, Komal Teru, Adam Santoro, Guillaume Lajoie, and Blake Richards. Tracing the representation geometry of language models from pretraining to post-training. *Advances in Neural Information Processing Systems*, 38:54691–54724, 2026.

Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. In *International Conference on Learning Representations*, volume 2024, pp. 39578–39601, 2024.

Bill Yuchen Lin, Ronan Le Bras, Kyle Richardson, Ashish Sabharwal, Radha Poovendran, Peter Clark, and Yejin Choi. Zebralogic: On the scaling limits of llms for logical reasoning. *arXiv preprint arXiv:2502.01100*, 2025.

Jack Lindsey, Wes Gurnee, Emmanuel Ameisen, Brian Chen, Adam Pearce, Nicholas L. Turner, Craig Citro, David Abrahams, Shan Carter, Basil Hosmer, Jonathan Marcus, Michael Sklar, Adly Templeton, Trenton Bricken, Callum McDougall, Hoagy Cunningham, Thomas Henighan, Adam Jermyn, Andy Jones, Andrew Persic, Zhenyi Qi, T. Ben Thompson, Sam Zimmerman, Kelley Rivoire, Thomas Conerly, Chris Olah, and Joshua Batson. On the biology of a large language model. *Transformer Circuits Thread*, 2025. URL `https://transformer-circuits.pub/2025/attribution-graphs/biology.html`.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. *Advances in neural information processing systems*, 34:18878–18890, 2021.

Nicole Ma and Nick Rui. Where's the plan? locating latent planning in language models with lightweight mechanistic interventions. *arXiv preprint arXiv:2605.07984*, 2026.

Jim Maar, Denis Paperno, Callum McDougall, and Neel Nanda. What's the plan? metrics for implicit planning in llms and their application to rhyme generation and question answering. In *International Conference on Learning Representations*, volume 2026, pp. 96387–96427, 2026.

Divyat Mahajan, Sachin Goyal, Badr Youbi Idrissi, Mohammad Pezeshki, Ioannis Mitliagkas, David Lopez-Paz, and Kartik Ahuja. Beyond multi-token prediction: Pretraining llms with future summaries. In *International Conference on Learning Representations*, volume 2026, pp. 64342–64364, 2026.

Giovanni Monea, Nathan Godey, Kianté Brantley, and Yoav Artzi. The state-prediction separation hypothesis. *arXiv preprint arXiv:2607.01218*, 2026.

Koyena Pal, Jiuding Sun, Andrew Yuan, Byron C Wallace, and David Bau. Future lens: Anticipating subsequent tokens from a single hidden state. In *Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL)*, pp. 548–560, 2023.

Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3): 127–239, 2014.

Nicky Pochinkov, Yulia Volkova, Anna Vasileva, and Sai VR Chereddy. Parascopes: What do language models activations encode about future text? *arXiv preprint arXiv:2511.00180*, 2025.

Nikhil Prakash, Tamar Shaham, Tal Haklay, Yonatan Belinkov, and David Bau. Fine-tuning enhances existing mechanisms: A case study on entity tracking. In *International Conference on Learning Representations*, volume 2024, pp. 8057–8082, 2024.

Jianing Qi, Hao Tang, and Zhigang Zhu. GRRR: The geometry of reshaping, rotation, and routing in decoder LLM post-training. *arXiv preprint arXiv:2609.22146*, 2026. doi: 10.48550/arXiv.2609.22146. URL https://arxiv.org/abs/2609.22146.

Mohammad Samragh, Arnav Kundu, David Harrison, Kumari Nishu, Devang Naik, Minsik Cho, and Mehrdad Farajtabar. Your llm knows the future: Uncovering its multi-token prediction potential. *arXiv preprint arXiv:2507.11851*, 2025.

ByteDance Seed, Yufeng Yuan, Yu Yue, Mingxuan Wang, Xiaochen Zuo, Jiaze Chen, Lin Yan, Wenyuan Xu, Chi Zhang, Xin Liu, et al. Seed-thinking-v1. 5: Advancing superb reasoning models with reinforcement learning. *arXiv preprint arXiv:2504.13914*, 4, 2025.

Jingyan Shen, Ang Li, Salman Rahman, Yifan Sun, Micah Goldblum, Matus Telgarsky, and Pavel Izmailov. Understanding reasoning from pretraining to post-training. *arXiv preprint arXiv:2607.16097*, 2026.

Zayne Sprague, Xi Ye, Kaj Bostrom, Swarat Chaudhuri, and Greg Durrett. Musr: Testing the limits of chain-of-thought with multistep soft reasoning. In *International Conference on Learning Representations*, volume 2024, pp. 14670–14728, 2024.

