ORTHOGONALIZATION BUYS A LEARNING RATE, NOT ROBUSTNESS, IN RL POST-TRAINING

Anonymous authors
Paper under double-blind review

ABSTRACT

Orthogonalized optimizers such as Muon are reported to collapse in RL with verifiable rewards (RLVR), and the spectral filter Pion to beat AdamW. Step size accounts for Muon's collapse and for Pion's stability. The public Pion code runs Muon at thirteen times AdamW's measured step, and Pion's filter drives toward zero every matrix whose leading direction holds less than $2/3$ of the energy. With each matrix's step matched in size to AdamW's, in GRPO on Qwen3, the flat-spectrum orthogonalized step tolerates within 60 steps a 2.0–$3.0\times$ larger learning rate at 0.6B and 1.4–$2.8\times$ at 1.7B; at 4B it is not determined, and over 200 steps without a penalty it is gone. That margin is not robustness: a flat step moves the layer outputs, on the inputs they receive, about three times less than an AdamW step of the same size. Matched in that output displacement it collapses at least as often as AdamW in 14 of 15 settings, and earlier in 11 of the 12 that differ ($p = 0.003$ over settings, 0.062 over groups of model, task and horizon); matched at every step in how much it changes the policy, its collapse fraction is within the registered ± 25 points of AdamW's, and at thirty-two seeds per arm within a tighter ± 20 ($+8$, 90% interval -4 to $+20$). At each arm's best learning rate the two end within a point of each other in accuracy (75.7% and 75.6%; under a practitioner's recipe, 72.3% and 71.6%), the flat step getting there sooner. The key tests were registered in advance, and the ones that failed are reported with the rest.

1 INTRODUCTION

Reinforcement learning with verifiable rewards (RLVR), usually with group-relative policy optimization (GRPO; Shao et al., 2024) and its variants (Liu et al., 2025b; Yu et al., 2025), has become the standard way to elicit reasoning from pretrained language models, and its optimizer is almost always AdamW (Kingma & Ba, 2015; Loshchilov & Hutter, 2019). In pretraining, orthogonalized updates are a strong alternative: Muon replaces the momentum matrix by an approximation of its polar factor (Jordan et al., 2024) and scales to large language models once its update magnitude is aligned with AdamW (Liu et al., 2025a). Whether orthogonalized updates also help in post-training is open.

The evidence so far is mixed. Fan et al. (2026) report that Muon collapses in RLVR post-training of Qwen3 models, whereas their high-pass variant Pion, which drives dominant singular values to one and suppresses the tail, improves over AdamW. In agentic RL, Ruan et al. (2026) find that Muon tolerates a more aggressive effective step than AdamW and that its late-stage gain disappears once the update RMS of the hidden matrices is matched to AdamW. Partial spectral flattening helps fine-tuning (Dong & Sawin, 2026), and RLVR updates are sparse yet nearly full-rank (Mukherjee et al., 2025) and avoid the principal directions of the pretrained weights (Zhu et al., 2025). RMS matching removes the magnitude confound but, as we show, leaves the flat direction a two- to threefold learning-rate headroom; whether that headroom is robustness or a further rescaling is the question this paper answers.

Direction and magnitude are confounded. An orthogonalized update has no intrinsic scale; its magnitude is fixed by a convention. The Moonlight convention multiplies the polar factor by $0.2\sqrt{d_{\max}}$ so that the update RMS matches the typical range of AdamW (Liu et al., 2025a); the

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.

AdamW-fused optimizer classes in the public Pion RLVR code default to 5, twenty-five times that. For filtered directions the output norm of the Pion suppression filter depends sharply on the momentum spectrum (Proposition 1): the same scale factor yields either an update of order $d_{\min}^{-1/2}$ of Muon's or one that is exactly zero in bfloat16. Comparing AdamW, Muon and Pion therefore changes the direction, the magnitude and a spectrum-dependent effective learning rate at once.

Questions. We disentangle these factors in GRPO post-training and ask:

- **(Q1)** Is the reported collapse of Muon in RLVR caused by orthogonalization or by the magnitude of the resulting update?
- **(Q2)** When the magnitude is matched per matrix and per step, how does the spectrum of the update direction affect stability and accuracy?
- **(Q3)** What sets the learning rate a direction tolerates, and the progress it makes there: the size of the step or of the weight change in some norm, or how much of the step lands where the activations are?

Contributions.

- **The comparison is confounded (Q1).** The public Pion RLVR code runs Muon at 25 times the Moonlight scale (13 times AdamW's measured step), and Proposition 1 shows its filter gates Pion's step by the momentum spectrum; at a moderate scale, orthogonalization does not destabilize GRPO (Section 5.2).
- **At matched magnitude the spectrum sets a learning rate (Q2).** The flat polar direction tolerates 2.0–3.0 times AdamW's rate within 60 steps on Qwen3-0.6B, and a larger rate than AdamW on Qwen3-1.7B, on MATH and, above the registered grid, on Llama-3.2-1B; near-rank-one directions tolerate less (Sections 5.3, 5.5 and 5.6).
- **Matched in displacement, the flat direction is no more stable (Q3).** Matched per matrix to the Adam step's input-aligned norm $\eta\sqrt{\operatorname{tr}(U_t\Sigma_t U_t^\top)/\operatorname{tr}(\Sigma_t)}$, with the rest of the network untouched, the flat direction collapses more often or earlier than AdamW in 11 of the 12 settings where they differ, less often in one cell at 4B (six runs) (Table 15), also for Muon's own momentum. That residual is one constant: shrunk by it the flat direction is as stable as AdamW, and rescaled at every step to AdamW's measured one-step KL its collapse fraction differs from AdamW's by $+8$ points (90% interval -4 to $+20$, inside the registered ± 25).
- **A ten-step measurement, and its reach.** Per direction, a/η is nearly constant over the first ten steps; registered in advance, it placed the threshold of the public Muon twenty times above any rate we had run it at, and it fails at Qwen3-4B, for near-rank-one directions and for random rotations (Section 5.7).

2 Related work

Muon and its variants. Muon orthogonalizes the momentum of each hidden matrix with a quintic Newton–Schulz iteration (Jordan et al., 2024); weight decay and an RMS-matching scale of $0.2\sqrt{d_{\max}}$ make AdamW learning rates transferable (Liu et al., 2025a). Analyses of spectral flattening concern the maximal stable step (Nguyen et al., 2026), momentum as a spectral filter (Li et al., 2026), a flat-spectrum bias (Beneventano et al., 2026) and progress along informative directions (Shen et al., 2026); adaptive variants combine orthogonalization with input or element-wise second moments (Du & Su, 2026; Si et al., 2025), neuron-wise normalization (Li et al., 2025), norm-based steps (Zhang et al., 2026) or module-wise signal-to-noise rescaling (Wen et al., 2026). Muon^p interpolates the spectrum with fractional powers $US^pV^\top$ (Dong & Sawin, 2026), and Massena et al. (2026) choose a per-layer Schatten-p geometry between SGD and Muon. These methods change the spectrum and the step size together; we match the per-matrix RMS to the Adam step and vary the spectrum of the direction on the singular vectors of the momentum, from flat to concentrated. A related line asks in which norm a step should be measured (Bernstein & Newhouse, 2024; Large et al., 2024); for pretraining, Filatov et al. (2025) find the output layer's operator norm to be the invariant along which optimal hyperparameters transfer, whereas in RLVR the operator norm is not what orders stable from collapsed runs. Newton-Muon uses the input second moment to precondition the step (Du & Su, 2026); we use it only to measure one.

Orthogonalized optimizers in RL post-training. Fan et al. (2026) attribute the failure of spectral whitening in RLVR to amplified noisy tail directions and lost per-head specialization, and propose Pion, a fixed polynomial high-pass iteration applied to the normalized momentum. We show that this filter's output norm is controlled by the momentum spectrum and that, once its magnitude is matched, its concentrated direction, whole-matrix or per attention head as in their main RLVR recipe, is less stable than the flat polar factor. Ruan et al. (2026) show in agentic RL that the late-stage gain of Muon is tied to its scale convention via an RMS-matched control; we find the same confound in RLVR and go past it: at matched RMS the spectrum still changes which learning rate a direction tolerates, what it changes is the input-aligned norm, and that norm accounts for the one-step policy change of every gradient-derived direction but not of an arbitrary one.

Update geometry in RLVR. RL fine-tuning updates a small subnetwork, yet the update to almost every matrix is nearly full-rank (Mukherjee et al., 2025), and RLVR learns off the principal directions of the pretrained weights with little spectral drift (Zhu et al., 2025). Because RLVR on Qwen models can improve even under spurious rewards (Shao et al., 2025), we compare optimizers through matched controls, not by absolute gains.

3 METHOD

3.1 SETTING

GRPO. For each prompt the policy samples g completions with binary verifiable rewards $r_1, \dots, r_g$ and standardized advantages $\hat{r}_i = (r_i - \bar{r})/(\mathrm{std}(r) + 10^{-6})$, with the sample standard deviation over the group (Shao et al., 2024). We take one optimizer step per rollout batch, so the PPO ratio is one at the update and clipping never binds. The surrogate for a batch of N completions is normalized by a constant,

$$\mathcal{L}(\theta) = -\frac{1}{NL_{\max}} \sum_{i=1}^{N} \sum_{j=1}^{|y_i|} \hat{r}_i \, \omega_{ij} \, \exp\big(\ell_{ij}(\theta) - \mathrm{sg}[\ell_{ij}(\theta)]\big), \qquad \omega_{ij} = \min\big(e^{\mathrm{sg}[\ell_{ij}] - \ell_{ij}^{\mathrm{roll}}}, \, C\big), \quad (1)$$

where ℓ_{ij} and $\ell_{ij}^{\mathrm{roll}}$ are the log-probabilities of token j under the trainer and the inference engine, sg is stop-gradient, and the truncated importance weight ω_{ij} corrects the engine–trainer mismatch. The completions are split into K equal micro-batches; with the constant normalizer, the step gradient of a matrix $W \in \mathbb{R}^{d_1 \times d_2}$ is the equal-weight average $G_t = \frac{1}{K} \sum_k \hat{g}_{t,k}$ of micro-batch gradients.

Per-matrix updates. The attention (`q, k, v, o`) and MLP (`gate, up, down`) projection matrices form the *matrix branch*; all other parameters always receive AdamW. For a matrix, all arms maintain the same Adam moments

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1-\beta_1) G_t, \quad & v_t &= \beta_2 v_{t-1} + (1-\beta_2) G_t^{\odot 2}, \\ \hat{m}_t &= \frac{m_t}{1-\beta_1^t}, & A_t &= \frac{\hat{m}_t}{\sqrt{v_t/(1-\beta_2^t)} + \epsilon}, \end{aligned} \quad (2)$$

and apply $W \leftarrow (1 - \eta\lambda)W - \eta \, \tilde{D}_t$ with decoupled weight decay λ and learning rate η. Arms differ only in $\tilde{D}_t$.

3.2 A MAGNITUDE-MATCHED FAMILY OF DIRECTIONS

Let $\mathrm{RMS}(X) = \|X\|_F / \sqrt{d_1 d_2}$. Given a nonzero direction D_t computed from the arm's Adam moments, we apply

$$\tilde{D}_t = \frac{\|A_t\|_F}{\|D_t\|_F} D_t, \qquad \text{so that} \qquad \mathrm{RMS}(\tilde{D}_t) = \mathrm{RMS}(A_t) \text{ for every matrix and step.} \quad (3)$$

All arms use the same moment rule, learning rate, weight decay and parameter routing, and each step takes the RMS of the Adam step computed from that arm's own moments; the moment values themselves differ across arms, whose rollouts differ. The three directions below keep the singular vectors of $\hat{m}_t$ and change only its singular values; the Adam direction also rotates them, so the comparison with AdamW is not purely spectral. The Adam direction $D_t = A_t$ is the AdamW update;

the AdamW arms run `torch.optim.AdamW`. The family contains three further directions. (i) *Polar*: $D_t = \mathrm{NS}_5(\hat{m}_t)$, the quintic Newton–Schulz approximation of the polar factor with Muon's coefficients, whose singular values are all close to one. (ii) *Pion filter*: $D_t = P_{k_s}(\hat{m}_t)$, the suppression filter of Proposition 1, evaluated in float64 so that its direction survives (Appendix C); it is close to rank one (Remark 2). (iii) *Edge-truncated polar*: $D_t = N_t = U_{k_t} V_{k_t}^\top$, the polar factor of $\hat{m}_t$ restricted to the k_t singular directions above a noise edge $e_t = c_{\mathrm{edge}}\, \sigma_{\max}(\Delta_t)\sqrt{w_t/K}$ estimated from the micro-batch gradient difference $\Delta_t = (\hat{g}_{t,1} - \hat{g}_{t,2})/\sqrt{2}$, with $c_{\mathrm{edge}} = 1.05$ and the exact noise weight w_t of the bias-corrected momentum (Lemma 1); k_t is typically one, and the step falls back to A_t when $k_t = 0$, which is frequent (Appendices F.1 and F.2 and Table 10). A per-matrix gate between A_t and N_t, a probe of online direction selection, is in Appendix F. Unlike Muon, the polar arm orthogonalizes the bias-corrected Adam momentum $\hat{m}_t$.

Operator-norm matching. As a control for Section 5.4, the operator-norm-matched variant sets $\tilde{D}_t = (\|A_t\|_2/\|D_t\|_2)D_t$, the largest singular value estimated by power iteration. By Proposition 2(d) this multiplies the step of Equation (3) by $\sqrt{r_s(D_t)/r_s(A_t)}$, where r_s is the stable rank: concentrated directions then take a smaller RMS step than AdamW and flat directions a larger one (Figure 2).

Three yardsticks for a step. Two steps of equal size can move the model differently, so we compare directions under three matchings, from the parameters to the policy. *RMS*: each matrix step has the Frobenius norm of its own Adam step (above). *Input-aligned norm*: what a step U_t of learning rate η does to a layer's outputs on the inputs it actually receives, $a_t = \eta\sqrt{\mathrm{tr}(U_t \Sigma_t U_t^\top)/\,\mathrm{tr}(\Sigma_t)}$ with Σ_t the second moment of those inputs; it lies between the Frobenius norm (isotropic inputs) and the operator norm (inputs along the top singular direction, Proposition 2), and a/η, the aligned norm per unit learning rate, is a property of a direction measured in the first ten steps. *Policy change*: the one-step KL between the policies before and after the step, measured at every step on the step's own samples. Section 5.4 matches in all three.

Contrast with scale conventions. Muon-type optimizers set $\tilde{D}_t = \kappa\sqrt{d_{\max}}\, O_t$ for the orthogonalized output O_t and $d_{\max} = \max(d_1, d_2)$; a full-rank polar factor has $\|O_t\|_F = \sqrt{d_{\min}}$, hence $\mathrm{RMS}(\tilde{D}_t) = \kappa$. This is 0.2 under the Moonlight convention (Liu et al., 2025a) and 5 in the public Pion RLVR code (Fan et al., 2026). We run that code, whose momentum and precision also differ from ours (Appendix C), as separate public-code arms.

4 Analysis

Both statements concern a single matrix; proofs are in Appendix A, and Appendix B checks each statement numerically, in part against the implementation.

The Pion suppression map. For a nonzero $M \in \mathbb{R}^{d_1 \times d_2}$ with singular values $\sigma_1 \geq \cdots \geq \sigma_{d_{\min}}$, let $s_i = \sigma_i/\|M\|_F$, so that $\sum_i s_i^2 = 1$; we call s_1^2 the *top energy share*. The suppression step $X \leftarrow (2.5XX^\top - 1.5(XX^\top)^2)X$ applied to $X = M/\|M\|_F$ maps the singular values through $f_s(s) = 2.5s^3 - 1.5s^5$ and leaves the singular vectors unchanged; k_s steps give $P_{k_s}(M) = \sum_i f_s^{k_s}(s_i) u_i v_i^\top$. The public RLVR configuration uses $k_s = 5$ suppression steps and no promotion steps (Fan et al., 2026).

Proposition 1 (Suppression collapse)**.** *Let* $s_\star = \sqrt{2/3}$.

(a) f_s *is nondecreasing on* $[0,1]$ *and maps* $[0,1]$ *into itself. Its fixed points there are* 0, $s_\star$ *and* 1, *with* $f_s'(0) = f_s'(1) = 0$ *and* $f_s'(s_\star) = 5/3$, *so* $s_\star$ *is repelling.*

(b) If $0 < s_0 < s_\star$, *the iterates* $s_j = f_s^j(s_0)$ *decrease strictly to* 0. *They satisfy* $s_j \leq h(s_0)^j s_0$ *with* $h(s) = 2.5s^2 - 1.5s^4 < 1$, *and* $s_j \leq \sqrt{2/5}\,(\sqrt{5/2}\, s_0)^{3^j}$, *a bound that decays doubly exponentially when* $s_0 < \sqrt{2/5}$. *If* $s_\star < s_0 \leq 1$, *then* $s_j \uparrow 1$.

(c) At most one singular value exceeds $s_\star$. *If the top energy share satisfies* $s_1^2 < 2/3$, *then* $P_{k_s}(M) \to 0$ *as* $k_s \to \infty$. *If* $s_1^2 > 2/3$, *then* $P_{k_s}(M) \to u_1 v_1^\top$, *and if* $s_1^2 = 2/3$, *then* $P_{k_s}(M) \to s_\star u_1 v_1^\top$. *In all cases* $\|P_{k_s}(M)\|_F^2 \leq 4/3$.

(d) Under the convention $\tilde{D} = \kappa\sqrt{d_{\max}}\,P_{k_s}(M)$*, the update satisfies* $\mathrm{RMS}(\tilde{D}) = \kappa\|P_{k_s}(M)\|_F/\sqrt{d_{\min}} \le \kappa\sqrt{4/3}/\sqrt{d_{\min}}$*, whereas a full-rank polar factor gives* $\mathrm{RMS} = \kappa$*.*

With $k_s = 5$ the leading value after five steps is 0.542 for $s_1 = 0.8$ but 7.2×10^{-6} for $s_1 = 0.7$ (Table 2), and for 256×768 matrices the bfloat16 output is exactly zero in every entry once the top energy share is at most 0.2 (Table 3). By part (d), surviving updates are small as well: for the 1024-dimensional side shared by all projection matrices of Qwen3-0.6B, $\kappa = 5$ gives $\mathrm{RMS}(\tilde{D}) \le 0.18$, against 5 for Muon under the same factor. The effective learning rate of public-code Pion is therefore set by the momentum spectrum, not by κ. Promotion steps lower the survival threshold but keep it relative to $\|M\|_F$ (Remark 3). When the filter output survives, it is nearly rank one (Remark 2).