Mirac Suzgun, Nathan Scales, Nathanael Schärli, Sebastian Gehrmann, Yi Tay, Hyung Won Chung, Aakanksha Chowdhery, Quoc Le, Ed H Chi, Denny Zhou, et al. Challenging big-bench tasks and whether chain-of-thought can solve them. In *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 13003–13051, 2023.

OLMo Team, Pete Walsh, Luca Soldaini, Dirk Groeneveld, Kyle Lo, Shane Arora, Akshita Bhagia, Yuling Gu, Shengyi Huang, Matt Jordan, et al. 2 olmo 2 furious. *arXiv preprint arXiv:2501.00656*, 2025.

Sihan Wang, Jiayi Zhao, Qingyan Cao, Hongbo Yao, and Lin Shu. Fishback: Pullback fisher geometry for optimal activation steering in transformers. *arXiv preprint arXiv:2605.17231*, 2026.

Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran Arulraj, Xuan He, Ziyan Jiang, et al. Mmlu-pro: A more robust and challenging multi-task language understanding benchmark. *Advances in Neural Information Processing Systems*, 37:95266–95290, 2024.

Wilson Wu, John X Morris, and Lionel Levine. Do language models plan ahead for future tokens? *arXiv preprint arXiv:2404.00859*, 2024.

Yongliang Wu, Yizhou Zhou, Ziheng Zhou, Yingzhe Peng, Xinyu Ye, Xinting Hu, Wenbo Zhu, Lu Qi, Ming-Hsuan Yang, et al. On the generalization of sft: A reinforcement learning perspective with reward rectification. In *International Conference on Learning Representations*, volume 2026, pp. 27550–27571, 2026.

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.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. *Advances in neural information processing systems*, 33: 5824–5836, 2020.

Thomas Zeng, Shuibai Zhang, Shutong Wu, Christian Classen, Daewon Chae, Ethan Ewer, Minjae Lee, Heeju Kim, Wonjun Kang, Jackson Kunde, et al. Versaprm: Multi-domain process reward model via synthetic reasoning data. *arXiv preprint arXiv:2502.06737*, 2025.

He Zhu, Junyou Su, Peng Lai, Ren Ma, Wenjia Zhang, Linyi Yang, and Guanhua Chen. Anchored supervised fine-tuning. In *International Conference on Learning Representations*, volume 2026, pp. 110364–110385, 2026a.

Wenhong Zhu, Ruobing Xie, Rui Wang, Samm Sun, Di Wang, and Pengfei Liu. Proximal supervised fine-tuning. In *International Conference on Learning Representations*, volume 2026, pp. 134993–135015, 2026b.

A Analysis Setup

A.1 Models and data

The analysis in Sections 2 and 3 uses Qwen3-4B for output interface validation and for the depth and horizon studies, and the OLMo-2 1B lineage (Base, SFT, DPO, and Instruct) for controlled comparisons across post training stages. Temporal analyses use correct chains of thought from VersaPRM. When checkpoints are compared, every checkpoint receives the same plain format input so that the input strings are byte identical; OLMo-2 Base has no chat template, and this is the only format shared by all four stages. The analyzed layer is the middle block of each model (Appendix C).

A.2 Anchor selection

Candidate positions are sampled by stride with a fixed maximum number per sequence. Positions whose next token distribution has very low entropy are excluded at selection time from a candidate population shared by all checkpoints, rather than at extraction time. The fraction of such positions ranges from under 1% in OLMo-2 Base to about 16% in Instruct, so filtering each checkpoint separately would compare different populations of positions.

A.3 Numerics

Jacobians, basis construction, and all interventions use float32. The smallest effects of random control directions lie near the numerical floor and can be hidden by bfloat16 noise.

B Readout Construction and Rank Analysis

Construction For each anchor we take the top K support S_t of the next token distribution, form B_t in Eq. 1 in float32, and keep the leading r right singular vectors.

Rank analysis Table 4 reports retained full distribution KL of the top K READOUT on Qwen3-4B as the rank grows. Most immediate output sensitivity is concentrated in a few directions: four directions already bring KL below 0.7, and $r = 64$ lies well into the saturated regime.

Table 4: Retained full distribution KL of the top K READOUT at the output interface of Qwen3-4B.

Rank r	1	2	4	8	16	32	64
Retained KL	3.4806	1.6309	0.6284	0.3277	0.1329	0.0281	0.0095

Why the uncentered factor is used The centered softmax Fisher factor removes the common mode logit direction. This suits curvature on the probability simplex but not the reconstruction of the full output from a retained span, where it degrades the global top K even when the local ranking is correct. We therefore build the READOUT from the uncentered factor.