Spectral concentration. Magnitude matching fixes a step's Frobenius norm, not its distribution over singular directions. At equal RMS a rank-one step has $\sqrt{d_{\min}}$ times the operator norm of a polar factor (32 for $d_{\min} = 1024$), so it moves a layer's output up to $\sqrt{d_{\min}}$ times further along its leading input direction; for isotropic inputs the output displacement depends only on the Frobenius norm, and for the inputs the model actually sees it is the input-aligned norm (Proposition 2, stated and proved in Appendix A). This is one step, not a stability guarantee (Remark 1).

5 Experiments

5.1 Protocol

Models, data and training. We train Qwen3-0.6B, with all arms, and Qwen3-1.7B, with key arms (Yang et al., 2025), on the GSM8K training set (Cobbe et al., 2021), with the chat template, thinking disabled, and a binary reward for a correct boxed answer. Each step samples 8 completions for each of 128 prompts (temperature 1, at most 1024 new tokens) from a vLLM server (Kwon et al., 2023), which reloads the bfloat16 policy after every step. The loss is Equation (1) with $N = L_{\max} = 1024$ and importance weights capped at $C = 2$; there is no entropy bonus, gradient clipping or KL penalty (except Section 5.6), so that instability is not masked. Clipping does not change what we measure: in a registered control, a global clip active on 40% of AdamW's steps leaves AdamW collapsing in 2 of 2 runs and polar-match in none (Table 20). Step statistics (update norms, stable rank, aligned norm, drift) are logged for *probe matrices*: the seven attention and MLP projections of the first, middle and last decoder layers (21 on Qwen3-0.6B), with input activations from each step's first micro-batch. Learning rates are constant and runs last 60 steps unless stated (the 200-step runs of Section 5.6; Table 5 and Appendix D).

Evaluation and collapse. Greedy GSM8K test accuracy (1319 problems) is evaluated every 10 steps and reported as the change from step 0 to the step-50 endpoint. The *collapse step* is the first step after step 10 at which the training reward falls below half of its mean over steps 1–10; runs are stopped once collapsed into the absorbing state observed on Qwen3-0.6B (five consecutive steps with reward ≤ 0.02 and $\ge 90\%$ truncated responses), and step 50 remains the endpoint for runs that reached it. Means and standard errors are over n seeds, with n stated everywhere. Greedy vLLM decoding is not deterministic, which bounds what single-seed comparisons resolve to about 2.5 points (Appendix D).

5.2 Q1: magnitude or orthogonalization?

It is the scale, not the orthogonalization. With $\kappa = 5$ and $\eta = 10^{-6}$ on Qwen3-0.6B ($n = 3$), GSM8K accuracy first rises by up to 8.6 points and the training reward then collapses in every seed (steps 44/40/43), with entropy and truncation rising together (Figure 4 and Table 1); on Qwen3-1.7B (one seed) at step 57; these collapses are analogous to the one Fan et al. (2026) report under a different protocol (Appendix C). The same public code at the Moonlight scale $\kappa = 0.2$ is stable ($n = 3$; accuracy change 3.4 points), although its step is then smaller than AdamW's own (0.2 against 0.305 per unit learning rate), so that arm isolates the scale inside the public code rather than showing that orthogonalization at AdamW's step size is safe; the matched control for that is the polar direction at AdamW's RMS (polar-match, $n = 3$ and 1 on the two sizes), which is stable in every seed at $\eta = 10^{-6}$ and which also replaces the EMA momentum and the Newton–Schulz precision.

Table 1: Accuracy and collapse (remaining learning rates: Table 9). Δacc: change of GSM8K test accuracy (points) from step 0 to the endpoint — step 50, or the last evaluation of a run stopped earlier by the collapse rule — mean $\pm$ s.e.m. over n seeds; Δacc_{60}: the same at step 60, which shows collapses falling after the endpoint; peak: largest change at any evaluation; collapse step: of the individual seeds. Differences below 2.5 points between single-seed arms are not resolved (Section 5.1); rows use each seed's first execution (repeats in Tables 8 and 27).

Arm	η	n	Δacc	Δacc_{60}	peak Δacc	collapse step
Qwen3-0.6B, reference						
AdamW	10^{-6}	3	5.9 ± 0.2	6.8	6.4	none
AdamW	$2{\cdot}10^{-6}$	3	10.2 ± 0.1	12.0	12.0	none
AdamW	$5{\cdot}10^{-6}$	3	-63.8 ± 0.0	-63.8	9.4	35/34/32
Qwen3-0.6B, public code, κ convention						
Muon, $\kappa = 5$	10^{-6}	3	-16.0 ± 2.8	–	8.6	44/40/43
Muon, $\kappa = 0.2$	10^{-6}	3	3.4 ± 0.5	3.7	3.6	none
Pion, $\kappa = 5$	10^{-6}	3	9.1 ± 0.8	9.4	9.5	none
Per-head Pion, $\kappa = 5$	10^{-6}	3	10.2 ± 0.9	8.0	10.6	none/60/none
Qwen3-0.6B, RMS matched to AdamW (Equation (3))						
Polar	10^{-6}	3	4.0 ± 0.2	4.2	4.5	none
Edge-truncated polar	10^{-6}	3	-14.8 ± 23.9	-32.3	10.7	57/51/35
Pion filter	10^{-6}	3	-40.6 ± 22.6	-61.9	10.3	36/34/50
Polar	$5{\cdot}10^{-6}$	3	11.4 ± 0.5	11.7	11.8	none
Edge-truncated polar	$5{\cdot}10^{-6}$	3	-62.8 ± 1.5	-64.4	1.5	16/11/16
Pion filter	$5{\cdot}10^{-6}$	3	-63.7 ± 0.4	-63.8	0.0	17/11/11

The effective step of public-code Pion is small. Public-code Pion ($\kappa = 5$) gains 9.1 ± 0.8 points without collapsing in all 3 seeds, with a learning-rate-normalized update RMS of 0.031 against 0.305 for AdamW at $2{\cdot}10^{-6}$ and an exactly zero bfloat16 update on 47% of probe matrices per step (Proposition 1(d); Table 3): its stability is consistent with a small, spectrum-gated step, and its gain lies within the resolution of what the accuracy line of the matched family gives at its displacement (Figure 8). Raised in learning rate, both public Pion recipes collapse in every run (per-head from $1.95{\cdot}10^{-6}$, whole-matrix from $3{\cdot}10^{-6}$; registered, Appendix E.7), so their usable range lies below $2{\cdot}10^{-6}$. At matched RMS the whole-matrix filter collapses, and so does the nearly as concentrated per-head filter of their main RLVR recipe (10.2 ± 0.9 points at $\kappa = 5$, $n = 3$): matched, it collapses in 3 of 3 runs at 10^{-6} and 0 of 3 at $2.5{\cdot}10^{-7}$, where the aligned norm placed it in a registered test (Appendix E.7).

5.3 Q2: SPECTRUM AT MATCHED MAGNITUDE

Within 60 steps, the flat polar direction remains stable where AdamW collapses. At $\eta = 5{\cdot}10^{-6}$ on Qwen3-0.6B (Table 1), polar-match gains 11.4 ± 0.5 points without collapse in all 3 seeds. AdamW at the same learning rate first gains up to 9.4 points and then collapses in all 3 seeds (steps 35/34/32); the two arms use the same moment rule, learning rate and weight decay, and each matches its step to the RMS of its own Adam step. A learning-rate scan (Table 27) brackets both thresholds: polar-match is stable in every seed at 2, 3 and $5{\cdot}10^{-6}$, collapses in 3 of 6 runs at $6{\cdot}10^{-6}$ and in almost all runs at $7{\cdot}10^{-6}$ and 10^{-5}; AdamW is stable at 1 and $2{\cdot}10^{-6}$, collapses in 1 of 6 at $2.5{\cdot}10^{-6}$, in all but one run at $3{\cdot}10^{-6}$ and in all at $5{\cdot}10^{-6}$ (the scan's original seeds, on which the fits of this section are frozen; the seeds added later for Table 15 give 5 of 35 at $2.5{\cdot}10^{-6}$). Taking the threshold as the lowest rate at which any run collapses within the horizon gives ($5{\cdot}10^{-6}$, $6{\cdot}10^{-6}$] against ($2{\cdot}10^{-6}$, $2.5{\cdot}10^{-6}$], a ratio between 2.0 and 3.0 (2.0–2.8 if most runs at a rate must collapse); these boundary collapses all fall in the last steps of the horizon (from step 54).

Concentrated directions collapse first. The edge-truncated polar and Pion-filter directions place the same RMS in few singular directions. At $\eta = 5{\cdot}10^{-6}$ they collapse at steps 16/11/16 and 17/11/11, earlier than every AdamW seed. At $\eta = 10^{-6}$, where AdamW and polar-match are stable in all seeds, both collapse in every seed of Table 1 (one repeated edge-truncated execution is stable), after peak gains of 10.7 and 10.3 points: the edge-truncated direction at steps 57/51/35 ($n = 3$) and the Pion filter at steps 36/34/50 ($n = 3$), two of them after the step-50 endpoint, so Table 1 understates the instability.

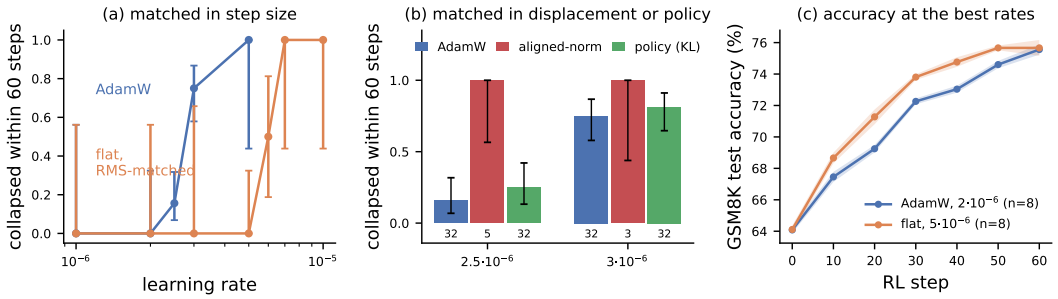


Figure 1: Qwen3-0.6B, GSM8K, 60 steps, one run per seed. (a) Matched in step size (RMS), the flat direction tolerates a larger learning rate than AdamW. (b) Matched per matrix in output displacement (aligned norm) it collapses more often than AdamW; matched at every step in one-step policy change (KL) about as often; numbers under the bars are runs. (c) At each arm's best stable rate both end at the same accuracy, the flat direction sooner (mean ± s.e.). Intervals are 95% Wilson intervals.

5.4 Q3: WHAT SETS THE LEARNING RATE A DIRECTION TOLERATES?

No norm of the step or of the weight change draws the boundary. Every RMS-matched arm takes the Frobenius norm of the Adam step, yet they range from stable to the earliest collapses. Nor does the operator norm: polar-match at 10^{-5} collapses with steps of operator norm 2.5×10^{-4} (per-step logs), below the 3.1×10^{-4} of stable AdamW at $2{\cdot}10^{-6}$, and operator-norm-matched controls collapse in a direction-dependent order (Table 7).

Near the threshold the outcome is noisy: repeats of one seed change outcome in 4 of 30 groups (Table 8) and a registered test for a separating statistic failed (Appendix E.8), so we model the *probability* of collapse.

The input-aligned norm. We use the norm of Section 3.2, averaged over the probe matrices and over steps 1–10, before any run collapses.

One step from one state. From 9 saved AdamW states we took eight directions, RMS-matched to the Adam step, and measured the KL from the old policy after one step (registered; Appendix E.5): it spans a factor 48.5–765.5 at matched RMS, 9.2–31.9 at matched operator norm and 2.1–6.4 at matched aligned norm (Figure 10), also in six Qwen3-1.7B states and, on Llama-3.2-1B, at step 2 only. A Haar-random basis at the same aligned norm moves the policy 743.1–4112.2 times less, against our registered expectation: the aligned norm bounds gradient-derived directions only.

Thresholds, and what is left. Over training the same quantity orders the collapses. With collapse within 60 steps logistic in a step size (shared slope, one offset per direction), the directions' half-probability values span a factor 10 in learning rate, 3.4 in operator norm and 1.4 in aligned norm (Table 26 and Figure 7), and it orders the time to collapse best (Table 32). Per-direction offsets still improve that pooled fit ($p = 0.007$), polar tolerating the smallest aligned norm (1.43×10^{-4} at probability $1/2$ against 2.02×10^{-4} for AdamW), and held out over all 9 families the aligned norm is no better than the learning rate or the operator norm (log-loss 0.71 against 0.68 and 0.66; Tables 26 and 28). It orders; it is not a sufficient statistic.

Matched per matrix, and matched in what the policy does. Those comparisons reach a common aligned norm through the learning rate, which also moves the Adam-updated non-matrix branch. Matching instead every matrix step to the aligned norm of its own Adam step, with every parameter at AdamW's learning rate, removes that confound (Table 15). The aligned norm, however, under-counts what the flat step does to the policy: inside the runs, before any rescaling, its one-step KL is 1.09–1.88 times the Adam step's over 60 steps and 1.45–1.57 times over 200, so the aligned-matched arm is in fact taking the larger step. Rescaling all matrix steps at every step so that this measured one-step KL equals that of the Adam step from the same state ("policy-matched"; the factor is 0.62–0.89, and the residual mismatch 0.96–1.04) removes the bias as well: over the two registered rates, thirty-two seeds per arm and rate and one run per seed (Table 15 also lists relaunched repeats of

some AdamW seeds), the flat direction then collapses in 34 of 64 runs and AdamW in 29 of 64; the rate-stratified difference is $+8$ percentage points (90% interval -4 to $+20$, inside the registered margin of ± 25; a tighter ± 20, registered at twenty-four seeds, was missed there and is met at thirty-two; Newcombe score interval -4 to $+19$; 95% interval -6 to $+22$; 11 cells in all, including 200 steps with the penalty and with the clip, where the two arms cross within a few steps of each other). One caveat bounds what this compares: the Adam-updated non-matrix branch (embeddings, head, norms), identical in both arms, carries 16–79% of AdamW's one-step KL in the 60-step cells and 14–20% over 200 steps, so the direction governs the rest. Freezing that branch, so that only the compared direction moves the policy (registered), leaves the picture unchanged: at 3 and $5{\cdot}10^{-6}$ the aligned-matched arm collapses earlier than AdamW in every seed pair (median steps 45.5 and 26 against 54 and 36), and the policy-matched arm with it (56 and 33; seven of eight runs against eight of eight; Table 15). On MATH, where the aligned-norm match leaves the widest gap (three of three collapse at $3{\cdot}10^{-6}$, AdamW none of three), the policy match leaves one of three; on Llama-3.2-1B, where the aligned-matched arm collapses at steps 6, 17 and 19 against AdamW's 30, 21 and 21, the policy-matched arm collapses at 29, 36 and 21 (registered). Matched in what the policy does, the flat direction is not more robust than AdamW, and we cannot show it is less.

The aligned-norm match itself, which ignores that bias, puts the flat direction behind AdamW almost everywhere: in 14 of 15 settings — three Qwen3 sizes, two tasks, 60 and 200 steps, with and without penalty and clip, and Llama-3.2-1B — it collapses at least as often as AdamW, as does the same arm on Muon's own momentum, and in 11 of the 12 settings whose medians differ its median collapse step is the earlier one (one-sided sign test $p = 0.003$; the settings share models and recipes, and over the 7 groups of model, task and horizon 6 point the same way, $p = 0.062$; the exception is Qwen3-4B at $5{\cdot}10^{-6}$, six runs); at the boundary cell with five seeds it collapses in 5 of 5 runs where AdamW collapses in 5 of 35 (one-sided Fisher $p = 3.8{\cdot}10^{-4}$). Two registered controls locate that gap in the yardstick and not in the direction: shrunk by the constant 0.88 of the saved states ("KL-matched", too little, as the in-run ratios show) it still collapses more often, while shrunk by 0.71, the fitted residual of Table 26, or by 0.50 — the range the policy match actually asks for — it is stable in every run at a rate where AdamW collapses in 5 of 35 (Table 17). For the near-rank-one Pion filter matching is not enough (it still collapses in 2 of 3 runs at $2{\cdot}10^{-6}$): there the residual is the direction.

Predicting an implementation before running it. Registered in advance from its ten-step a/η, the threshold of the public Muon ($\kappa = 0.2$, run only at 10^{-6} until then) was placed at $2.04{\cdot}10^{-5}$: at 0.7 times that rate 1 of 3 seeds collapse, at 1.4 times 3 of 3. For the two nearly rank-one public Pion filters the same rule gets the order but places the threshold too high (Table 37).

A held-out direction. $US^{0.5}V^{\top}$, flatter than AdamW but with a larger aligned norm per unit learning rate, collapses in a registered test where flatness predicted stability (Appendix D).

Accuracy: the headroom buys speed, not an endpoint. At each arm's best stable rate, eight seeds each (registered), the flat direction and AdamW end at the same accuracy at step 60: 75.7% for polar-match at $5{\cdot}10^{-6}$ and 75.6% for AdamW at $2{\cdot}10^{-6}$ (difference $+0.1 \pm 0.6$, one-sided Welch $p = 0.433$), all sixteen runs stable. The flat direction gets there sooner: at step 50, the endpoint of Table 1, it leads by $+1.1 \pm 0.3$ points (75.7% against 74.6%; not registered). Under a practitioner's recipe (200 steps, warmup and cosine decay, KL penalty and clip; registered) neither arm collapses and the two end at 72.3% $\pm$ 0.3 (flat, $n = 8$) and 71.6% $\pm$ 0.2 (AdamW, $n = 8$; difference $+0.6$, 95% interval -0.1 to $+1.3$; Table 30). Public Pion at the best of three rates ($1.5{\cdot}10^{-6}$; registered) ends at 73.5%, 2.1 points below AdamW's best. That is what their displacement predicts: polar-match at its best rate moves the layer outputs slightly less than AdamW at its own, and the accuracy line, fitted beforehand, gives a difference of -0.3 against the observed $+0.1$. Across the 37 runs that stay stable for 60 steps, one line in the aligned norm accounts for the gain (3.7 points per decade, residual spread 1.6); per-direction offsets ($F(3, 32) = 1.29$, $p = 0.30$) and slopes ($p = 0.12$) do not improve it, and every offset is below AdamW's (Figure 8 and Table 31). The line is not exact, though: with new seeds ($n = 5$), polar-match at $5{\cdot}10^{-6}$ sits $+1.9 \pm 0.6$ points above it and public Pion ($\kappa = 5$), at a displacement about seven times below the fitted range, $+2.3 \pm 0.2$ points above it (registered; per-head Pion $+1.0 \pm 1.5$). We therefore do not claim that the Pion filter has no accuracy effect at small steps, only that at its best rate it ends below AdamW's best; the per-head recipe collapses in 3 of its 8 seeds within 60 steps (Table 29).

5.5 SCALE: QWEN3-1.7B

On Qwen3-1.7B (Table 6) the two directions separate as on Qwen3-0.6B, at a twice larger learning rate: at 10^{-5} AdamW collapses in all 3 seeds while polar-match is stable in all 3; both thresholds and both aligned-norm constants are higher than at 0.6B. Registered bracketing runs place AdamW's threshold in $(5, 7]\cdot10^{-6}$ (3 of 5 collapse at $7\cdot10^{-6}$) and polar-match's in $(1, 1.4]\cdot10^{-5}$ (3 of 5 at $1.4\cdot10^{-5}$, where we had predicted at most one of three), a ratio in $(1.4, 2.8]$ (Table 30). Matched at every step in policy change at 10^{-5} the flat direction collapses in two of three runs, AdamW in three of three (registered).

A second family. On Llama-3.2-1B-Instruct the registered grid (up to 10^{-5}) brackets no threshold, so its prediction fails; an exploratory extension finds the ordering one step higher: at $2\cdot10^{-5}$ AdamW collapses in 3 of 3 runs and polar-match in 0 of 3, and at $4\cdot10^{-5}$ both AdamW runs reach the absorbing state by step 4 while one polar-match run of two collapses, at step 54 (a/η: 149 and 36; Table 18).

Qwen3-4B. The registered extrapolation of the half-probability aligned norm from the two smaller models fails (1 of 8 decided runs wrong; Table 25): the two constants differ fourfold, the thresholds do not. With added seeds (registered), AdamW's collapses at 4B are not monotone in the learning rate — two of three at $5\cdot10^{-6}$, 1 of 3 at 10^{-5} (three of five with two later seeds), 3 of 3 at $2\cdot10^{-5}$ — and polar-match collapses in 0 of 3 at 10^{-5}, one of four at $1.5\cdot10^{-5}$ (registered) and 1 of 3 at $2\cdot10^{-5}$. By the rule of Section 5.3 the headroom at 4B is not determined. Matched per matrix in aligned norm, the flat direction is stable in three of three at $5\cdot10^{-6}$ where AdamW collapses in two, the one matched cell pointing the other way, and collapses in three of three at 10^{-5} against AdamW's one of three (registered; Table 15).

5.6 A SECOND TASK AND A LONGER HORIZON

MATH. On MATH (levels 3–5, evaluated on MATH-500; Table 12) the per-direction constants carry over ($a/\eta = 23$ and 64 against 24 and 72) and so does the Q2 ordering: at $5\cdot10^{-6}$ AdamW collapses in 3 of 3 runs while polar-match is stable in 3 of 3, and the edge-truncated direction collapses already at $2\cdot10^{-6}$. The GSM8K relation, not refitted, decides every MATH run as registered (0 of 11 wrong), though learning-rate thresholds do as well (Table 37).

The order of two directions reverses with what is held fixed. With a per-token KL penalty ($\beta = 0.04$) we mark instability by the first step with per-token KL to the initial policy above 1, where entropy also rises (Tables 22 and 23). Over 200 steps at a common $\eta = 5\cdot10^{-6}$ AdamW reaches it at steps 62/66/73 and polar-match at 146/135/147 (Figure 6a); polar-match at $1.5\cdot10^{-5}$, with AdamW's aligned norm, at 50/48/48 (Figure 6b); and the polar direction matched per matrix at AdamW's learning rate at 51/44/48, earlier than every AdamW run (registered; Figure 6c). With a gradient clip of 1.0 added the three arms cross at 58/63/67, 144/149/129 and 50/45/50 (Table 19), and matched in policy change the flat direction crosses with AdamW, with or without the clip (Table 15). A wider range of stable learning rates is not a more robust direction.

5.7 LIMITS

The primary evidence is Qwen3-0.6B on GSM8K over 60 steps at a constant learning rate without penalty or clip; over 200 steps without a penalty every run collapses, polar-match at $5\cdot10^{-6}$ before AdamW at $2\cdot10^{-6}$ (77/65 against 86/83). Table 37 lists what did not hold.

6 CONCLUSION

The learning-rate headroom of an orthogonalized update in RLVR is a rescaling: its step moves the layer outputs less per unit learning rate, and once that is matched — in displacement, where it collapses at least as often as AdamW in 14 of 15 settings, or step by step in policy change, where its 90% interval excludes an advantage of more than four points — no robustness is left (Figure 1 and Table 15). For practice: an RMS-matched orthogonalized step is safe and reaches AdamW's accuracy sooner, but its wider stable range is a rescaled learning rate, not a margin against collapse.

AI USE STATEMENT

Generative AI tools, in the form of a large language model coding assistant, were used in this work for research ideation and experiment design, for searching and summarizing related work, for writing and debugging the training and optimizer code, for orchestrating and monitoring the training runs, for the analysis scripts that turn the logs into every number, table and figure of this paper, for the scripts that numerically verify the analysis, for drafting and editing the text of the paper, and for adversarial review of intermediate drafts against the raw logs. Generative AI tools were not used to generate or modify any experimental result: every number reported here comes from executed code and logged training runs, and the scripts that produce them are released with the paper. No AI-generated data, images or human-subject material is involved, so the remaining categories of required disclosure do not apply. We have reviewed all AI-assisted work: the authors checked the optimizer implementation and the analysis scripts against the definitions in the paper, re-derived the propositions of Section 4 and Appendix F, verified the citations against the cited papers, and confirmed every quoted number against the generated macro files. We take responsibility for the final content of this work, including text, claims and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies optimizer stability in reinforcement learning post-training. The experiments use the public GSM8K (Cobbe et al., 2021) and MATH (Hendrycks et al., 2021) datasets and the openly released Qwen3-0.6B, Qwen3-1.7B and Qwen3-4B models (Yang et al., 2025) and Llama-3.2-1B-Instruct, under their respective licences; no human subjects, personal data or crowdworkers are involved, and no new dataset is released. The trained checkpoints are intermediate artifacts of a stability study and are not released. The results are negative or diagnostic for existing optimizers rather than capability-increasing, and we see no direct path from them to harmful applications beyond those already associated with language-model post-training in general. The study consists of 611 executed training runs (failed and interrupted executions included), each on a single GPU, totalling about 1219 GPU-hours of logged training steps (Appendix D). We have no conflicts of interest to declare.

REPRODUCIBILITY STATEMENT

The optimizer, the GRPO training script, the scripts that generate every quoted number, table and figure from the training logs, and the script that numerically verifies every statement of Section 4 and Appendix F will be released. The supplementary material also contains the step-by-step logs of all 611 executed runs, including failed and interrupted executions (the scripts use the first completed execution of each configuration and seed; directories of runs that failed before their first step are in the archive but not in the count), from which those scripts reproduce every generated table of this paper byte for byte. Appendix A contains all proofs, Appendix B the output of the verification script, Appendix C the implementation details of every arm, Appendix D the full training and evaluation protocol, and Appendix F the specification of the gate probe. The code is provided as anonymized supplementary material.

REFERENCES

Florent Benaych-Georges and Raj Rao Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012. arXiv:1103.2221.

Pierfrancesco Beneventano, Mahmoud Abdelmoneum, and Tomaso Poggio. The spectral dynamics and noise geometry of Muon. *arXiv preprint arXiv:2606.08388*, 2026.

Jeremy Bernstein and Laker Newhouse. Old optimizer, new norm: An anthology. *arXiv preprint arXiv:2409.20325*, 2024.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John

Schulman. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Yihe Dong and Will Sawin. Muon^p: Muon with fractional spectral powers. *arXiv preprint arXiv:2606.13867*, 2026.

Zhehang Du and Weijie Su. The Newton-Muon optimizer. *arXiv preprint arXiv:2604.01472*, 2026.

Chongyu Fan, Gaowen Liu, Mingyi Hong, Ramana Rao Kompella, and Sijia Liu. Rethinking Muon beyond pretraining: Spectral failures and high-pass remedies for VLA and RLVR. *arXiv preprint arXiv:2605.19282*, 2026.

Oleg Filatov, Jiangtao Wang, Jan Ebert, and Stefan Kesselheim. Optimal scaling needs optimal norm. *arXiv preprint arXiv:2510.03871*, 2025.

Dan Hendrycks, Collin Burns, Saurabh Kundu, Akul Arora, Steven Basart, Eric Tang, Dawn Sun, and Jacob Steinhardt. Measuring mathematical problem solving with the MATH dataset. In *Advances in Neural Information Processing Systems Datasets and Benchmarks Track*, 2021.

Keller Jordan, Yuchen Jin, Vlado Boza, Jiacheng You, Franz Cesista, Laker Newhouse, and Jeremy Bernstein. Muon: An optimizer for hidden layers in neural networks. Blog post, 2024. URL `https://kellerjordan.github.io/posts/muon/`.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. arXiv:1412.6980.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with PagedAttention. In *Proceedings of the 29th Symposium on Operating Systems Principles*, 2023. arXiv:2309.06180.

Tim Large, Yang Liu, Minyoung Huh, Hyojin Bahng, Phillip Isola, and Jeremy Bernstein. Scalable optimization in the modular norm. *arXiv preprint arXiv:2405.14813*, 2024.

Xianliang Li, Zihan Zhang, Weiyang Liu, and Han Bao. Denoise first, orthogonalize later: Understanding momentum in Muon via spectral filtering. *arXiv preprint arXiv:2606.03899*, 2026.

Zichong Li, Liming Liu, Chen Liang, Weizhu Chen, and Tuo Zhao. NorMuon: Making Muon more efficient and scalable. *arXiv preprint arXiv:2510.05491*, 2025.

Hunter Lightman, Vijay Korthikanti, 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*, 2024.

Jingyuan Liu, Jianlin Su, Xingcheng Yao, Zhejun Jiang, Guokun Lai, Yulun Du, Yidao Qin, Weixin Xu, Enzhe Lu, Junjie Yan, et al. Muon is scalable for LLM training. *arXiv preprint arXiv:2502.16982*, 2025a.

Zichen Liu, Changyu Chen, Wenjun Li, Penghui Qi, Tianyu Pang, Chao Du, Wee Sun Lee, and Min Lin. Understanding R1-Zero-like training: A critical perspective. *arXiv preprint arXiv:2503.20783*, 2025b.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. arXiv:1711.05101.

Thomas Massena, Corentin Friedrich, and Mathieu Serrurier. From SGD to Muon: Adaptive optimization via Schatten-p norms. *arXiv preprint arXiv:2605.19781*, 2026.

Sam McCandlish, Jared Kaplan, Dario Amodei, and OpenAI Dota Team. An empirical model of large-batch training. *arXiv preprint arXiv:1812.06162*, 2018.

Sagnik Mukherjee, Lifan Yuan, Dilek Hakkani-Tur, and Hao Peng. Reinforcement learning finetunes small subnetworks in large language models. In *Advances in Neural Information Processing Systems*, 2025. arXiv:2505.11711.

Tien-Phat Nguyen, Truong Nguyen, Minh-Phuc Truong, Tuc Nguyen, James Bailey, and Trung Le. Spectral flattening is all Muon needs: How orthogonalization controls learning rate and convergence. *arXiv preprint arXiv:2605.13079*, 2026.

Kai Ruan, Jinghao Lin, Zihe Huang, Ziqi Zhou, Qianshan Wei, Xuan Wang, and Hao Sun. When does Muon help agentic reinforcement learning? *arXiv preprint arXiv:2607.16169*, 2026.

Rulin Shao, Shuyue Stella Li, Rui Xin, Scott Geng, Yiping Wang, Sewoong Oh, Simon Shaolei Du, Nathan Lambert, Sewon Min, Ranjay Krishna, Yulia Tsvetkov, Hannaneh Hajishirzi, Pang Wei Koh, and Luke Zettlemoyer. Spurious rewards: Rethinking training signals in RLVR. *arXiv preprint arXiv:2506.10947*, 2025.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Tianqi Shen, Jinji Yang, Runze Shi, Jianhao Ma, Jiaye Teng, and Ziye Ma. Towards understanding the power and limits of the Muon optimizer: A river-valley perspective. *arXiv preprint arXiv:2606.21514*, 2026.

Chongjie Si, Debing Zhang, and Wei Shen. AdaMuon: Adaptive Muon optimizer. *arXiv preprint arXiv:2507.11005*, 2025.

Ziqing Wen, Zhouyang Liu, Jiahuan Wang, Ping Luo, Li Shen, Dongsheng Li, and Tao Sun. Revealing modular gradient noise imbalance in LLMs: Calibrating Adam via signal-to-noise ratio. *arXiv preprint arXiv:2605.05794*, 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.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Lingjun Liu, et al. DAPO: An open-source LLM reinforcement learning system at scale. *arXiv preprint arXiv:2503.14476*, 2025.

Minxin Zhang, Yuxuan Liu, and Hayden Schaeffer. Adam improves Muon: Adaptive moment estimation with orthogonalized momentum. *arXiv preprint arXiv:2602.17080*, 2026.

Hanqing Zhu, Zhenyu Zhang, Hanxian Huang, DiJia Su, Zechun Liu, Jiawei Zhao, Igor Fedorov, Hamed Pirsiavash, Zhizhou Sha, Jinwon Lee, David Z. Pan, Zhangyang Wang, Yuandong Tian, and Kai Sheng Tai. The path not taken: RLVR provably learns off the principals. *arXiv preprint arXiv:2511.08567*, 2025.

A PROOFS

Throughout, $\langle X, Y\rangle = \operatorname{tr}(X^\top Y)$ and $\|\cdot\|_F$ is the induced norm.

A.1 PROOF OF LEMMA 1

Unrolling the recursion gives $m_t = (1-\beta)\sum_{i=1}^{t}\beta^{t-i}G_i$, hence $\omega_{t,i} = (1-\beta)\beta^{t-i}/(1-\beta^t)$ and $\sum_i \omega_{t,i} = (1-\beta)\frac{1-\beta^t}{1-\beta}\cdot\frac{1}{1-\beta^t} = 1$. Using $1-\beta^2 = (1-\beta)(1+\beta)$ and $1-\beta^{2t} = (1-\beta^t)(1+\beta^t)$,

$$\sum_{i=1}^{t}\omega_{t,i}^2 = \frac{(1-\beta)^2}{(1-\beta^t)^2}\sum_{j=0}^{t-1}\beta^{2j} = \frac{(1-\beta)^2(1-\beta^{2t})}{(1-\beta^t)^2(1-\beta^2)} = \frac{1-\beta}{1+\beta}\cdot\frac{1+\beta^t}{1-\beta^t}.$$

The map $x \mapsto (1+x)/(1-x)$ is increasing on $[0,1)$ and β^t decreases in t, so w_t decreases. We have $w_1 = 1$ and $w_t \to (1-\beta)/(1+\beta)$. If the ξ_i are uncorrelated with covariance Σ, then $\operatorname{Cov}(\sum_i \omega_{t,i}\xi_i) = \sum_{i,j}\omega_{t,i}\omega_{t,j}\operatorname{Cov}(\xi_i,\xi_j) = w_t\Sigma$. □

A.2 PROOF OF COROLLARY 1

Let $\operatorname{vec}\xi_{i,k} \sim \mathcal{N}(0,\Sigma)$ be independent over steps i and micro-batches k. The noise of G_i is $\xi_i = \frac{1}{K}\sum_k \xi_{i,k} \sim \mathcal{N}(0,\Sigma/K)$, independent across i. By Lemma 1 and closure of Gaussians under linear maps, $\sum_i \omega_{t,i}\xi_i \sim \mathcal{N}(0, w_t\Sigma/K)$. Since $\hat{g}_{t,1}$ and $\hat{g}_{t,2}$ have the same conditional mean, $\Delta_t = (\xi_{t,1} - \xi_{t,2})/\sqrt{2} \sim \mathcal{N}(0,\Sigma)$, so $\sqrt{w_t/K}\,\Delta_t \sim \mathcal{N}(0, w_t\Sigma/K)$. The two random matrices therefore have the same distribution, and so do their top singular values, because $\sigma_{\max}$ is a measurable function. The statement concerns marginal distributions only; the two matrices are dependent, since both contain $\xi_{t,1}$ and $\xi_{t,2}$. Without Gaussianity, the first two moments still agree. □

A.3 PROOF OF PROPOSITION 1 AND REMARK 3

Matrix form. Write $X = \sum_i s_i u_i v_i^\top$. Then $XX^\top = \sum_i s_i^2 u_i u_i^\top$ and $(2.5XX^\top - 1.5(XX^\top)^2)X = \sum_i (2.5s_i^2 - 1.5s_i^4)s_i u_i v_i^\top = \sum_i f_s(s_i)u_i v_i^\top$. Induction gives P_{k_s}; transposing a tall matrix changes neither singular values nor this identity. The implementation divides by $\|M\|_F + 10^{-7}$, which multiplies every s_i by the same factor below one. Since f_s^k is nondecreasing on $[0,1]$, this can only lower the output spectrum.

(a) $f_s'(s) = 7.5s^2(1-s^2) \geq 0$ on $[0,1]$, $f_s(0) = 0$ and $f_s(1) = 1$, so $f_s([0,1]) \subseteq [0,1]$. With $u = s^2$, $f_s(s) - s = -1.5\,s\,(u-1)(u-\frac{2}{3})$, so the fixed points in $[0,1]$ are 0, $s_\star$ and 1. Moreover $f_s'(0) = f_s'(1) = 0$ and $f_s'(s_\star) = 7.5\cdot\frac{2}{3}\cdot\frac{1}{3} = \frac{5}{3}$.

(b) For $s \in (0,1]$ write $f_s(s) = s\,h(s)$ with $h(s) = 2.5s^2 - 1.5s^4 > 0$. Since $h'(s) = s(5-6s^2)$, h increases on $(0,\sqrt{5/6})$ and decreases on $(\sqrt{5/6},1]$; with $h(s_\star) = h(1) = 1$ this gives $h < 1$ on $(0,s_\star)$ and $h > 1$ on $(s_\star,1)$. If $s_0 \in (0,s_\star)$, induction shows $s_j \in (0,s_0]$ and $s_{j+1} = s_j h(s_j) \leq h(s_0)\,s_j$, so $s_j \leq h(s_0)^j s_0 \to 0$ and the sequence decreases strictly. In addition $f_s(s) \leq 2.5s^3$, so $r_j = \sqrt{5/2}\,s_j$ satisfies $r_{j+1} \leq \sqrt{5/2}\cdot\frac{5}{2}s_j^3 = r_j^3$, hence $r_j \leq r_0^{3^j}$, which is the stated bound. If $s_0 \in (s_\star,1)$, then $s_{j+1} = s_j h(s_j) > s_j$, and $s_{j+1} = f_s(s_j) < f_s(1) = 1$ because f_s is strictly increasing on $(0,1)$. The sequence is increasing and bounded, so it converges to a fixed point in $(s_\star,1]$, which must be 1.