C Choice of Layer

A hidden state can influence later positions only through attention in the blocks above it, so the layer at which we measure determines how much future influence there is to decompose. Figure 5 sweeps the analyzed layer of Qwen3-4B-Base with the logit support held fixed. We measure two quantities from the current Jacobian J_0 and the future Jacobian J_w at the same hidden state: the transport $\tau = \|J_w\|_F / \|J_0\|_F$, which measures how strongly the state still influences later positions,

and the separation $\varphi = \|J_w(I - P_t)\|_F^2 / \|J_w\|_F^2$, which measures how much of that influence lies outside the READOUT. Transport falls monotonically with depth and reaches exactly zero at the final layer, where no block remains to carry information forward. Separation rises monotonically. Early layers therefore carry strong future influence that is largely entangled with the current output, and late layers separate the two roles but retain little future influence. Their product $C = \varphi\tau$, the future influence that is both present and separable, is largest at the middle of the network, which is where we measure. The post trained Qwen3-4B shows the same maximum. For the longest horizon ($w = 4$), the maximum shifts toward shallower layer, consistent with longer range influence requiring more remaining depth.

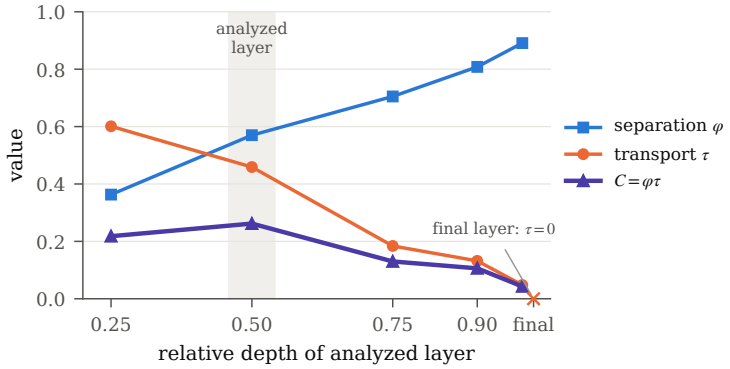

Figure 5: Depth sweep on Qwen3-4B-Base. Transport τ falls and separation φ rises with depth; their product peaks at the middle layer used in our analyses.

Table 5 gives the same trend in the main pipeline at relative depths 0.5, 0.75, and 0.9 ($w = 1$). Future sensitivity decays far faster than current sensitivity (about $220\times$ against $25\times$ between the middle and the latest layer), and the top directions of the two geometries diverge. The decreasing overlap should therefore not be read as pure specialization, because late layers simultaneously have fewer paths through which $h_{\ell,t}$ can influence later positions.

Table 5: Depth trend in the main pipeline on Qwen3-4B ($w = 1$). Traces are medians over anchors.

Relative depth	0.50	0.75	0.90
Current Fisher trace	0.114	0.021	0.005
Future Fisher trace	3.0×10^{-3}	1.0×10^{-4}	1.4×10^{-5}
Top span cosine	0.962	0.842	0.693
Ω_1	0.758	0.510	0.386

D SIGNED SHARING: DEFINITIONS, CONTROLS, AND ROBUSTNESS

Energy accounting For the components in Eqs. 2 and 3, define

$$E_R = \frac{\|R_w\|_2^2}{\|g_w\|_2^2}, \qquad E_S = \frac{\|S_w\|_2^2}{\|g_w\|_2^2}, \qquad E_C = \frac{\|C_w\|_2^2}{\|g_w\|_2^2}. \tag{10}$$

Because R_w is orthogonal to the READOUT and S_w and C_w occupy disjoint retained modes, $E_R + E_S + E_C = 1$ and $E_S + E_C = \Omega_w^2$. These identities are used for description only.

Permutation null For anchor i and horizon w, we keep $g_0^{(i)}$ fixed and replace its paired future gradient with $g_w^{(j)}$ from another anchor at the same horizon, $\chi_w^{\text{null}} = \cos\left(g_0^{(i)}, g_w^{(j)}\right)$ with $j \neq i$. This preserves hidden space anisotropy and the marginal structure of gradient norms while breaking the semantic and temporal pairing that defines a true record. If true pairs were no more related than shuffled pairs, about 5% of observed records would fall below the null q_{05} and 5% above q_{95}. A large fraction in both tails therefore indicates stronger signed dependence in both directions, not a global shift toward conflict.

Mode level coexistence The record level null does not quantify how often substantial SYNERGY and CONFLICT coexist inside one shared gradient. For this we measure the fraction of records that are globally positive yet mode conflicted,

$$\Pr\left[g_0^\top g_w \geq 0 \ \wedge \ E_C > \delta\right], \tag{11}$$

for a fixed threshold δ, together with the joint distribution of (E_S, E_C) for true and donor shuffled future gradients in the same current basis. A non trivial value of Eq. 11 means that a scalar rule such as PCGrad (Yu et al., 2020) would treat the pair as non conflicting even though conflicting modes remain inside the READOUT. Table 6 lists the three tests and what each establishes. Rows 2 and 3 use the donor control run of Appendix F. Because donor gradients place little energy in the READOUT, their low rate in Eq. 11 partly reflects their smaller shared energy; the conflict share, which is normalized by the shared energy, is close to one half for both, so the evidence for coupling beyond chance is the record level enrichment and the size of the shared energy, not a preference for one sign.