(c) If $s_1, s_2 > s_\star$ then $s_1^2 + s_2^2 > 4/3 > 1 = \sum_i s_i^2$, a contradiction. If $s_1^2 < 2/3$, all $s_i < s_\star$ and every term vanishes by (b); if $s_1^2 > 2/3$, the first term tends to $u_1 v_1^\top$ and all others vanish. If $s_1^2 = 2/3$, then $s_1 = s_\star$ is a fixed point, so the first term equals $s_\star u_1 v_1^\top$ for every k_s, while $s_i^2 \leq 1/3 < 2/3$ for $i \geq 2$ and all other terms vanish. For the norm bound: if $s_1 \leq s_\star$, then $f_s^k(s_i) \leq s_i$ for all i and $\|P_{k_s}\|_F^2 \leq \sum_i s_i^2 = 1$. Otherwise $f_s^k(s_1) \leq 1$ and $f_s^k(s_i) \leq s_i$ for $i \geq 2$, so $\|P_{k_s}\|_F^2 \leq 1 + (1-s_1^2) < 4/3$.

(d) $\operatorname{RMS}(\kappa\sqrt{d_{\max}}P) = \kappa\sqrt{d_{\max}}\|P\|_F/\sqrt{d_1 d_2} = \kappa\|P\|_F/\sqrt{d_{\min}}$. A full-rank polar factor has $\|O\|_F = \sqrt{d_{\min}}$, which gives $\operatorname{RMS} = \kappa$.

Promotion. $f_p'(s) = 1.875(1-s^2)^2 \geq 0$ and $f_p(1) = 1$, so f_p maps $[0,1]$ onto itself and is strictly increasing. Write $f_p(s) = s\,h_p(s)$ with $h_p(s) = 1.875 - 1.25s^2 + 0.375s^4$. Then $h_p'(s) = s(1.5s^2 - 2.5) < 0$ on $(0,1]$ and $h_p(1) = 1$, so $h_p > 1$ on $[0,1)$: $f_p(s) > s$ on $(0,1)$, the only fixed points are 0 and 1, and $f_p'(0) = 15/8$. Since $f_p^{k_p}$ is a continuous increasing bijection of $[0,1]$, τ_{k_p} is well defined. Part (b) applies to each value $f_p^{k_p}(s_i) \in [0,1]$ separately (it does not use $\sum_i s_i^2 = 1$), which gives the survival criterion: $f_s^k(f_p^{k_p}(s_i))$ tends to 0 if $f_p^{k_p}(s_i) < s_\star$, stays at $s_\star$ if equality holds, and tends to 1 otherwise. After promotion the values no longer have unit ℓ_2 norm, so several of them can survive. □

Statement. The main text uses the following facts about one step of a fixed magnitude.

Proposition 2 (Spectral concentration at a fixed magnitude)**.** *Let $D \in \mathbb{R}^{d_1 \times d_2}$, $D \neq 0$, $\|D\|_F = F$, and let $r_s(D) = F^2/\|D\|_2^2$ be its stable rank.* ***(a)*** *$1 \leq r_s(D) \leq \operatorname{rank}(D) \leq d_{\min}$, i.e., $F d_{\min}^{-1/2} \leq \|D\|_2 \leq F$, with equality on the left iff all $d_{\min}$ singular values are equal (e.g., a polar factor) and on the right iff $\operatorname{rank}(D) = 1$.* ***(b)*** *If $D = cU_kV_k^\top$ with $U_k^\top U_k = V_k^\top V_k = I_k$ (e.g., N_t), then $\|D\|_2 = Fk^{-1/2}$; at equal* RMS *this is $(d_{\min}/k)^{1/2}$ times the polar factor's (32 for $d_{\min} = 1024$, $k = 1$).* ***(c)*** *For $y = Wx$ and $W \leftarrow W - \eta D$, $\max_{\|x\|=1} \|\eta D x\| = \eta\|D\|_2$, attained at the top right singular vector, and $\|\eta D X^\top\|_F \leq \eta\|D\|_2\|X\|_F$ for inputs X in rows, whereas $\mathbb{E}\|\eta D x\|^2 = \eta^2 F^2\,\mathbb{E}\|x\|^2/d_2$ for isotropic random x.* ***(d)*** *$\tilde{D} = (\|A\|_2/\|D\|_2)D$ equalizes the bounds in (c) across directions, and $\|\tilde{D}\|_F = \|A\|_2\, r_s(D)^{1/2}$.*

A.4 Proof of Proposition 2

Let $\sigma_1 \geq \cdots \geq \sigma_{d_{\min}} \geq 0$ be the singular values of D, $\varrho = \operatorname{rank}(D)$, and v_1 a top right singular vector, so $\|D\|_2 = \sigma_1 > 0$.

(a) $F^2 = \sum_{i \leq \varrho} \sigma_i^2$ lies between σ_1^2 and $\varrho\,\sigma_1^2$, so $1 \leq r_s(D) \leq \varrho \leq d_{\min}$; dividing F^2 by these bounds and taking square roots gives $F/\sqrt{d_{\min}} \leq \sigma_1 \leq F$. The right bound holds with equality iff $\sigma_i = 0$ for all $i \geq 2$, i.e., $\varrho = 1$. The left bound holds with equality iff $r_s(D) = d_{\min}$. This forces $\varrho = d_{\min}$ and $\sigma_i = \sigma_1$ for all $i \leq d_{\min}$; conversely, equal singular values give $F^2 = d_{\min}\sigma_1^2$. The polar factor $UV^\top$ of a full-rank matrix has all $d_{\min}$ singular values equal to one.

(b) Up to the sign of c, $cU_kV_k^\top$ is a thin SVD with k singular values equal to $|c|$. Hence $F = |c|\sqrt{k}$ and $\|D\|_2 = |c| = F/\sqrt{k}$. The polar factor rescaled to the same Frobenius norm is the case $k = d_{\min}$, which gives the ratio $\sqrt{d_{\min}/k}$. At RMS ρ, $F = \rho\sqrt{d_1 d_2}$, so the two operator norms are $\rho\sqrt{d_1 d_2/k}$ and $\rho\sqrt{d_{\max}}$. More generally, by (a) every direction of rank k has $\|D\|_2 \geq F/\sqrt{k}$, with equality iff it has the form $cU_kV_k^\top$. The polar arm uses NS_5 with Muon's coefficients, whose singular values lie in a band around one rather than at one; its stable rank, and hence the ratio, is then smaller (Appendix B).

(c) For every x, $\|Dx\|^2 = x^\top D^\top D x \leq \sigma_1^2\|x\|^2$, with equality iff x lies in the eigenspace of $D^\top D$ for the eigenvalue σ_1^2, in particular for $x = v_1$. Summing over the rows $x_i^\top$ of X gives $\|DX^\top\|_F^2 = \sum_i \|Dx_i\|^2 \leq \sigma_1^2\|X\|_F^2$, with equality iff every row lies in that eigenspace. If $\mathbb{E}\|x\|^2 < \infty$, linearity of trace and expectation gives $\mathbb{E}\|Dx\|^2 = \mathbb{E}\operatorname{tr}(Dxx^\top D^\top) = \operatorname{tr}(D\Sigma D^\top) = \langle D^\top D, \Sigma\rangle$. For isotropic x, $\Sigma = (\mathbb{E}\|x\|^2/d_2)I$ and this is $F^2\mathbb{E}\|x\|^2/d_2$, because $\operatorname{tr}(D^\top D) = F^2$; multiplying by η^2 gives the statement. For inputs aligned with the top right singular vector, $\Sigma = \tau^2 v_1 v_1^\top$, it is instead $\tau^2\|Dv_1\|^2 = \sigma_1^2\operatorname{tr}(\Sigma)$, so at equal F and $\mathbb{E}\|x\|^2$ the aligned and isotropic root mean squares differ by the factor $\sqrt{d_2/r_s(D)}$, which ranges up to $\sqrt{d_2}$. More generally, for positive semidefinite P with eigendecomposition $P = \sum_i p_i w_i w_i^\top$, $\langle P, \Sigma\rangle = \sum_i p_i\, w_i^\top \Sigma w_i$ lies between $\lambda_{\min}(\Sigma)\operatorname{tr} P$ and $\|\Sigma\|_2 \operatorname{tr} P$. Applying this with $P = D^\top D$, and with the roles of $D^\top D$ and Σ exchanged, gives

$$\lambda_{\min}(\Sigma)\,F^2 \;\leq\; \mathbb{E}\|Dx\|^2 \;\leq\; \min\left\{\|\Sigma\|_2\,F^2,\; \|D\|_2^2\operatorname{tr}\Sigma\right\}.$$

If all eigenvalues of Σ lie within a factor γ of each other, then $\mathbb{E}\|Dx\|^2$ lies within a factor γ of $F^2\operatorname{tr}(\Sigma)/d_2$ for every direction. Only an input distribution concentrated on few directions separates directions of equal F, by up to the factor $\sqrt{d_2/r_s(D)}$ in the root mean square.

(d) r_s is invariant under scaling, so $\|\tilde{D}\|_F^2 = r_s(\tilde{D})\|\tilde{D}\|_2^2 = r_s(D)\|A\|_2^2$. With $\|\tilde{D}\|_2 = \|A\|_2$, part (c) gives $\max_{\|x\|=1} \|\eta\tilde{D}x\| = \eta\|A\|_2$ and $\|\eta\tilde{D}X^\top\|_F \le \eta\|A\|_2\|X\|_F$ for every direction. By (a), $\|A\|_2 \le \|\tilde{D}\|_F \le \sqrt{d_{\min}}\,\|A\|_2$, with the lower value for rank-one directions and the upper value for flat spectra. Relative to RMS matching (Equation (3)), the step is multiplied by $\sqrt{r_s(D)/r_s(A)}$. □

Remark 1 (Scope of Proposition 2 and the aligned regime)**.** Proposition 2 concerns one linear map and one step. It does not bound the change of the loss, of the KL divergence to the rollout policy, or of the training dynamics over many steps, and we do not claim a stability theorem. It isolates a quantity that differs across directions at a fixed RMS. By (c), whether this quantity matters depends on the inputs, and two facts point to the aligned regime. First, a batch of tokens contributes $G = \sum_i \delta_i x_i^\top$ to the gradient of W, with backpropagated errors δ_i. The right singular vectors of G lie in $\mathrm{span}\{x_i\}$. If the δ_i are mutually orthogonal, then $G^\top G = \sum_i \|\delta_i\|^2 x_i x_i^\top$ is an error-weighted second moment of the inputs, whose top eigenvectors are the top right singular vectors of G. The momentum is a weighted sum of such gradients. A truncated direction with small k_t keeps only the leading singular directions of the momentum, so its row space points at input directions that have carried large error-weighted energy. If the input distribution changes slowly, these are directions in which the input second moment Σ of the proof of (c) is concentrated. Second, the hidden activations of trained transformers are far from isotropic, with a few directions carrying much of the energy. The training logs record the RMS, the operator norm and the stable rank of the applied step for probe matrices, and the operator-norm-matched arms equalize $\|\tilde{D}_t\|_2$. The operator-norm-matched arms test this: at equal per-step operator norm, the polar direction with its larger RMS collapses before AdamW and the concentrated directions collapse before every AdamW seed or at steps that bracket AdamW's (Section 5.4), so the operator norm of a step alone does not determine stability.

Remark 2 (Pion at a fixed magnitude)**.** Let M, s_i and $P_{k_s}(M)$ be as in Proposition 1, with $M \neq 0$. (i) For all i and k_s, $f_s^{k_s}(s_i)/f_s^{k_s}(s_1) \le s_i/s_1$. Hence $r_s(P_{k_s}(M)) \le r_s(P_{k_s-1}(M)) \le \cdots \le r_s(M)$, and $r_s(P_{k_s}(M)) \to 1$ as $k_s \to \infty$ if $s_1^2 > 2/3$ or $s_1 > s_2$; in the latter case $P_{k_s}(M)/\|P_{k_s}(M)\|_F \to u_1 v_1^\top$. (ii) Rescaled to Frobenius norm F, the Pion direction has operator norm $F/\sqrt{r_s(P_{k_s}(M))} \ge F/\sqrt{r_s(M)}$. This is Proposition 2(b) with the effective number of directions $r_s(P_{k_s}(M))$ in place of k, so at a fixed RMS the magnitude-matched Pion arm concentrates its step. (iii) Under the convention $\tilde{D} = \kappa\sqrt{d_{\max}}\,P_{k_s}(M)$, $\|\tilde{D}\|_2 = \kappa\sqrt{d_{\max}}\,f_s^{k_s}(s_1) \le \kappa\sqrt{d_{\max}}$, which is the operator norm of the same convention applied to the exact polar factor, and $\mathrm{RMS}(\tilde{D}) \le \kappa\sqrt{4/3}/\sqrt{d_{\min}}$ by Proposition 1(d). The public-code Pion step is therefore small in both norms.

Proof. Because f_s is nondecreasing on $[0,1]$ and $f_s(s) = s\,h(s) > 0$ for $s \in (0,1]$, the singular values of $P_j = P_j(M)$ are $f_s^j(s_i)$ in the original order, $\|P_j\|_2 = f_s^j(s_1) > 0$, and $r_s(P_j) = \sum_i \big(f_s^j(s_i)/f_s^j(s_1)\big)^2$. We first show that $0 < x \le y \le 1$ and $x^2 + y^2 \le 1$ imply $h(x) \le h(y)$ and $f_s(x)^2 + f_s(y)^2 \le 1$. If $y \le \sqrt{5/6}$, then $h(x) \le h(y)$ because h is increasing on $[0, \sqrt{5/6}]$ (proof of Proposition 1(b)). Otherwise $x^2 \le 1 - y^2 < 1/6$, so $h(x) \le h(1/\sqrt{6}) = 3/8 < 1 \le h(y)$, since $h \ge 1$ on $[s_\star, 1]$. For the second inequality, let $g(u) = u^3(5/2 - 3u/2)^2$, so that $f_s(s)^2 = g(s^2)$ and g is nondecreasing on $[0,1]$. With $u = y^2$ and $x^2 \le 1 - u$,

$$f_s(x)^2 + f_s(y)^2 \le g(1-u) + g(u) = 1 - \tfrac{15}{4}\,u^2(1-u)^2 \le 1.$$

Now fix $i \ge 2$ with $s_i > 0$, and let $x_j = f_s^j(s_i)$ and $y_j = f_s^j(s_1)$. Since $s_i^2 + s_1^2 \le \sum_l s_l^2 = 1$, induction shows that every pair (x_j, y_j) satisfies the hypotheses, and $x_{j+1}/y_{j+1} = (x_j/y_j)\,h(x_j)/h(y_j) \le x_j/y_j$. Summing the squares over i gives (i). If $s_1^2 > 2/3$, then $P_{k_s}(M) \to u_1 v_1^\top$ by Proposition 1(c), and r_s is continuous at nonzero matrices, so $r_s(P_{k_s}(M)) \to 1$. If $s_1 > s_2$ and $s_1^2 \le 2/3$, then $y_j \le s_1$ for all j and $x_j \to 0$, since $f_s(s) < s$ on $(0, \sqrt{2/3})$ and $x_0 < \sqrt{2/3}$. If $y_j \to y > 0$, the ratio x_j/y_j tends to 0. Otherwise $y_j \to 0$, and $h(x_j)/h(y_j) = (x_j/y_j)^2\,(5/2 - 3x_j^2/2)/(5/2 - 3y_j^2/2)$ gives $x_{j+1}/y_{j+1} \le c_j\,(x_j/y_j)^3$ with $c_j \to 1$; as $x_0/y_0 < 1$, again $x_j/y_j \to 0$. Hence every singular-value ratio of P_j to its largest tends to 0, which gives $r_s(P_j) \to 1$ and $P_j/\|P_j\|_F \to u_1 v_1^\top$. Part (ii) is the definition of r_s combined with (i). Part (iii) follows from $f_s^{k_s}(s_1) \le 1$ and from Proposition 1(d). □

B NUMERICAL VERIFICATION OF THE ANALYSIS

The script `theory/verify_props.py` checks every statement of Section 4 and Appendix F.4 on CPU, in part against the optimizer implementation itself. It runs in about one minute:

```
CUDA_VISIBLE_DEVICES="" python3 theory/verify_props.py
```

With PyTorch 2.9.1, NumPy 2.2.6 and Python 3.13.9, all 67 checks pass; the full output is in `theory/verify_output.txt`. We summarize it below.

Proposition 1 (suppression collapse). The fixed points, the derivative $f_s'(s_\star) = 5/3$ and the monotonicity of f_s are confirmed in 120-digit arithmetic. Table 2 lists the iterates and the doubly exponential bound. The monotone bound $s_j \le h(s_0)^j s_0$ and the bound of part (b) hold at every listed s_0. Over 20,000 random unit-norm spectra, at most one value exceeds $s_\star$, and the largest value of $\|P_5\|_F^2$ is $1.0000 \le 4/3$. Table 3 applies the implemented filter (`pion_filter`, $k_p = 0$, $k_s = 5$) to 256×768 matrices with a prescribed top energy share. The thresholds of Remark 3 are $\tau_0 = 0.8165$, $\tau_1 = 0.5230$, $\tau_2 = 0.2957$ and $\tau_3 = 0.1605$.