Table 6: Signed sharing validation. The first test is reported in Section 3.3; the other two are the mode level controls, from the donor control run on the OLMo-2 lineage (ranges over the four checkpoints; donors from different problems).

Test	Evidence sought	Result
Observed χ_w versus permutation null	Both tails enriched beyond the 5% null expectation	Section 3.3
Joint (E_S, E_C) versus donor shuffled control	Mode energy beyond incidental sign mixing	$E_S + E_C = 0.29$ to 0.36 versus 0.05 at $w = 1$; conflict share 0.41 to 0.48 versus 0.48 to 0.50 over $w = 1, 2, 4$
Eq. 11 with $\delta = 0.05$	Conflict energy that a positive global cosine hides	0.59 to 0.66, 0.53 to 0.56, 0.39 to 0.46 at $w = 1, 2, 4$, versus at most 0.02 for donors

Basis stability Mode wise S_w/C_w labels are invariant to sign flips of individual basis vectors but can change under rotations inside nearly degenerate singular subspaces, so we use them only through aggregate energies and never interpret individual modes.

E TEMPORAL BILL DETAILS

Suffix identity Rerunning the transformer suffix from an unperturbed $h_{\ell,t}$ reproduces the full forward pass up to logit differences of 10^{-4}, and every record satisfies $Q_w \perp R_w$ and $\|Q_w\|_2^2 + \|R_w\|_2^2 = \|g_w\|_2^2$ to a relative error below 10^{-6}.

Absolute and normalized bills Normalized bills alone can assign a sharply peaked profile to a direction whose absolute effect is negligible, so every normalized result is paired with absolute KL.

Random controls Isotropic controls are unit directions drawn uniformly from the complement of the current READOUT. Their absolute bills lie at the numerical floor at every offset, so all interventions are run in float32.

Relay profiles and selectivity Table 7 gives the normalized bills of RELAY directions on a representative OLMo-2 checkpoint; the profiles reproduce within 1% across runs. Table 8 gives their absolute KL relative to isotropic controls across the four checkpoints, and Table 9 gives the target horizon ratio plotted in Figure 2c.

F DONOR CONTROLS

A future gradient that lies largely inside the READOUT, or a RELAY direction that acts at one offset, could reflect the geometry of the construction rather than the pairing between a position and its own future loss. To separate the two, we replace the future gradient of each anchor with that of a donor

Table 7: Normalized temporal bills of RELAY directions on a representative OLMo-2 checkpoint. The target offset is in bold.

Direction	$k = 0$	$k = 1$	$k = 2$	$k = 4$
R_1	0.000	**0.978**	0.019	0.003
R_2	0.001	0.075	**0.916**	0.008
R_4	0.001	0.067	0.061	**0.871**

Table 8: Temporal selectivity of RELAY directions across four OLMo-2 checkpoints. Ratios compare absolute KL against isotropic directions in the complement of the current READOUT.

Target	Peak at target	Target/random	Present/random
$w = 1$	4/4	770–1118×	3.2–4.1×
$w = 2$	4/4	670–963×	2.5–3.4×
$w = 4$	4/4	492–704×	2.1–2.9×

anchor at the same horizon, drawn from a different problem, and decompose it with the READOUT of the original anchor. The donor components share every property of the construction but not the pairing. As a second control we draw donors from the same problem at least 16 tokens away, and as a reverse control we decompose the anchor's own future gradient with the READOUT of the donor; both give the same conclusions, and we report donors from different problems.

Setup This control is a separate run on the OLMo-2 lineage with VersaPRM dataset. Anchors are placed by position alone: starting from the first response token, every eighth position whose farthest target is still a response token is a candidate, and at most eight are kept per sequence at even spacing. Because the rule uses no model output and all checkpoints share a tokenizer and receive byte identical plain format inputs, the four checkpoints are measured at the same positions. We use $K = r = 64$ and report $\varepsilon = 0.01$ in Eq. 4; $\varepsilon = 0.003$ gives the same conclusions, and at both values the change of the target loss along the true RELAY agrees with its first order prediction to within 4%. With a zero perturbation the logits are reproduced exactly, and $|\cos(Q_w, R_w)|$ and the energy identity error stay below 10^{-5}. Because anchors and numerics differ from the main analysis, Table 10 reports comparisons within this run rather than absolute values.