Table 2: Iterates $s_j = f_s^j(s_0)$ (exact arithmetic) and the bound $B_5 = \sqrt{2/5}\,(\sqrt{5/2}\,s_0)^{3^5}$, which applies for $s_0 < \sqrt{2/5}$.

s_0	share s_0^2	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	B_5
0.80	0.640	0.788	0.768	0.732	0.666	0.542	–
0.75	0.563	0.699	0.603	0.429	0.175	0.0132	–
0.70	0.490	0.605	0.433	0.180	0.0143	7.2×10^{-6}	–
0.60	0.360	0.423	0.169	0.0119	4.2×10^{-6}	1.9×10^{-16}	1.7×10^{-6}
0.50	0.250	0.266	0.0449	2.3×10^{-4}	2.9×10^{-11}	5.9×10^{-32}	1.0×10^{-25}
0.40	0.160	0.145	7.5×10^{-3}	1.0×10^{-6}	2.8×10^{-18}	5.7×10^{-53}	2.8×10^{-49}
0.30	0.090	0.0639	6.5×10^{-4}	6.8×10^{-10}	8.0×10^{-28}	1.3×10^{-81}	1.2×10^{-79}
0.10	0.010	2.5×10^{-3}	3.8×10^{-8}	1.4×10^{-22}	7.0×10^{-66}	8.7×10^{-196}	1.4×10^{-195}

Table 3: Implemented Pion filter ($k_p = 0$, $k_s = 5$) on 256×768 matrices: Frobenius norm and number of nonzero entries (out of 196,608) of the output in bfloat16, float32 and float64, and RMS of the $\kappa\sqrt{d_{\max}}$ update relative to κ (Muon gives 1; the bound of Proposition 1(d) is 0.0722).

top share	$\|O\|_F$ bf16	nnz	$\|O\|_F$ f32	nnz	$\|O\|_F$ f64	$f_s^5(s_1)$	RMS $/\kappa$
0.95	1.000	196608	1.000	196608	1.000	1.000	6.3×10^{-2}
0.70	0.970	196608	0.970	196608	0.970	0.970	6.1×10^{-2}
0.60	0.135	196608	0.136	196608	0.136	0.136	8.5×10^{-3}
0.50	2.6×10^{-5}	196608	2.6×10^{-5}	196608	2.6×10^{-5}	2.6×10^{-5}	1.6×10^{-6}
0.30	6.2×10^{-24}	196608	6.2×10^{-24}	196608	6.2×10^{-24}	6.2×10^{-24}	3.9×10^{-25}
0.20	0	0	2.3×10^{-43}	2270	2.8×10^{-42}	2.8×10^{-42}	1.8×10^{-43}
0.10	0	0	0	0	2.7×10^{-76}	2.7×10^{-76}	1.7×10^{-77}
0.05	0	0	0	0	1.0×10^{-111}	1.0×10^{-111}	6.2×10^{-113}

Lemma 1 and Corollary 1 (noise weight and edge). For $\beta \in \{0.9, 0.95\}$ and $t \in \{1, 2, 3, 5, 10, 20, 40, 80, 500\}$, the closed form matches $\sum_i \omega_{t,i}^2$ to 10^{-12}, and a Monte Carlo estimate with 2×10^5 draws lies within four standard errors. Examples for $\beta = 0.9$: $w_{10} = 0.1090$ (Monte Carlo 0.1092 ± 0.0003), 2.07 times the asymptotic value, and $w_{40} = 0.0542$, 1.03 times the asymptotic value. For 128×384 Gaussian noise with $K = 8$ and $\beta = 0.9$, the ratio of the top singular value of the momentum noise to $\sigma_{\max}(\Delta_t)\sqrt{w_t/K}$ averages 0.989, 0.992, 1.000 and 1.005 at $t = 1, 3, 10, 40$ over 10 repetitions each, with extremes 0.954 and 1.033. Single-sample fluctuations are therefore comparable to the safety factor $c_{\text{edge}} = 1.05$.

Proposition 2 and Remark 2 (operator norm at a fixed magnitude). Norms are computed by exact float64 SVDs. We test 28 matrices of shapes 64×64, 48×144, 144×48 and 256×768 with flat, near-flat, rank-one, rank-4 (equal and unequal values), heavy-tailed ($s_i = 1/i$) and Gaussian spectra. On all of them the bounds of part (a) hold, $r_s = d_{\min}$ only for the flat spectrum, and $r_s = 1$ only for rank one. The exact polar factor of a Gaussian matrix has $r_s = d_{\min}$. NS_5 with Muon's coefficients is not exactly flat: for the rectangular Gaussian matrices its singular values lie in $[0.68, 1.13]$ and $r_s/d_{\min}$ is between 0.57 and 0.71. For part (b), the truncated polar factor of a 1024×3072 Gaussian matrix at a common RMS has operator norm $F/\sqrt{k}$, and its ratio to the polar factor is 2, 8, 16 and 32 for $k = 256$, 16, 4 and 1, to 10^{-9}. For part (c) (128×384, $\|D\|_F = 1$), $\|Dv_1\| = \|D\|_2$ and none of $20{,}000$ random unit inputs exceeds it. The batch bound holds, with equality for rows along v_1. For every spectrum, the isotropic mean $\mathbb{E}\|Dx\|^2 = F^2\mathbb{E}\|x\|^2/d_2$ agrees with a Monte Carlo estimate from 4×10^4 draws. The ratio of the aligned to the isotropic root mean square equals $\sqrt{d_2/r_s(D)}$, from 1.73 (flat) to 19.6 (rank one). The two-sided bound on $\mathbb{E}\|Dx\|^2$ in the proof holds for 35 random low-rank-plus-ridge covariances. For part (d), operator-norm matching to a synthetic Adam direction with $r_s(A) = 40$ gives $\|\tilde{D}\|_F = \|A\|_2\sqrt{r_s(D)}$. This is 0.16 (rank one) to 1.79 (flat) times the RMS-matched step. For Remark 2, the polynomial identity and the pairwise inequalities (10^6 random pairs) hold, and $r_s(P_j(M))$ is nonincreasing for $j \le 5$ over $20{,}000$ random spectra. For the implemented float64 filter on 256×768 matrices, $r_s(P_5(M)) = 1.000$ at every top share from 0.3 to 0.9. The magnitude-matched Pion step therefore has $\sqrt{d_{\min}} = 16$ times the operator norm of the polar factor, whereas $\|\kappa\sqrt{d_{\max}}P_5\|_2/(\kappa\sqrt{d_{\max}}) = f_s^5(s_1)$ falls from 1 to 6×10^{-24}.

Implementation. `DirOpt` runs for 12 steps on a 96×288 matrix with $K = 4$ micro-batches of a rank-4 signal plus anisotropic noise, with the same gradient stream for every arm. For `magnitude="spec"`, the implementation estimates both operator norms with 50 float32 power iterations from the all-ones vector. The exact operator norm of the applied step is within 0.82% of $\|A_t\|_2$ for the polar arm and within 2×10^{-7} for the truncated arm, and the estimate of $\|A_t\|_2$ is within 0.03%. In this stream $k_t \in \{1, 2\}$, except for $k_t = 0$ at the first two steps, where the step falls back to A_t. Under `magnitude="match"`, both arms have $\mathrm{RMS}(\tilde{D}_t) = \mathrm{RMS}(A_t)$, and the truncated step has 7.5 ($k_t = 1$) or 5.3 ($k_t = 2$) times the operator norm of the polar step. This equals $\sqrt{r_s/k_t}$ for the NS_5 polar step, whose stable rank is $0.587\, d_{\min}$, so the ratio is $0.77\sqrt{d_{\min}/k_t}$ rather than $\sqrt{d_{\min}/k_t}$ (9.8 and 6.9).

C Implementation details

Cost and memory. Per matrix and step, NCSH adds one thin SVD of $\hat{m}_t$ ($O(d_1 d_2 d_{\min})$), J matrix–vector products with Δ_t and its transpose ($O(Jd_1d_2)$), and a few inner products. A randomized low-rank SVD with rank cap k_{cap} is available as an option. Beyond Adam's two moments, memory grows by one gradient-sized buffer, used only during the first two micro-batches, and by the factors (U_k, V_k) of $\hat{N}_t$. When $k(d_1 + d_2) \ge d_1 d_2$ we store the dense $\hat{N}_t$ instead, so the extra persistent memory never exceeds one parameter-sized tensor per matrix. On CPU, one step for a 1024×3072 matrix takes 0.84 s with a full SVD and 0.09 s with the low-rank SVD, compared with 0.06 s for AdamW. The polar direction adds five Newton–Schulz iterations per matrix, and operator-norm matching adds 50 power iterations for each of the two operator norms. Table 10 reports wall-clock time per GRPO step, its training part (forward, backward and optimizer step, excluding rollout generation and weight reloading) and peak trainer memory on Qwen3-0.6B. Runs were distributed over several GPU models, so times differ across machines by more than across optimizers and are indicative only. The peak memory of polar-match equals that of AdamW up to 0.1 GB (14.5 vs. 14.6 GB), and the edge-truncated arm, which keeps the SVD factors, adds about 1.5 GB. On Qwen3-1.7B at $\eta = 5{\cdot}10^{-6}$, the training part takes 81 s for polar-match and 81 s for AdamW, with peak memory 34.8 and 34.8 GB.

Routing. A parameter enters the matrix branch if it is two-dimensional and its name contains one of `q_proj`, `k_proj`, `v_proj`, `o_proj`, `gate_proj`, `up_proj` or `down_proj`. All other parameters receive AdamW with the same hyperparameters. Tied tensors (the output head and the input embedding of Qwen3-0.6B/1.7B) are optimized once.

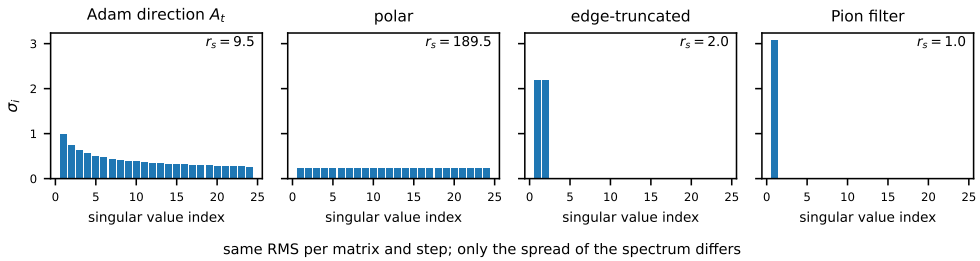


Figure 2: The four directions after the rescaling of Equation (3), computed with the maps used in training on a momentum matrix whose stable rank matches the measured Adam step. All four carry the same RMS and differ only in how it is spread, hence in operator norm (0.22, 2.18 and 3.08 times AdamW's). For this spectrum the Pion filter falls below the floating-point floor, as Proposition 1 predicts, so its panel shows the normalized direction.

Micro-batch protocol and noise sample. Gradients are zeroed at the start of every step. Each micro-batch loss is divided by the constant $NL_{\max}$, so micro-batch k adds $\hat{g}_{t,k}/K$ to the gradient buffer. After the first backward pass the optimizer copies the buffer ($= \hat{g}_{t,1}/K$). After the second, it forms $2 \cdot \text{copy} - \text{buffer} = (\hat{g}_{t,1} - \hat{g}_{t,2})/K$ in place, stores its top singular value as a scalar, and frees the copy. Since $\sigma_{\max}(\Delta_t) = K\,\sigma_{\max}\big((\hat{g}_{t,1} - \hat{g}_{t,2})/K\big)/\sqrt{2}$, no further memory is needed. Later micro-batches only increment a counter. If $K < 2$, no edge is available and the truncated and gated arms use A_t.

Power iteration. We run J iterations of $u \leftarrow \Delta v/\|\Delta v\|$, $v \leftarrow \Delta^\top u$, $\hat{\sigma} \leftarrow \|v\|$, $v \leftarrow v/\hat{\sigma}$, starting from a fixed, seeded unit vector for each matrix shape. The estimate is a lower bound on $\sigma_{\max}$. With $J = 20$ it underestimates the top singular value of Gaussian matrices by about 1.5–2% on average, which is why all runs use $J = 50$.

Truncation and storage. The SVD of $\hat{m}_t$ uses `torch.linalg.svd` in float32. An optional randomized low-rank SVD (`torch.svd_lowrank` with $q = k_{\text{cap}} + 10$ and two subspace iterations, seeded per step without disturbing global random streams) caps k_t at $k_{\text{cap}} = 256$ and logs when the cap binds. $\hat{N}_t$ is stored as the factors (U_k, V_k), or as a dense matrix when $k(d_1 + d_2) \geq d_1 d_2$. $\hat{A}_{t-1}$ is not stored; it is recomputed from the moments before they are updated. The cross-fitted estimate is $\hat{a}_t = \langle \hat{N}_{t-1}, G_t \rangle = \operatorname{tr}(U_k^\top G_t V_k)/\sqrt{k}$, computed from the stored factors. When $k_t = 0$, $\hat{N}_t$ is discarded, so no cross-fitted estimate is made at step $t + 1$, and the EMAs keep their previous values.

Arms. The polar arm uses five quintic Newton–Schulz steps with coefficients $(3.4445, -4.7750, 2.0315)$ on the Frobenius-normalized input ($+10^{-7}$), transposed to the wide orientation. The Pion arm applies k_p steps of $X \leftarrow 1.875X + (-1.25XX^\top + 0.375(XX^\top)^2)X$ followed by k_s steps of $X \leftarrow (2.5XX^\top - 1.5(XX^\top)^2)X$ ($k_p = 0$, $k_s = 5$). This is our reimplementation of the singular-value maps of Fan et al. (2026); we check that its output equals $Uf(S)V^\top$ in float64. Under matched magnitude, the filter is evaluated in float64 and normalized. A bfloat16 evaluation is used only to log whether the public precision would return a null update, and a float64 output of norm zero falls back to A_t. Under the κ convention, the filter runs in float32. For all matched arms, a zero direction falls back to A_t so that the step size is preserved; in particular, the Pion arm falls back to A_t if its float64 direction still vanishes. Under operator-norm matching (Section 3.2), $\|A_t\|_2$ and $\|D_t\|_2$ are estimated with 50 float32 power iterations from the all-ones vector (Appendix B).

Logged diagnostics. For every matrix and step we log σ_t, $\sigma_t^\star$, k_t, e_t, $\sigma_{\max}(\Delta_t)$, $\bar{a}_t$, $\bar{b}_t$, ρ_t, $\operatorname{RMS}(A_t)$ and $\operatorname{RMS}(\tilde{D}_t)$, together with the flags *no edge*, $k = 0$, *cap reached*, *too few estimates*, *Adam fallback* and *null bfloat16 Pion output*.

Differences from the public Muon/Pion implementation. Our shared-moment arms deliberately differ from the public implementation of Fan et al. (2026), so that only one variable

changes at a time: (i) *momentum*: an EMA with $\beta = 0.95$ without bias correction (Nesterov off by default) there, versus the bias-corrected Adam momentum here; (ii) *precision*: Newton–Schulz and Pion iterations in bfloat16 there, versus float32 (float64 for the matched Pion direction) here; (iii) *magnitude*: $\kappa\sqrt{d_{\max}}$ with $\kappa = 5$ by default in the AdamW-fused classes there (`RL/verl/verl/utils/muon.py`, commit `8796501`: `scale_factor: float = 5.0` in `MuonAdamW`, `DefaultPionAdamW` and `PerHeadMuonAdamW`, lines 349, 438 and 535; the functional helpers of the same file default to 0.2). The released RLVR scripts use $\kappa = 5$: `examples/grpo_trainer/run_qwen3-4b_fsdp2_muon.sh` selects `MuonAdamW` without overriding the scale, so the class default applies, and `examples/gmpo_trainer/run_qwen3-1.7b_fsdp2_perhead_pion.sh` sets `scale_factor=5.0` explicitly (line 70). Our polar arm, by contrast, uses Equation (3); (iv) *routing*: when not configured, the public classes send every parameter with at least two dimensions to the orthogonalized branch, including embeddings and the output head. In our public-code arms we pass explicit parameter groups, so exactly the projection matrices listed above are orthogonalized.

Assumptions of the noise estimate in GRPO. Completions are shuffled across prompts before being split into micro-batches. Micro-batches 1 and 2 hold 16 completions each out of 1024 (128 prompts $\times$ 8 completions), so on average $16 \cdot 16 \cdot 7/1023 \approx 1.75$ pairs across the two micro-batches share a prompt and a group baseline. The independence in Assumption 1 therefore holds only approximately. The implementation also assumes unsharded tensors; sharded training would need to gather each matrix before the SVD, the power iteration and the inner products.

D Experimental details

Arms. Table 5 lists every arm, its momentum, precision and magnitude rule, and the learning rates tested on GSM8K with Qwen3-0.6B.

Endpoints. Every run is evaluated at step 0 and every 10 steps; the common accuracy endpoint is step 50. A run that meets the collapse-stop criterion is stopped and evaluated at that step; step 50 remains its endpoint if it reached step 50, and the stopping evaluation is the endpoint only for runs stopped earlier, since the collapsed state (near-zero reward with almost all responses at the length limit) is absorbing on Qwen3-0.6B. Runs launched before the stop rule was introduced were not stopped; some of them collapsed and continued to step 60 or to the end of their log, so their final checkpoint lies long after the collapse (Table 11). Some early runs have logs that end at step 58 because a prompt-indexing error in the training script terminated them during step 59; this happened after their step-50 evaluation, so their endpoint is unaffected, their stability window covers steps 49–58 as for all other runs, and they have no final checkpoint. The collapse step reported in Tables 1 and 11 is defined separately in Section 5.1 and does not depend on whether a run was stopped. One operator-norm-matched polar run had been stopped by hand before the automatic rule applied; it is excluded; the run reported in Table 1 is its rerun under the automatic rule. An early duplicate of one gate-probe seed, launched before the stop rule existed, ended at step 49 without a step-50 evaluation and is excluded as well; it collapsed at step 24 rather than 14, so the collapse step is not determined by the seed alone.

Evaluation noise. All 84 step-0 evaluations of Qwen3-0.6B use identical pretrained weights, greedy decoding and a fixed sampling seed. Their accuracy ranges from 63.0 to 65.0% with a standard deviation of 0.47 points, and two of them differ on 72 of the 1319 problems on average. Repeated evaluations on the same machine vary as much as evaluations on different machines, so computing accuracy changes within a run on one GPU does not remove this noise. Each accuracy change is the difference of two evaluations and carries the noise of both.

Policy entropy. The logged policy entropy is the mean per-token entropy of the trainer's distribution over the completions of the first micro-batch of each step (16 of 1024 completions).

Drift measurement. For Table 11, we load the checkpoint saved at the last step T of a run, in bfloat16, and the pretrained weights, and compute for every attention and MLP projection matrix of

Table 4: Training and evaluation protocol of the GSM8K runs. The MATH runs (Section 5.6) differ in data, answer check and $K = 128$ micro-batches of 8; the 200-step runs evaluate every 20 steps. The Qwen3-4B runs use $K = 512$ micro-batches of 2 (memory), so their aligned norm and entropy, taken from the first micro-batch, rest on 2 sequences instead of 16, which may contribute to the fourfold a/η at 4B (Table 25); the Llama-3.2-1B policy-matched runs use micro-batches of 2 and measure the one-step KL on the first 8 (16 sequences, against 32 elsewhere). The gradient is the same sum over the same 1024 completions in every case.