Sharing The donor future gradient places 4 to 5% of its energy in the READOUT, 1.3 to 1.6 times chance, whereas the anchor's own future gradient places 3.8 to 8.5 times more. READOUTs of different positions overlap about twice as much as random subspaces of the same rank, so the READOUT is conditioned on the position rather than shared across contexts.

RELAY Donor RELAY directions have no horizon: their normalized bills peak at $t + 1$ whatever w they come from, their centroids stay between 1.5 and 1.8, and their absolute effect at the target is within a factor of 2.5 of isotropic directions in the same complement. The true RELAY exceeds its donor at the target by a factor of 344 to 492, so the horizon selectivity of Section 3.2 comes from the anchor's own future loss. Its own horizon also dominates the other later offsets: the absolute effect of R_w at $t + w$ exceeds its largest effect at another later offset by a factor of 9 to 13 for $w = 2, 4$ and 34 to 52 for $w = 1$. Cross horizon effects are not absent, however, and remain 20 to 60 times those of isotropic directions, so the separation is one of relative strength.

Shared component Any direction inside the READOUT acts mainly on the present: a random one spends 89 to 93% of its normalized bill at t and a donor shared component 89 to 92%. The present facing character of Q_w is therefore expected from the construction. What the construction does not produce is the future effect of Q_w, which exceeds that of its donor at the target by a factor of 27 to 53. For donor components the ratio of RELAY to shared effect at the target stays below 0.5 at every horizon.

Table 9: Target horizon ratio $\rho_w = \mathcal{B}_w(R_w)/\mathcal{B}_w(Q_w)$ of absolute KL across the OLMo-2 lineage.

Horizon	Base	SFT	DPO	Instruct
$w = 1$	1.20	1.59	1.44	1.53
$w = 2$	2.58	3.03	3.11	3.23
$w = 4$	3.99	4.95	5.26	5.06

Table 10: Donor control on the OLMo-2 lineage, ranges over the four checkpoints ($\varepsilon = 0.01$). Ratios are medians of per anchor ratios for Ω^2 and ratios of medians for absolute KL at the target offset $t + w$.

	$w = 1$	$w = 2$	$w = 4$
Ω^2_w, true / donor	6.6–8.5	5.1–6.6	3.8–4.7
RELAY peak rate at $t + w$, true	0.81–0.92	0.68–0.80	0.63–0.73
RELAY peak rate at $t + w$, donor	0.40–0.47	0.26–0.30	0.18–0.20
RELAY target KL, donor / isotropic	1.7–2.5	1.9–2.3	2.1–2.3
RELAY target KL, true / donor	378–492	439–472	344–369
Shared target KL, true / donor	32–44	33–53	27–40
ρ_w of donor components	0.08–0.14	0.14–0.25	0.33–0.48

G Additional Post Training Evidence

Table 11 gives the values plotted in Figure 3 together with the bill centroids. Current rank_{99} is the number of eigendirections needed to capture 99% of the current output's Fisher energy at an anchor. Local conflict cost is the median, over records with $g_0^\top g_w < 0$ at $w = 1$, of the increase in current loss per unit decrease in future loss after the step $h'_{\ell,t} = h_{\ell,t} - \varepsilon \|h_{\ell,t}\|_2 \, g_w / \|g_w\|_2$ with $\varepsilon = 0.01$. Local synergy gain is the corresponding decrease in current loss on records with $g_0^\top g_w > 0$.

Table 11: Across the OLMo-2 lineage, broad geometric organization is stable while current output sensitivity becomes more concentrated and more costly.

Metric	Base	SFT	DPO	Instruct
Span overlap	0.199	0.190	0.196	0.202
Top span cosine	0.901	0.884	0.893	0.899
Bill centroid ($w = 1/2/4$)	1.09/1.95/3.39	1.08/1.97/3.38	1.10/1.98/3.32	1.10/1.98/3.29
Current rank_{99}	24	15	12	10
Top eigenvalue fraction	0.543	0.592	0.649	0.666
Local conflict cost	0.317	0.316	0.516	0.580
Local synergy gain	1.07	1.40	1.40	1.49

The donor control run of Appendix F, which measures all four checkpoints at the same 796 positions, removes selection effects induced by differing entropy distributions and gives the same picture: bill centroids of the RELAY stay within 0.05 of 1.07, 1.96, and 3.24 for $w = 1, 2, 4$ across stages, and the target horizon ratio at $w = 4$ rises from 3.4 in Base to 4.4 to 4.7 after post training.