Setting	Value
Models	Qwen3-0.6B (all arms); Qwen3-1.7B (public-code Muon and per-head Pion with $\kappa = 5$, AdamW, polar-match)
Training data	GSM8K train (MATH: levels 3–5 of the MATH training set (Hendrycks et al., 2021)); prompt order determined by the seed
Prompt	chat template, thinking disabled, "put your final answer within \boxed{}"
Reward	1 if the final answer matches the reference, else 0 (no format reward)
Prompts per step / completions per prompt	128 / 8
Sampling	temperature 1.0, top-p 1.0, max. 1024 new tokens, vLLM (Kwon et al., 2023)
Advantages	$(r_i - \bar{r})/(\mathrm{std}(r) + 10^{-6})$ within each group, sample standard deviation
Loss normalizer	constant $NL_{\max} = 1024 \times 1024$
Importance weights	truncated at $C = 2$; one optimizer step per rollout batch
KL penalty / entropy bonus / gradient clipping	none (k3, $\beta = 0.04$, in Tables 22 and 23) / none / none
Micro-batches	$K = 64$, 16 completions each
Precision	float32 master weights, bfloat16 autocast, gradient checkpointing
Weight sync	bfloat16 weights reloaded into the inference server after every step
Learning rate / schedule	grid of Table 5 / constant
Weight decay / (β_1, β_2) / ϵ	0.01 (decoupled) / $(0.9, 0.999)$ / 10^{-8}
Steps	60 (200 for the runs of Table 13)
Edge-truncated polar and gate	$c_{\mathrm{edge}} = 1.05$, $J = 50$, $\beta_s = 0.8$, $n_{\min} = 2$, full SVD (float32)
Operator-norm matching	50 float32 power iterations per operator norm
Public-code Muon/Pion	$\kappa \in \{5, 0.2\}$, momentum 0.95, 5 iterations, $k_p = 0$ (Pion)
Collapse stop	from step 15: 5 consecutive steps with reward ≤ 0.02 and $\geq 90\%$ truncated
Evaluation	GSM8K test (1319) or MATH-500 (Lightman et al., 2024), greedy, every 10 steps and at the stopping step
Answer extraction	last \boxed{} expression; commas, \$, % and \text{} removed; first number compared with the reference to relative tolerance 10^{-6}
Seeds	0, 1, 2 in the scans of Table 1; further seeds (up to 31) where a registered test adds them, as stated with each test
Software	PyTorch 2.9.1, Transformers 4.57.6, vLLM 0.20.2
Hardware	NVIDIA RTX 5090 (Qwen3-0.6B runs, five machines); NVIDIA RTX PRO 6000 Blackwell (Qwen3-1.7B runs, the 200-step runs and one held-out-direction run)

the first, middle and last transformer layer the operator norm (exact SVD, float32), the Frobenius norm and the stable rank of $W_T - W_0$; the table reports means over these matrices. Rounding to the bfloat16 grid can lower the operator norm of the drift by a factor of up to 3.4 (see *Cumulative drift per run* below). Newer runs also log these quantities in float32 during training, which gives the drift at the collapse step and the curves of Figure 5. The trainer–sampler log-probability gap is the mean absolute difference between the log-probabilities of the sampled tokens under the trainer (float32 master weights, bfloat16 autocast) and under the bfloat16 inference engine, over all tokens of a step.

Evaluation noise. Greedy vLLM decoding is not deterministic: 84 evaluations of the same pre-trained Qwen3-0.6B weights range from 63.0 to 65.0% (standard deviation 0.47 points). With the seed-to-seed spread, differences below about 2.5 points between single-seed arms are not resolved; differences between collapsed and stable runs, of tens of points, are unaffected.

Table 5: Arms and tested learning rates on Qwen3-0.6B. "Shared $\hat{m}$": the bias-corrected Adam momentum shared by all arms of our implementation; "EMA 0.95": the momentum of the public Pion implementation. Magnitude: κ convention (update $= \kappa\sqrt{d_{\max}}\cdot$direction), RMS matching (Equation (3)) or operator-norm matching (Section 3.2). On Qwen3-1.7B we run public-code Muon ($\kappa = 5$) and per-head Pion ($\kappa = 5$) at 10^{-6}, and AdamW and RMS-matched polar at 10^{-6}, $5{\cdot}10^{-6}$, 10^{-5} and $2{\cdot}10^{-5}$. The gate probe is specified in Appendix F.

Arm	Momentum	Direction	Magnitude	η
AdamW	Adam	A_t	Adam	1, 2, 2.5, 3, 5 ($\times 10^{-6}$)
Muon (public code)	EMA 0.95	NS_5 (bfloat16)	$\kappa \in \{5, 0.2\}$	10^{-6}
Pion (public code)	EMA 0.95	f_s^5 (bfloat16)	$\kappa = 5$	10^{-6}
Polar	shared $\hat{m}$	NS_5 (float32)	RMS; operator norm	1, 2, 3, 5, 6, 7, 10 ($\times 10^{-6}$); $5{\cdot}10^{-6}$
Pion filter	shared $\hat{m}$	f_s^5 (float64)	RMS; operator norm	0.3, 0.5, 0.7, 1, 5 ($\times 10^{-6}$); $5{\cdot}10^{-6}$
Edge-truncated polar	shared $\hat{m}$	$U_{k_t}V_{k_t}^\top$, else A_t	RMS; operator norm	0.3, 0.5, 0.7, 1, 5 ($\times 10^{-6}$); $5{\cdot}10^{-6}$
$US^pV^\top$ (held out)	shared $\hat{m}$	exact SVD, $p \in \{0.5, 0.25\}$	RMS	$p = 0.5$: 1.2, 2.4, 4; $p = 0.25$: 4, 6 ($\times 10^{-6}$)

Table 6: Accuracy and collapse on Qwen3-1.7B, columns as in Table 1 (Section 5.5).

Arm	η	n	Δacc	peak Δacc	collapse step
Muon (public), $\kappa = 5$	10^{-6}	1	2.7	3.9	57
Per-head Pion (public), $\kappa = 5$	10^{-6}	1	2.6	3.1	none
AdamW	10^{-6}	1	1.0	1.8	none
AdamW	$5{\cdot}10^{-6}$	2	5.1 ± 0.8	5.8	none
Polar, RMS matched	10^{-6}	1	0.7	1.4	none
Polar, RMS matched	$5{\cdot}10^{-6}$	2	3.1 ± 0.5	3.7	none
AdamW	10^{-5}	3	-68.3 ± 14.4	2.8	41/44/41
Polar, RMS matched	10^{-5}	3	3.8 ± 0.2	4.0	none
AdamW	$2{\cdot}10^{-5}$	3	-69.9 ± 13.2	0.0	43/18/51
Polar, RMS matched	$2{\cdot}10^{-5}$	3	-54.2 ± 29.0	2.1	37/34/none

Run-level reproducibility of the collapse step. Rollout generation is not deterministic, so two runs of the same configuration and seed are not identical. Table 8 lists the 30 configurations that were run twice (a run relaunched on another machine, or rerun after the activation probe was added). The collapse step differs by a median of 6 steps and by up to 24, which is larger than the seed-to-seed spread reported in Table 1; the collapse steps in the tables should therefore be read with that resolution. The stable/collapsed outcome is the same for all repeats in all but 4 of the 30 groups; in those 4 groups the collapsing repeats collapse at step 52 or later, so the flips sit at the end of the 60-step window, at learning rates next to the stability threshold. Outcome labels at such boundary settings are therefore themselves noisy, which is one more reason the run-level separation tests of Section 5.4 are reported as failing rather than tuned. Runs that log the aligned norm were launched after the probe was added, so Figure 11 and Table 1 need not come from the same execution of a configuration.

Compute. Every run uses one GPU (Table 4), with the trainer and the inference server sharing it. Summing the logged per-step wall-clock time of every collected run gives 1219 GPU-hours over 611 runs; this excludes the periodic evaluations, the model loading and the runs that were discarded before analysis.

Additional results. Table 10 lists, for the main Qwen3-0.6B arms, the accuracy change at step 60 for runs that reached it, the per-step update RMS and operator norm of the probe matrices ($U = -(W_{t+1} - (1 - \eta\lambda)W_t)/\eta$, averaged over the probe matrices and over the steps before collapse), the fraction of matrix-steps in which the edge-truncated arm keeps no singular direction and falls back to A_t, and the cost per step. Per-step update norms were logged only for runs launched after this logging was added ("–" otherwise).

Cumulative drift per run. Table 11 lists the drift of every run with a final checkpoint, computed from bfloat16 checkpoints. Qwen3 weights are stored in bfloat16, so per-element changes below the rounding step of a weight vanish in the checkpoint; in the 60 runs that also log float32 drift at the

Table 7: Operator-norm-matched arms on Qwen3-0.6B (Section 3.2), columns as in Table 1; every step has the operator norm of the AdamW step at $5{\cdot}10^{-6}$, whose own collapse steps are 35/34/32 (Section 5.4).

Arm	η	n	Δacc	peak Δacc	collapse step
Edge-truncated polar	$5{\cdot}10^{-6}$	2	-64.4 ± 0.5	6.8	29/21
Pion filter	$5{\cdot}10^{-6}$	2	-63.4 ± 0.4	10.2	19/36
Polar	$5{\cdot}10^{-6}$	2	-62.1 ± 0.2	3.5	17/11

Table 8: Configurations run twice with the same seed: the collapse step of each repeat, and their spread. "none": no collapse within the run. Repeats are grouped by model, arm, learning rate and seed; Llama runs are left out, being scored under their own registered rule.

Arm	model	seed	collapse steps of the repeats	spread
Pion filter, op.-norm-matched (5e-6)	0.6B	0	43/19	24
Gate probe (5e-6)	0.6B	0	36/14	22
Edge-truncated, RMS-matched	0.6B	2	53/35	18
Polar, op.-norm-matched (5e-6)	0.6B	0	17/29	12
AdamW (5e-6)	0.6B	0	35/38/44	9
Edge-truncated, op.-norm-matched (5e-6)	0.6B	0	21/29	8
Pion filter, RMS-matched	0.6B	0	42/36	6
Pion filter, RMS-matched (5e-6)	0.6B	0	11/17/17	6
Edge-truncated, RMS-matched	0.6B	1	51/57	6
Fixed gate 0.5	0.6B	0	59/55	4
AdamW (3e-6)	0.6B	0	56/60	4
Polar-match (1e-5)	0.6B	0	39/42	3
Edge-truncated, RMS-matched (5e-6)	0.6B	0	17/16/19	3
Muon (public code, κ=5)	0.6B	0	44/42	2
Polar-match (7e-6)	0.6B	2	53/52	1
Polar-match	0.6B	0	none/none	–
Polar-match (5e-6)	0.6B	0	none/none/none	–
Polar-match (7e-6)	0.6B	0	57/none	–
Polar-match (7e-6)	0.6B	1	54/54	0
Edge-truncated, RMS-matched	0.6B	0	57/none/52	–
Muon (public code, κ=0.2)	0.6B	0	none/none/none	–
Per-head Pion (public code, κ=5)	0.6B	0	none/none	–
Pion (public code, κ=5)	0.6B	0	none/none/none	–
AdamW	0.6B	0	none/none	–
AdamW (2.5e-6)	0.6B	0	none/none	–
AdamW (2.5e-6)	0.6B	1	none/none	–
AdamW (2.5e-6)	0.6B	2	none/58	–
AdamW (2e-6)	0.6B	0	none/none/none	–
AdamW (3e-6)	0.6B	1	52/52	0
AdamW (3e-6)	0.6B	2	56/none	–

final step, the bfloat16 operator-norm drift is 0.29–1.05 times the float32 value, lowest for AdamW at $2{\cdot}10^{-6}$. The table is therefore not used to compare runs (Section 5.4).

MATH and 200-step runs. Tables 12 and 13 list every finished run of Section 5.6. "aligned norm": the input-aligned norm of the step over steps 1–10 (mean over the runs of the row); "predicted": the registered GSM8K prediction per run (C: collapse, S: stable, ?: inside the band between probability 0.1 and 0.9); "observed": collapse within the horizon. MATH-500 has 500 problems, so its greedy evaluation is noisier than GSM8K's; step-0 accuracy ranges over a few points across runs.

Two hundred steps, in full. Stability within 60 steps does not carry over to longer training (Table 13). Without the penalty every 200-step run collapses: AdamW at $2{\cdot}10^{-6}$ at steps 86/83, polar-match at $6{\cdot}10^{-6}$ at step 78 and at $5{\cdot}10^{-6}$ at steps 77/65, although the last has the smallest aligned norm of the three. We had registered that polar-match at $5{\cdot}10^{-6}$ would outlast AdamW at $2{\cdot}10^{-6}$; in

Table 9: The remaining learning rates of the Qwen3-0.6B scan, in the format of Table 1.

Arm	η	n	Δacc	Δacc_{60}	peak Δacc	collapse step
AdamW	$3{\cdot}10^{-6}$	3	7.9 ± 1.1	-58.1	10.8	56/52/56
Polar	$2{\cdot}10^{-6}$	3	6.8 ± 0.5	8.0	8.0	none
Polar	$3{\cdot}10^{-6}$	3	8.2 ± 0.5	9.2	9.1	none
Polar	$7{\cdot}10^{-6}$	3	8.1 ± 1.1	-38.5	11.1	57/54/53
Polar	10^{-5}	3	-63.2 ± 0.7	–	9.9	39/43/39

Table 10: Additional statistics on Qwen3-0.6B (means over runs). n_{60}: runs evaluated at step 60; Δacc_{60}: change of test accuracy from step 0 to step 60 (points); RMS and $\|U\|_2$: per-step update of the probe matrices before collapse; fallback: percentage of matrix-steps of the edge-truncated arm that use A_t; time: wall-clock seconds per step (total and training part); memory: peak GPU memory of the trainer (GB).

Arm	η	n_{60}	Δacc_{60}	RMS	$\|U\|_2$	fallback (%)	time (s)	train (s)	memory (GB)
AdamW	10^{-6}	2	6.8	0.304	156	–	72	53	14.6
AdamW	$2{\cdot}10^{-6}$	3	12.0	0.305	156	–	92	68	14.6
AdamW	$5{\cdot}10^{-6}$	1	-63.8	0.381	217	–	106	76	14.6
Muon (public code, $\kappa = 5$)	10^{-6}	0	–	4.133	276	–	113	78	13.0
Pion (public code, $\kappa = 5$)	10^{-6}	2	9.4	0.031	43	–	81	60	13.0
Polar-match	10^{-6}	2	4.2	0.303	19	–	64	47	14.6
Polar-match	$5{\cdot}10^{-6}$	3	11.7	0.306	20	–	95	69	14.5
Edge-truncated	10^{-6}	2	-32.3	0.354	380	26	111	84	16.2
Edge-truncated	$5{\cdot}10^{-6}$	1	-64.4	0.488	461	39	111	83	16.1
Pion filter	$5{\cdot}10^{-6}$	1	-63.8	0.535	783	–	148	94	14.6

both seeds it collapses first. Over this horizon the aligned norm does not order collapse times across directions, and the thresholds of Section 5.3 are thresholds for a 60-step horizon.

A held-out direction, in full. Stable rank is a competing explanation, which $US^{0.5}V^\top$ separates from the aligned norm: it is flatter than AdamW (stable rank 24.6 against 11.2) but takes a larger aligned norm (101 against 72 per unit learning rate); both predictions were registered before its runs. At $2.4{\cdot}10^{-6}$ it collapses in 3 of 3 runs, as the aligned norm predicted and against flatness; at $1.2{\cdot}10^{-6}$ it collapses in 2 of 2 where the aligned norm predicted stability (Fisher $p = 0.018$ against AdamW at nearby rates): the ordering holds, the level is too optimistic (Tables 21 and 24).

E Additional analysis and discussion

E.1 Promotion steps and relative thresholds

Remark 3 (Promotion steps and relative thresholds)**.** The promotion map $f_p(s) = 1.875s - 1.25s^3 + 0.375s^5$ is nondecreasing on $[0, 1]$, satisfies $f_p(s) \geq s$, and has fixed points 0 (repelling, $f_p'(0) = 15/8$) and 1. With k_p promotion steps before suppression, a singular value does not vanish in the limit $k_s \to \infty$ iff $s_i \geq \tau_{k_p} = (f_p^{k_p})^{-1}(s_\star)$; it tends to one iff $s_i > \tau_{k_p}$ and stays at $s_\star$ if $s_i = \tau_{k_p}$. Numerically, $\tau_0 = 0.816$, $\tau_1 = 0.523$, $\tau_2 = 0.296$ and $\tau_3 = 0.160$. All of these thresholds are relative to $\|M\|_F$. For i.i.d. noise of scale τ, $\|M\|_F \approx \tau\sqrt{d_1 d_2}$, while the top noise singular value is $\approx \tau(\sqrt{d_1} + \sqrt{d_2})$. A 1024×3072 noise matrix therefore has $s_1 \approx 0.05$, and a signal several times above the noise edge can still fall far below τ_{k_p}. The edge in Section 3.2, by contrast, is measured on the absolute scale of the noise.

The proof is given together with that of Proposition 1 in Appendix A.

E.2 Gradient noise and spiked random matrices

The gradient noise scale relates the gradient covariance to the useful batch size (McCandlish et al., 2018). MoLS (Wen et al., 2026) uses module-level signal-to-noise ratios to calibrate Adam. For low-rank signals in rectangular noise, random matrix theory predicts a phase transition: a signal

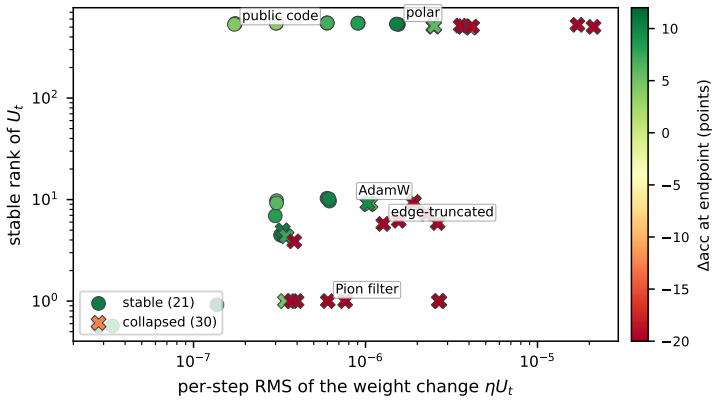


Figure 3: Magnitude and spectrum of the applied step decide stability together. One marker per Qwen3-0.6B run that logs update norms (51 runs): x is the RMS of the per-step weight change ηU_t on the probe matrices, y the stable rank of U_t (1: rank one, $d_{\min}$: flat), both averaged before collapse; colour is the endpoint accuracy change.

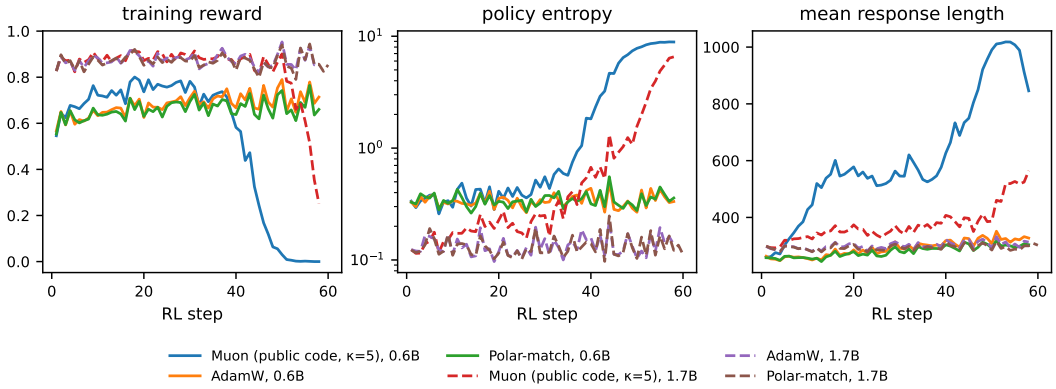


Figure 4: Training reward, policy entropy (log scale) and mean response length at $\eta = 10^{-6}$, seed 0: public-code Muon ($\kappa = 5$), AdamW and polar-match on Qwen3-0.6B (solid) and on Qwen3-1.7B (dashed).

singular value separates from the noise bulk, and its singular vectors become informative, only above a threshold set by the noise level and the aspect ratio (Benaych-Georges & Nadakuditi, 2012). The isotropic noise edge $\tau(\sqrt{d_1} + \sqrt{d_2})$ assumes i.i.d. entries. Gradient noise in language models is strongly anisotropic, so we estimate the edge from an independent noise sample instead (Appendix F.1).

Trainer–sampler log-probability gap. Rollouts are generated by a vLLM copy of the policy in bfloat16, while the gradient is computed from the float32 master weights, so the sampling distribution differs slightly from the one the loss differentiates. We log the mean absolute difference between the trainer and sampler log-probabilities of the sampled tokens. It stays flat in stable runs and rises before the collapse in collapsed runs, together with the drift and the entropy. Truncated importance sampling with cap $C = 2$ corrects the first-order effect of this mismatch, but the runs here cannot separate whether the growing gap contributes to the collapse or only reflects it.

How the drift accumulates. Up to weight decay, $W_T - W_0 = -\eta \sum_{t \le T} \tilde{D}_t$, so $\|W_T - W_0\|_2 \le \sum_{t \le T} \eta \|\tilde{D}_t\|_2$. One might expect concentrated directions, built from the slowly changing leading singular directions of the momentum, to accumulate coherently and approach this bound. The float32 logs do not show this. The ratio of the operator-norm drift to the summed per-step operator norms, at the collapse step or at the end of the run, is 0.50–0.61 for the Pion filter, 0.55–0.65

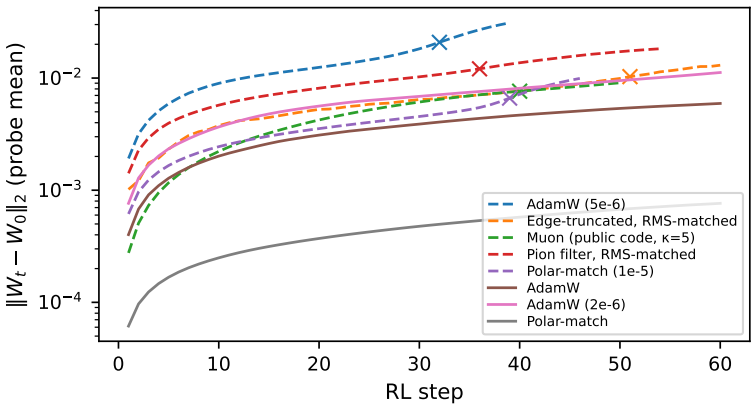


Figure 5: Operator-norm drift $\|W_t - W_0\|_2$ logged in float32 during training (mean over probe matrices) for the RMS-matched and public-code arms of Section 5.4, one seed each; the statistics in the text use all 41 non-gate-probe Qwen3-0.6B runs that log float32 drift. Dashed: collapsed runs, marked at the collapse step. The stable AdamW runs end above the drift at which polar-match at 10^{-5} collapses.

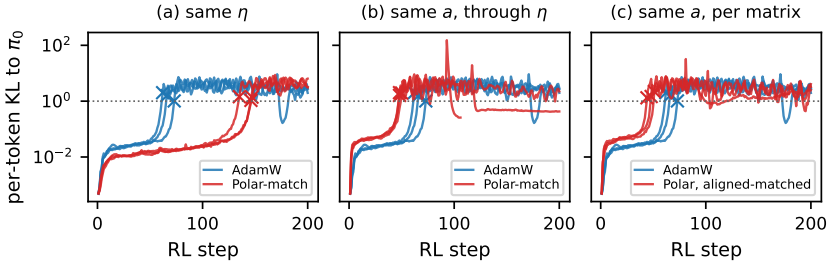


Figure 6: Which direction is more stable depends on what is held fixed (200 steps, GSM8K, Qwen3-0.6B, KL penalty $\beta = 0.04$; per-token KL to the initial policy, one line per run, cross at the first step above the dotted line; runs in Table 23). (a) Both arms at $\eta = 5{\cdot}10^{-6}$, where the polar step carries a third of AdamW's aligned norm: it crosses about twice as late. (b) The same AdamW runs against polar-match at $1.5{\cdot}10^{-5}$, which reaches their aligned norm through the learning rate: the order reverses. (c) The same AdamW runs against the polar direction matched to the Adam step in aligned norm matrix by matrix, at AdamW's own learning rate: the order reverses with the rest of the network untouched.

for AdamW and 0.59–0.72 for the polar direction, and in all collapsed runs the operator-norm drift grows roughly as T^α with α between 0.48 and 0.92. Concentrated directions reach a large operator-norm drift because each of their steps has a large operator norm (Proposition 2), not because successive steps align more. The effect of the spectrum on stability therefore lies in the individual step rather than in how steps add up, and it is not captured by the size of the step or of the drift in either norm.

E.3 Predicting collapse from the size of the step

Table 27 lists, for each RMS-matched direction, the aligned norm per unit learning rate, the learning rate at which the pooled logistic fit of Figure 7b gives collapse probability $1/2$, and the observed fraction of collapsed runs at every learning rate. Under RMS matching, every per-unit-learning-rate statistic of the early steps is nearly constant within a direction (the RMS is the same for all directions, the operator norm varies by a few percent), so this constancy is not specific to the aligned norm; what differs between quantities is how far the directions' constants bring their thresholds together. Table 26 gives the held-out log-loss per direction for a fit on the other families, for the aligned norm, the operator norm, the geometric mean of RMS and operator norm, and the learning rate (which, under RMS matching, is equivalent to the RMS of the step), together with the base

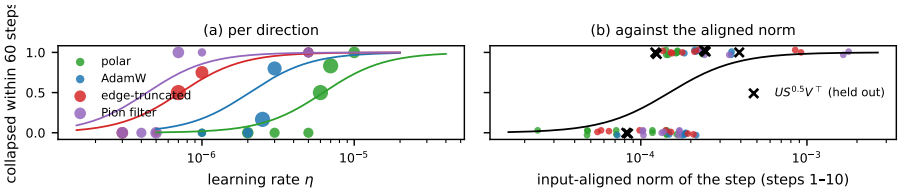


Figure 7: Collapse within 60 steps against the size of the step (Qwen3-0.6B, GSM8K). (a) Fraction of RMS-matched runs that collapse at each learning rate (marker area: number of runs), with one logistic fit in the aligned norm over all runs, evaluated at η times each direction's aligned norm per unit learning rate; deviations at the boundary cells reflect the direction offsets. (b) Every RMS-matched run (jittered) against its aligned norm over steps 1–10, same fit; crosses: the held-out direction $US^{0.5}V^\top$, not used in the fit.

rate, and the per-direction half-probability aligned norms of the fit with direction offsets. Table 28 pools all 86 runs (43 collapsed, 9 direction families, counting the public-code arms, the gate probes and each operator-norm-matched arm as families). Repeated executions of one configuration and seed enter as separate runs; they are independent draws of the nondeterministic rollout and differ in outcome in 4 of 30 groups. The pooled advantage of the aligned norm comes mostly from the near-rank-one families: held out it reaches a log-loss of 0.48 and 0.48 on the Pion filter and the edge-truncated direction against a base rate of 0.70, but only 0.66 on AdamW.

A dimensionless form. The aligned norm is an absolute displacement, averaged over matrices whose outputs differ in scale. In the 47 runs that also log the relative perturbation $\eta\|U_tX\|/\|W_tX\|$, which is dimensionless, the same logistic fit rises from 0.1 to 0.9 within a factor of 36.0 and orders the collapsed runs with $|\rho| = 0.42$. Before seeds 1 and 2 of these runs were launched, we registered the rule that the form separates stable from collapsed runs only if every stable run lies below every collapsed one, together with our expectation that it would not; it does not (largest stable value 1.89×10^{-4}, smallest collapsed value 4.63×10^{-5}), for the same reason as the aligned norm: at polar-match $7{\cdot}10^{-6}$, three seeds share the value to within 0.2% and differ in outcome.

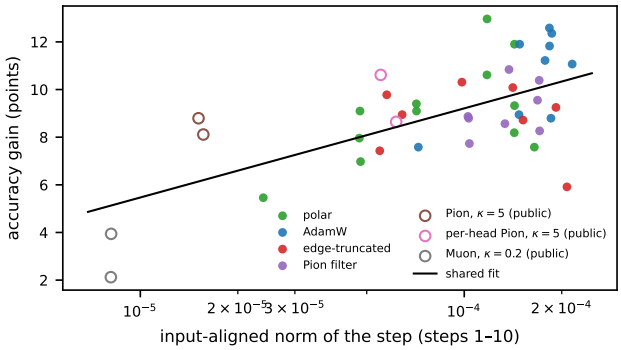


Figure 8: Accuracy gain over 60 steps against the input-aligned norm of the step, for every run that stays stable (Qwen3-0.6B, GSM8K). The line is the fit with no direction term; the four directions interleave along it. Runs that collapse have no end-of-training accuracy and are absent, so the sample is conditioned on stability.

Without the horizon. The binary label of every analysis above is decided at step 60: a run that collapses at step 58 and one that would collapse at step 62 fall on opposite sides of it. To check that the ordering does not depend on that cut, we treat the collapse step as an event time and the runs that never collapse as right-censored, and fit a Cox proportional-hazards model (Breslow ties) with one covariate, the logarithm of a step size measured over steps 1–10, on the 69 main-family runs (32 collapses). Table 32 gives the result: the ordering of the four candidates is the one the binary

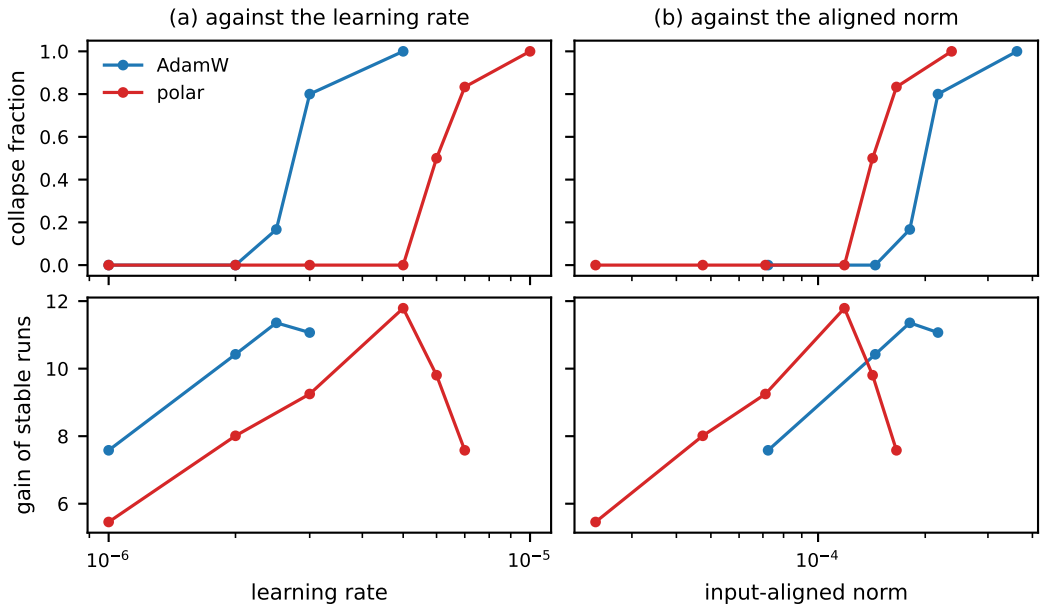


Figure 9: The usable window of AdamW and of the flat polar direction (Qwen3-0.6B, GSM8K, 60 steps, RMS-matched; every run of the main scan): the fraction of runs that collapse and the mean accuracy gain of the stable runs, per learning rate. (a) Against the learning rate the two windows are a factor of about three apart. (b) Against the aligned norm, i.e. the learning rate times each direction's constant a/η, they overlap, and the flat direction's collapses begin slightly earlier.

fits give, and the aligned norm reaches concordance 0.85 against 0.70 for the learning rate alone. This is an in-sample comparison on the same runs, so it tests the horizon and not generalisation; the held-out test is Table 26, where its margin over the base rate is smallest on AdamW (0.66 against 0.70).

E.4 The direction constant at Qwen3-4B

The aligned norm per unit learning rate is nearly constant within a direction on Qwen3-0.6B (coefficient of variation 0.9% for polar-match and 2.1% for AdamW over the scanned rates). On Qwen3-4B it is not: AdamW gives 153 at $2{\cdot}10^{-6}$ and 188 at $2{\cdot}10^{-5}$, a spread of 23%. The Qwen3-4B rates span a factor of 10 against 5 on Qwen3-0.6B, and a step large enough to move the activations within the first ten steps changes the second moment against which the aligned norm is measured. We therefore state the constancy as an approximation over a scanned range, not as a property of a direction; the extrapolated threshold in Section 5.5 uses the constant measured on each Qwen3-4B run itself, so it does not rely on that approximation.

Is the shared line the right shape? Per-direction offsets would absorb curvature if the gain were not linear in the logarithm of the aligned norm, so the non-significant offsets of Figure 8 could in principle hide a non-linear relation. A quadratic term does not improve the fit either ($F(1, 34) = 1.81$, $p = 0.19$), and the directions' aligned-norm ranges overlap rather than tile the axis (polar $2.4{\cdot}10^{-5}$–$1.7{\cdot}10^{-4}$, AdamW $7.2{\cdot}10^{-5}$–$2.2{\cdot}10^{-4}$, edge-truncated $5.5{\cdot}10^{-5}$–$1.9{\cdot}10^{-4}$, Pion filter $1.0{\cdot}10^{-4}$–$1.7{\cdot}10^{-4}$), so the comparison does not rest on extrapolating one direction into another's range.

The trajectory is a symptom, not a warning. The aligned norm is not constant over training: in the ten steps before a collapse it is 1.20 times its steps-1–10 value, while over the last ten steps of a stable run it is 0.58 times it (31 collapsed and 37 stable runs, AUROC 0.93). That comparison uses the collapse step to place its window, so it says only that the quantity is already elevated when a run is failing. The prospective version is a null: among the runs whose outcome is still open at step 20 (28 that later collapse, 37 that do not), the ratio of the aligned norm over steps 11–20 to steps 1–10

does not separate them (AUROC 0.52, $p = 0.75$), whereas on the same runs the steps-1–10 aligned norm itself reaches AUROC 0.84. Watching the quantity change buys nothing over measuring it once at the start, which is also why steps 1–10 is the window we use.

E.5 One step from a shared state

States. Three AdamW runs at $2{\cdot}10^{-6}$ (Qwen3-0.6B, GSM8K, full protocol, seeds 0–2) saved, immediately before the optimizer step at steps 2, 20 and 40, the float32 weights, the gradient of that step, the Adam moments before it, the training batch, and 4 completions sampled from the current policy for each of 64 held-out test prompts. The seed-0 run was stopped by the operating system after writing its step-20 state; its step-40 state comes from a second seed-0 run, whose trajectory is not bit-identical to the first.

Directions. From each state we rebuild the bias-corrected moments exactly as `torch.optim.AdamW` does and compute, for every matrix of the matrix branch: the Adam step A; the momentum $\hat{m}$; the polar arm's $\mathrm{NS}_5(\hat{m})$; the exact flat spectrum in the momentum basis $U_m V_m^\top$ and in the Adam basis $U_A V_A^\top$; $U_m S_m^{0.5} V_m^\top$; the Pion filter (normalized in float64, as the trained arm does); and the rank-one $u_1 v_1^\top$ of $\hat{m}$. The Pion filter coincides with the rank-one direction at every state. Every direction is rescaled to the Frobenius norm of A for that matrix. Only the matrix branch moves, so every direction perturbs the same parameters.

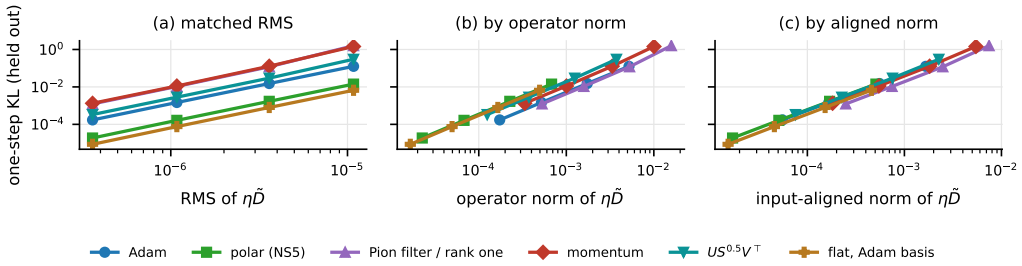


Figure 10: One step from a shared state: KL from the old policy on held-out completions after one step of each direction, against three sizes of the step (geometric mean over 9 saved AdamW states; all directions RMS-matched to the Adam step, matrix branch only). The directions spread over two orders of magnitude at matched RMS (a) and fall onto one curve against the input-aligned norm (c).

Why one step is the easy case. Let δ be the change of the logits caused by one step. The KL from the old policy is $\frac{1}{2}\mathbb{E}[\delta^\top F \delta] + O(\delta^3)$ with F the Fisher matrix of the old policy, and δ is, to first order, a sum over the perturbed matrices of their output displacements propagated through the frozen network. The input-aligned norm is exactly the root-mean-square output displacement of a matrix on the inputs it receives, so a quantity that ignores Σ_t (the RMS, the operator norm) is not expected to account for the one-step KL and one that includes it is. The purpose of this measurement is therefore not to discover that the aligned norm matters at one step, but to check that the eight directions – which differ in spectrum by up to a factor $\sqrt{d_{\min}}$ in operator norm at equal RMS – are not separated by anything else at one step, and to separate the spectrum from the singular vectors under a shared state. Whether the same quantity governs many steps, when the activations, the moments and the policy have all moved, does not follow from this argument; that is what Section 5.4 tests, and where the residuals are.

Measurement. For learning rates 10^{-6}, $3{\cdot}10^{-6}$, 10^{-5} and $3{\cdot}10^{-5}$ we set $W \leftarrow W - \eta D$ and score the first 256 tokens of the 256 held-out completions in float32. The KL from the old policy is the k_3 estimate $\mathbb{E}[e^\delta - 1 - \delta]$, $\delta = \log \pi_{\text{new}} - \log \pi_{\text{old}}$, on tokens sampled from the old policy. Step sizes are measured on the probe matrices, with the aligned norm taken on inputs from the held-out batch. For each step size we fit $\log \mathrm{KL} = c_d + b \log x$ with a shared slope and one offset per direction; the factor $\exp(\max_d c_d - \min_d c_d)$ is the spread of the KL of two directions with the same step size.