H TPC Implementation

H.1 Splitting the backward pass

Let L be the number of blocks. The analyzed layer is $\ell = \lfloor L/2 \rfloor$, and the rule acts at the normalized input of the attention block of layer $\ell + 1$, the first block that can carry the state of layer ℓ to later positions. A custom autograd function returns two views of this input with identical values. Forward pre hooks feed one view to the key and value projections and the other to the query projection and the rest of the attention module. In the backward pass the function receives the two branch gradients separately and returns their sum, modified by the rule when it is active. The residual connection around the block belongs to neither branch. The rule is applied only at positions whose next token is

a response target, which are exactly the positions over which the cross entropy is averaged; all other gradients pass through unchanged.

H.2 Estimating the sensitivity metric

Before fine tuning we run 512 training sequences of at most 4096 tokens through forward and backward passes without any optimizer step. At every response position we accumulate $a_t a_t^\top$ in float64, normalize by the number of positions, and keep the leading r eigenvectors and eigenvalues. The basis is stored in the normalized input space in which the rule acts. On a subsample of (a_t, b_t) pairs we also record two diagnostics for each candidate rank, the share of future path energy inside the basis, $\phi_r = \|U_r^\top b_t\|_2^2 / \|b_t\|_2^2$, and the fraction of future path credit that the rule would change, $\chi_r = \|d_t^\star - b_t\|_2 / \|b_t\|_2$. A rank is usable only if the median χ_r reaches 0.01. On held out VersaPRM chains the median ϕ_{64} is 0.046 on Qwen3-4B-Base and 0.038 on Llama-3.2-3B-Instruct, against random baselines of 0.025 and 0.021, and the median χ_{64} is 0.052 and 0.068 (Appendix H.3).

Why a fixed subspace A per token metric built from the instantaneous direction of current path credit, $M_t = a_t a_t^\top$, is the natural rank one alternative, but it barely changes training. On Qwen3-4B-Base its median χ stayed below 5×10^{-4}, even though future path credit was 2.6 to 2.8 times larger than current path credit, because future path credit is nearly orthogonal to any single current direction. Since $\chi \le \sqrt{\phi}$ for every λ, no choice of λ can compensate. A subspace estimated over many positions captures the directions the current computation relies on across the data, which is what Eq. 6 provides.

H.3 Checking the metric against the Readout

Table 12 pairs each object of the analysis with its counterpart in TPC.

Table 12: Correspondence between the analysis of Sections 2 and 3 and TPC.

Analysis	TPC	Relation
Residual state $h_{\ell,t}$	Attention input x_t of block $\ell+1$	Where later positions first read $h_{\ell,t}$; the residual bypass is left untouched
Future gradient g_w	Future path credit b_t	Approximation 1 (first hop): b_t sums the key and value credit of all later losses at block $\ell+1$; credit through higher blocks stays in the residual path; self credit is 1.6 to 6.4% of $\|b_t\|_2$
Readout U_t	Fixed basis U of $\mathbb{E}[a_t a_t^\top]$	Approximation 2 (data-averaged basis): U captures 16 to 23 times the site sensitivity captured by a random basis
Shared $Q_w = P_t g_w$	$UU^\top b_t$	Shrunk by s_i; by linearity, for all horizons at once
Relay $R_w = (I - P_t) g_w$	$(I - UU^\top) b_t$	Passed unchanged
Sign of $a_i b_i$	Not used	Every mode keeps its sign (Section 3.3)
Current gradient g_0	Not used	a_t serves only to estimate U

Table 13 compares the fixed basis with an activation PCA basis of the attention input and a random basis, all of rank 64 and carrying the same shrink factors, on 40 held out VersaPRM problems per model before fine tuning. The site Readout is computed from the Jacobian of the current top K logits with respect to the attention input. The basis captures 16 to 23 times more of this sensitivity than a random basis and 6 to 9 times more than activation PCA, so it reflects the output function rather than activation variance. The residual Readout, the Readout of Section 2.1 at the input of the same block pushed to the attention input through the Jacobian of its normalization, overlaps the basis within a factor of two of chance, because it also reflects the residual path that bypasses the attention block; the metric should therefore be read as the Readout of the rule site.

How exact is the split The key and value path carries a median self credit of 1.6% and 6.4% of $\|b_t\|_2$, so b_t is almost entirely future credit. The query path is less pure: 99% and 86% of the norm of a_t comes from later losses that reach position t through its attention output. Because such credit

passes through the attention output of position t, 82% and 90% of the energy of a_t still lies in the site READOUT, which is why $\mathbb{E}[a_t a_t^\top]$ locates the directions position t relies on even though a_t is not current credit.

Table 13: Rule site check before fine tuning (medians). Uniform shrinks b_t along itself, so its output sensitivity ratio is exactly one.

	Qwen3-4B-Base			Llama-3.2-3B-Instruct		
Quantity	TPC	PCA	Random	TPC	PCA	Random
Site READOUT energy captured	0.400	0.046	0.025	0.496	0.085	0.021
Overlap with site READOUT	0.212	0.037	0.025	0.231	0.042	0.021
Overlap with residual READOUT	0.041	0.062	0.025	0.041	0.047	0.021
Energy of a_t inside basis	0.436	0.049	0.025	0.560	0.084	0.021
Energy of b_t inside basis (ϕ_{64})	0.046	0.053	0.024	0.038	0.055	0.021
Correction inside site READOUT	0.405	0.052	0.025	0.761	0.182	0.023
Output sensitivity of correction, relative to b_t	8.74	0.40	0.14	17.83	3.24	0.09

Direction of the correction The TPC correction places 41% and 76% of its energy in the site READOUT, against 9% and 11% for b_t and hence for the Uniform correction, and per unit norm it moves the current prediction 8.7 and 17.8 times more than a change along b_t. The two rules of Section 5.3 therefore differ in direction by design, and the difference lies along the directions the current output is sensitive to.

After fine tuning Table 14 repeats the measurement on the fine tuned models with the basis estimated before training. The share of site sensitivity captured and the direction of the correction change little under any of the three procedures. The ratio $\|b_t\|_2/\|a_t\|_2$ falls on Qwen3-4B-Base, but it falls under Vanilla SFT as well. The TPC checkpoints are measured on 40 problems and the others on 200.

Table 14: Rule site check after fine tuning with the basis estimated before training (medians).

	Qwen3-4B-Base				Llama-3.2-3B-Instruct			
Quantity	Before	Vanilla	Uniform	TPC	Before	Vanilla	Uniform	TPC
Site READOUT energy captured	0.400	0.366	0.366	0.370	0.496	0.476	0.476	0.486
Correction inside site READOUT	0.405	0.393	0.395	0.397	0.761	0.746	0.739	0.755
Output sensitivity of correction	8.74	6.11	6.47	7.33	17.83	12.97	12.23	15.59
$\|b_t\|_2/\|a_t\|_2$	1.98	1.56	1.52	1.66	1.82	1.81	1.80	1.82

H.4 ALGORITHM

Algorithm 1 summarizes the procedure. Vanilla SFT runs the same loop but returns $a_t + b_t$ unchanged, skipping lines 12 and 13. Uniform replaces line 12 with $d_t \leftarrow (1 - q_t)\, b_t$, where $q_t = \min\big(1, \|d_t^\star - b_t\|_2/\|b_t\|_2\big)$ is computed from the TPC solution of the same token.

I TPC DERIVATION

I.1 CLOSED FORM AND SIGN PRESERVATION

Write $b_t = b_{t,\perp} + \sum_{i=1}^r y_i u_i$ with $y_i = u_i^\top b_t$. The objective in Eq. 7 separates over the orthonormal basis $\{u_i\}_{i=1}^r$ and its complement. The complement carries no penalty, so its optimum is $b_{t,\perp}$. For retained coordinate i, minimizing $\frac{1}{2}(z_i - y_i)^2 + \frac{\lambda\hat{\mu}_i}{2} z_i^2$ gives $z_i = s_i y_i$ with $s_i = 1/(1+\lambda\hat{\mu}_i) \in (0, 1]$ because $\lambda \geq 0$ and $\hat{\mu}_i \geq 0$. Hence $u_i^\top d_t^\star$ has the sign of $u_i^\top b_t$, and so does its product with $u_i^\top a_t$. After the ramp and the trust region, the retained coordinate becomes $y_i\big(1 - \kappa\alpha_t(1 - s_i)\big)$ with $\kappa, \alpha_t \in [0, 1]$. The factor lies in $[s_i, 1]$, so the sign is again preserved.

Algorithm 1 Supervised fine tuning with Temporal Proximal Credit

Require: model θ, data $\mathcal{D}$, rule site x_t, rank r, coefficient λ, trust region ϵ_{TR}, ramp length K_{ramp}

Measurement, once before fine tuning

$M \leftarrow 0$
for each of N sequences from $\mathcal{D}$ **do**
  run forward and backward; at each response position t, $M \leftarrow M + a_t a_t^\top$
end for
$(U, \mu) \leftarrow$ leading r eigenpairs of M/count; $\quad s_i \leftarrow 1/\big(1 + \lambda\, \mu_i/\bar{\mu}\big)$

Fine tuning

for update $k = 1, \ldots, K$ **do**
  $\kappa \leftarrow \min(1, k/K_{\mathrm{ramp}})$
  for each micro batch in the accumulation group **do**
    run the forward pass unchanged, then the backward pass until x_t
    for each response position t **do**
      $(a_t, b_t) \leftarrow$ current and future path credit at x_t
      $d_t \leftarrow (I - UU^\top)\, b_t + U\, \mathrm{Diag}(s)\, U^\top b_t$ ▷ Uniform: $d_t \leftarrow (1 - q_t)\, b_t$
      $\Delta_t \leftarrow \kappa\, (d_t - b_t)$; $\quad \alpha_t \leftarrow \min\big(1, \epsilon_{\mathrm{TR}} \|a_t + b_t\|_2 / \|\Delta_t\|_2\big)$
      return $a_t + b_t + \alpha_t \Delta_t$ as the gradient of x_t
    end for
    continue the backward pass and accumulate parameter gradients
  end for
  clip the gradient and take one optimizer step
end for

I.2 Forward identity

The two views of x_t have identical numerical values, and the split exists only in the autograd graph. Logits and loss are therefore identical to Vanilla SFT up to numerical precision before the first optimizer update, and with no rule attached the backward pass returns the SFT gradient.

J Training Details

J.1 Training data

We use the math subset of Mixture of Thoughts (Fein-Ashley et al., 2025), discard examples whose response exceeds 4096 tokens, and sample 20k of the remaining examples uniformly at random. Each example is a single user turn followed by a single assistant turn, rendered with the model's chat template, and prompt tokens are masked from the loss.

J.2 Configuration

Table 15 lists the configuration. Vanilla SFT, DFT, Uniform, and TPC use the same settings and seed and therefore see identical batches in the same order. DFT (Wu et al., 2026) follows the original formulation and multiplies the cross entropy of each token by its predicted probability with the gradient stopped through the weight; it has no additional hyperparameters.

K Evaluation Details

K.1 Benchmarks

Table 16 summarizes the evaluation suite. MATH500, Minerva, the AIME sets, and BeyondAIME match the training domain. MMLU-Pro, BBH, MuSR, and ZebraLogic test whether changes transfer to knowledge intensive and logical reasoning outside mathematics.

MATH500 (Lightman et al., 2024) is a 500-problem subset of MATH, Minerva Math (Lewkowycz et al., 2022) contains undergraduate and graduate STEM problems from MIT OpenCourseWare, and

Table 15: Training configuration shared by all methods. The last four rows apply to TPC (and to Uniform, which reuses its correction norm).

	Qwen3-4B-Base	Llama-3.2-3B-Instruct
Blocks	36	28
Rule site (attention input of block)	19	15
Maximum sequence length	4096	
Optimizer	AdamW, weight decay 0	
Learning rate	2×10^{-6}, cosine schedule, 3% warmup	
Batch	16	
Epochs	2	
Gradient clipping	1.0	
Precision	float32, gradient checkpointing	
Metric	rank $r = 64$, estimated from 512 sequences, fixed during training	
Proximal coefficient	$\lambda = 0.25$	
Trust region	$\epsilon_{\mathrm{TR}} = 0.1$	
Ramp	linear over the first 5% of updates	

Table 16: Evaluation benchmarks

Benchmark	Domain	Items	Subset used	Decoding
MATH500	competition math	500	full	greedy
Minerva Math	university STEM	272	full	greedy
AIME 2025	olympiad math	30	full	sampled, $k \in \{8, 16\}$
AIME 2026	olympiad math	30	full	sampled, $k \in \{8, 16\}$
BeyondAIME	olympiad math	100	full	sampled, $k \in \{8, 16\}$
MMLU-Pro	14 subjects	560	40 per subject	greedy
BBH	deductive reasoning	750	full	greedy
MuSR	narrative reasoning	756	full	greedy
ZebraLogic	logic grid puzzles	1000	random sample	greedy

BeyondAIME (Seed et al., 2025) contains olympiad style problems with integer answers. MMLU-Pro (Wang et al., 2024) is a ten option multiple choice benchmark from which we sample 40 questions per subject, from BBH (Suzgun et al., 2023) we use the deductive reasoning tasks, MuSR (Sprague et al., 2024) covers murder mysteries, object placements, and team allocation, and ZebraLogic (Lin et al., 2025) consists of logic grid puzzles, of which we evaluate 1000 at random from mc mode.

K.2 Decoding and answer extraction

AIME 2025, AIME 2026, and BeyondAIME contain only 30, 30, and 100 problems, so a single greedy run is too noisy; we evaluate them with $k \in \{8, 16\}$ sampled completions per problem. All other benchmarks use greedy decoding at temperature 0. Sampled evaluation draws k independent completions per problem at temperature 0.7, top p 0.95, and top k 20. Every completion is limited to 4096 generated tokens. Avg@k is the mean accuracy over the k completions, averaged over problems, and Pass@k is the fraction of problems for which at least one completion is correct.

We extract the final answer from the last `\boxed{}` in the response and compare it with the reference answer. If a response contains no `\boxed{}`, we append `Answer:\boxed{` to the end of the response and let the model complete it, so that every response yields an extractable answer. The same extraction procedure is applied to all methods.