A second model family. We repeated the procedure on Llama-3.2-1B-Instruct (three seeds, states at steps 2 and 20; registered, Appendix E.7). Llama rarely writes the required boxed answer at first (accuracy 0% at step 0, 12.5% at step 2), so its early steps mostly learn the format. Table 34 gives the spreads. At step 2 all three states behave as on Qwen3; at step 20 one does, and two do not: in them the largest learning rates move the policy by 10^2 to 10^5 nats per token in one step, and even restricted to KL below 1 (an exploratory analysis we did not register) one state keeps an aligned-norm spread of about 27, the rank-one and momentum directions changing the policy far less than their aligned norm implies. The spectrum-over-rotation ordering holds in all six states.

An arbitrary rotation. The two bases compared above, the singular vectors of the momentum and of the Adam step, are computed from the same momentum. To ask what an arbitrary rotation does we rescored the nine states with two further directions, registered beforehand with the prediction that they would behave like the others (Appendix E.7): $U_r V_r^\top$, a flat spectrum in a Haar-random basis, and $U_r S_A V_r^\top$, the Adam step's spectrum in that basis, with U_r, V_r from the QR of a float64 Gaussian seeded by the parameter's name, and both RMS-matched to the Adam step as every other direction is. Both registered predictions failed, and not narrowly. At the same aligned norm the flat random-basis direction changes the policy by a factor 743.1–4112.2 less than the flat direction in the Adam basis, and the Adam spectrum in the random basis by a factor 14772.1–79580.6 less than the Adam step itself, their first-order surrogate gain is zero to the resolution of the measurement, and changing the spectrum inside the random basis does nothing (a factor 1.0–1.1). With all ten directions in one fit the aligned-norm spread rises from below 10 to 1338.4–7474.7.

Exploratory, not registered (Appendix E.7 records when it was added): writing the one-step KL as the squared aligned norm times a constant, relative to the Adam step's, that constant is 1.18 for the polar direction (range 0.81–1.62 over the nine states), 0.81 for the momentum and 0.42 for the Pion filter; it varies by a factor 13.4 across the eight directions the moments provide and is 1026–5302 times smaller for the random ones. The aligned norm is the input half of the step's effect: it measures the displacement a step produces at a layer's output on the inputs the layer receives, and says nothing about how much the rest of the network responds to a displacement in that direction. Directions built from the gradient produce displacements the network is sensitive to; a random direction, in a space where almost all directions are inert, does not. Within the family an optimizer can actually produce – every step here is a function of the Adam moments – the input half suffices, and that is the claim the paper makes; as a general statement about weight perturbations it is false, which this control shows.

A middle size. Within the Qwen3 family, the one-step procedure was repeated on Qwen3-1.7B (AdamW at $5{\cdot}10^{-6}$, its largest stable rate there, states at steps 2 and 20 of three seeds; registered, Appendix E.7), because the training-time equivalence of the two directions at a matched aligned norm holds at 0.6B and fails at 4B. Table 35 gives the spreads: held (aligned $<$ operator $<$ RMS), held (aligned spread below 10) and held (spectrum above basis) over the 6 states.

E.6 Matching the aligned norm matrix by matrix

The magnitude option `aligned` rescales the step D of every matrix so that $\|DX^\top\|_F/\|X\|_F$, on 256 input rows of that matrix sampled from the first micro-batch of the step, equals the same quantity for the Adam step A of the arm (unit-tested to a relative error below 10^{-5}). The learning rate is AdamW's, so the non-matrix branch takes the AdamW step. The prediction and its two readings were registered before any of these runs (Appendix E.7).

Split learning rate. The option `other_lr` gives the non-matrix branch its own learning rate while the matrix branch keeps the main one (unit-tested). Registered beforehand (Appendix E.7): polar-match with the matrix branch at $6{\cdot}10^{-6}$ and the rest at $2{\cdot}10^{-6}$ collapses in 1 of 3 runs (steps 56/–/–; prediction held); AdamW with the matrix branch at $2{\cdot}10^{-6}$ and the rest at $6{\cdot}10^{-6}$ collapses in 0 of 3 (prediction failed). On Qwen3-4B, polar-match with the matrix branch at $4{\cdot}10^{-5}$ and the rest at 10^{-5}: failed.

E.7 REGISTERED PREDICTIONS

Several tests in this paper were written down, with a refutation rule, before the runs they concern had finished; the files are released with the code. They are local files, and their hand-written clock times contain errors that we have corrected in the files from the run logs; except where noted, the order of registration and outcome is therefore not externally verifiable. Table 38 lists every registered prediction, including the ones that failed.

Table 38: Every registered prediction and its outcome (Qwen3-0.6B unless noted).

Registration	Prediction	Outcome	Reported in
E3	Edge-truncated at $3{\cdot}10^{-7}$ is stable	held (3 of 3 stable)	Table 27
E3	Pion filter at $5{\cdot}10^{-7}$ collapses	failed (1 of 1 stable)	Table 27
E1/E2	Aligned norm separates: AdamW $2.5{\cdot}10^{-6}$ collapses, polar $6{\cdot}10^{-6}$ is stable	failed	Appendix E.8
v8	The relative perturbation does not separate stable from collapsed runs	held	Appendix E.3
M1	MATH: a learning rate at which AdamW collapses in every seed and polar-match in none	held at $5{\cdot}10^{-6}$	Table 12
M2	MATH: edge-truncated collapses no later than AdamW at the same learning rate	held at $5{\cdot}10^{-6}$ (registered) and at $2{\cdot}10^{-6}$	Table 12
M3	MATH: polar-match tolerates a larger learning rate than AdamW	held (stable $3/3$ at $5{\cdot}10^{-6}$ where AdamW collapses $3/3$)	Table 12
M4	MATH: GSM8K aligned-norm band predicts each run outside it (registered after the first five runs had started)	held; matched by learning-rate thresholds	Section 5.6
L1, L3	200 steps: polar-match $5{\cdot}10^{-6}$ is stable, or outlasts AdamW $2{\cdot}10^{-6}$	failed (collapses earlier)	Section 5.6
L4	200 steps: polar $6{\cdot}10^{-6}$ collapses within $1.5\times$ of AdamW $2{\cdot}10^{-6}$	held ($n = 1$; see L3)	Section 5.6
H1†	Held-out $US^{0.5}V^\top$ at $2.4{\cdot}10^{-6}$: at least 2 of 3 collapse (aligned norm), at most 1 (spectral flatness)	aligned norm held (3 of 3)	Section 5.4
H1c†	Held-out $US^{0.5}V^\top$ at $1.2{\cdot}10^{-6}$: stable (both)	failed (collapse at step 59)	Section 5.4
H2†	Held-out $US^{0.25}V^\top$ at $6{\cdot}10^{-6}$: both runs collapse	held (2 of 2)	Table 21
G1†	Direction offsets in the accuracy gain stay non-significant when the added seeds are included	held ($p = 0.30$)	Table 31
G2†	Matched-window gains of two directions differ by less than 3 points	held (1.8)	Table 31
S1†	Qwen3-4B: both direction constants exceed their Qwen3-1.7B values, ratio at least 3.0	held (ratio 4.04)	Table 25
S2†	Qwen3-4B: the aligned norm extrapolated from two scales, not refitted, predicts every run outside its band	failed (1 of 8 wrong)	Table 25
S3†	Qwen3-4B: a common rate at which AdamW collapses and polar-match does not; threshold ratio above 1	failed	Table 25
S4†	Qwen3-4B: that threshold ratio within a factor 1.5 of the constant ratio	failed (ratio in $[0.50, 2]$)	Table 25
S5, S6†	Qwen3-4B, seeds added: at $2{\cdot}10^{-5}$ both directions collapse in at least 2 of 3; at 10^{-5} both stable in 3 of 3	failed; failed (AdamW 1 of 3 at 10^{-5})	Section 5.5
AL3†	Qwen3-4B, polar aligned-matched per matrix at 10^{-5}: collapses in at least as many runs as AdamW	held (3 of 3)	Table 14
L1–L3†	The same on Llama-3.2-1B-Instruct, six states	failed, failed, held	Table 34
M1–M3 (1.7B)†	The same on Qwen3-1.7B, six states	held, held, held	Table 35

Registration	Prediction	Outcome	Reported in
AL1, AL2†	Qwen3-1.7B, polar aligned-matched per matrix: at least 2 of 3 collapse at 10^{-5}; 0 of 2 at $5{\cdot}10^{-6}$	held (3 of 3); failed	Table 14
F1, F2†	Llama-3.2-1B-Instruct, grid up to 10^{-5}: a rate at which polar-match is stable and AdamW collapses; threshold ratio within a factor two of the constant ratio	failed; failed (no threshold bracketed; located by an exploratory extension)	Table 18
PP1, PP2†	Public-code Muon ($\kappa = 0.2$), threshold predicted from its ten-step constant as $2.04{\cdot}10^{-5}$: at most 1 of 3 collapse at $0.7\times$, at least 2 of 3 at $1.4\times$	held (1 of 3); held (3 of 3)	Section 5.4
RC1, RC2†	200 steps with clip 1.0 and the KL penalty: polar-match crosses later than every AdamW run; no more than one aligned-matched run crosses later than the latest AdamW run	held; held	Table 19
PP3, PP4†	Public per-head Pion, threshold predicted as $2.79{\cdot}10^{-6}$: at most 1 of 3 collapse at $0.7\times$, at least 2 of 3 at $1.4\times$	failed (3 of 3); held (3 of 3)	Section 5.4
MX1–MX3†	Aligned-matched polar: at $2.5{\cdot}10^{-6}$ at least 1 of 3 collapse; over 200 steps without penalty at most one run later than AdamW's step 86; on Llama-3.2-1B at $2{\cdot}10^{-5}$ at least 2 of 3	held (3 of 3); held (all three earlier); held (3 of 3)	Table 15
KM1, KM2†	KL-matched polar (matched step $\times 0.88$): at least 2 of 3 collapse at 2.5 and at $3{\cdot}10^{-6}$	held (2 of 3); held (3 of 3)	Table 15
NB1–NB3†	Seeds 3–4 at the boundary cells: at $2.5{\cdot}10^{-6}$ aligned- and KL-matched polar collapse in at least 1 of 5	held (5 of 5, Fisher $p = 3.8{\cdot}10^{-4}$); held (4 of 5)	Table 15
MM1, MM2†	Polar of Muon's own momentum, aligned-matched: at least 1 of 3 collapse at $2.5{\cdot}10^{-6}$, at least 2 of 3 at $3{\cdot}10^{-6}$	held (2 of 3); held (3 of 3)	Table 15
MT1, MT2†	MATH, aligned-matched polar: at least 2 of 3 collapse at $5{\cdot}10^{-6}$; at $3{\cdot}10^{-6}$ at least as many as AdamW	held (3 of 3); held (3 of 3, AdamW 0 of 2)	Table 15
RS1, RS2†	Aligned-matched polar times 0.71: at most 1 of 3 collapse at $2.5{\cdot}10^{-6}$; times 0.50: none	held (0 of 2, one requeued); held (0 of 3)	Table 17
KP1†	Polar rescaled every step to AdamW's one-step KL: at least 2 of 3 collapse at $3{\cdot}10^{-6}$ (AdamW 5 of 6)	held, narrowly (steps 58, 60); with seeds 3–5, 5 of 6 (KP6)	Table 15
KP2, KP3†	The same at $2.5{\cdot}10^{-6}$ (at least 1 of 3) and in the 200-step KL recipe (onset in at least 2 of 3, median no later than AdamW's)	KP2 held (1 of 3, AdamW 1 of 6); KP3 held (onset 62/58/64 against 62/66/73)	Table 15
KP4	The same at $2{\cdot}10^{-6}$, descriptive (AdamW 0 of 5)	reported in Table 15	Table 15
KP5–KP7†	Seeds 3–5 at $2.5{\cdot}10^{-6}$ and $3{\cdot}10^{-6}$: pooled collapse fraction not different from AdamW's (two-sided Fisher $p > 0.05$); at $3{\cdot}10^{-6}$ at least 4 of 6; 200-step KL + clip recipe: onset in at least 2 of 3	KP5 held at six seeds ($p = 0.68$; sixteen seeds: EQ1); KP6 held (5 of 6, AdamW 5 of 6); KP7 held (onset 63/61/62 against 58/63/67)	Table 15
PW1–PW3†	AdamW with distinct seeds 3–5 keeps its counts (at most 3 of 9 at $2.5{\cdot}10^{-6}$, at least 5 of 9 at $3{\cdot}10^{-6}$); eight policy-matched seeds per rate: Fisher $p > 0.05$; 1.7B policy-matched at 10^{-5}: at least 2 of 3	PW1 held (2 of 9, 8 of 9); PW2 held at eight seeds (10 of 16 against 8 of 12, $p = 1.00$; sixteen seeds: EQ1); PW3 held (2 of 3 against 3 of 3)	Table 15
KP8, KP9†	Policy match on MATH (at most 1 of 3 at $3{\cdot}10^{-6}$; at least 2 of 3 at $5{\cdot}10^{-6}$) and on Llama-3.2-1B at $2{\cdot}10^{-5}$ (at least 2 of 3, median no earlier than AdamW's minus 5)	held (1 of 3; 3 of 3); held (29/36/21 against 30/21/21)	Table 15
EQ1†	Sixteen seeds per arm and rate (at registration): the rate-stratified difference in collapse fraction, policy-matched minus AdamW, has a 90% interval inside ± 25 points	held (17 of 32 against 15 of 32; $+6.2$ points, 90% interval -9.7 to $+22.2$)	Table 15

Registration	Prediction	Outcome	Reported in
EQ2†	Twenty-four seeds per arm and rate: the same 90% interval inside ± 20 points	failed ($+10.4$ points, 90% interval -2.8 to $+23.7$; still inside EQ1's ± 25)	Table 15
EQ3, H4d†	Thirty-two seeds per arm and rate (interval reported); Qwen3-4B polar-match at $1.5{\cdot}10^{-5}$ collapses in at most half its runs	EQ3: $+7.8$ points, 90% interval -4.0 to $+19.6$; H4d held (1 of 4; AdamW at 10^{-5} 3 of 5)	Table 15
PB1†	Public Pion ($\kappa = 5$) at its best of 1, 1.5, $2{\cdot}10^{-6}$ does not beat AdamW's best accuracy by more than one point	held (best 73.5% at $1.5{\cdot}10^{-6}$, AdamW's best 75.6%)	Tables 29 and 30
FH1, FH2†	Non-matrix branch frozen, 3 and $5{\cdot}10^{-6}$, four seeds: aligned-matched collapses at least as often as AdamW; policy-matched not different (Fisher $p > 0.05$)	held (8 of 8 against 8 of 8, earlier in every seed pair); held (7 of 8, $p = 1.00$)	Table 15
PR1†	Warmup, cosine decay, KL penalty and clip over 200 steps, best rate against best rate: polar-match does not beat AdamW by more than one point	held at three seeds ($+1.0 \pm 0.8$); PR3 at eight seeds held ($+0.6$, 95% interval -0.1 to $+1.3$); no collapse, no KL onset	Table 30
BB1, BB2†	Best rate against best rate, 8 seeds each: polar-match at $5{\cdot}10^{-6}$ more accurate than AdamW at $2{\cdot}10^{-6}$ (one-sided Welch $p < 0.05$), by what the frozen accuracy line predicts	BB1 failed (75.7% against 75.6%, $p = 0.433$); BB2 held	Figure 8
AP1†	Public Pion, per-head Pion and polar-match with seeds 3–7: mean residual from the frozen accuracy line within 2 s.e. of zero	failed for public Pion ($+2.3 \pm 0.2$) and polar-match ($+1.9 \pm 0.6$); held for per-head Pion	Figure 8
H4a, H4b†	Qwen3-4B: polar-match stable at $1.5{\cdot}10^{-5}$ (0 of 2); aligned-matched polar at $5{\cdot}10^{-6}$ collapses at least as often as AdamW	H4a failed (1 of 2, step 42); H4b failed (0 of 2 against AdamW's 1 of 1); H4c failed (0 of 3 against 2 of 3)	Table 14
B17a, B17b†	Qwen3-1.7B: polar-match at $1.4{\cdot}10^{-5}$ collapses in at most 1 of 3; AdamW at $7{\cdot}10^{-6}$ reported	B17a failed (2 of 3); with five seeds (B17c) 3 of 5 and AdamW 3 of 5; ratio bracketed in $(1.4, 2.8]$	Table 6
GC3†	Clip 1.0 at $3{\cdot}10^{-6}$: AdamW collapses in at least as many of 2 runs as polar-match	held (AdamW 1 of 2, polar-match 0 of 2)	Table 20
GA1†	Public Pion ($\kappa = 5$) at 3 and $6{\cdot}10^{-6}$: stable runs' gain within the accuracy bound of the line	unscorable: all four runs collapse (threshold predicted $1.07{\cdot}10^{-5}$, observed in $(10^{-6}, 3{\cdot}10^{-6}]$)	Figure 8
GC1, GC2†	With a global gradient clip of 1.0 and of 0.15 at $5{\cdot}10^{-6}$: AdamW collapses in at least 1 of 2 runs and polar-match in none	held; held	Table 20
K1, K2†	200 KL-penalized steps, polar matched per matrix in aligned norm at AdamW's learning rate: no run crosses KL > 1 later than the latest AdamW run, and all cross within 200 steps	held; held	Figure 6c
RR1, RR2†	A Haar-random basis behaves like the momentum-derived ones: aligned spread below 10; spectrum effect above rotation effect	failed; failed	Appendix E.5
C1–C3†	One step from a shared state: aligned $<$ operator $<$ RMS spread; aligned spread < 10; spectrum effect $>$ basis effect, in every state	held, held, held (9 states)	Table 33
R1/R2†	Aligned-matched polar at $2{\cdot}10^{-6}$: at most 1 of 3 collapses (R1: residual is the non-matrix branch)	R1 on the registered 3 seeds; 2 of 5 with seeds added (NB3), against AdamW's 0 of 5	Table 36
A3†	Aligned-matched polar at $3{\cdot}10^{-6}$: at least 2 of 3 collapse	held	Table 36
A4†	Aligned-matched Pion filter at $2{\cdot}10^{-6}$, R1/R2 rule	R2 (residual kept)	Table 36

Registration	Prediction	Outcome	Reported in
P1†	Per-head Pion at matched RMS: $2.5{\cdot}10^{-7}$ at most 1 of 3 collapses; 10^{-6} at least 2 of 3	held; held	Section 5.2
N1–N3†	Split learning rate: polar matrix $6{\cdot}10^{-6}$/rest $2{\cdot}10^{-6}$ at most 1 of 3; AdamW $2{\cdot}10^{-6}$/rest $6{\cdot}10^{-6}$ at least 2 of 3; 4B polar $4{\cdot}10^{-5}$/rest 10^{-5} stable	held; failed; failed	Section 5.4

†Sent as a timestamped message to an external channel before the tested runs were launched, and also committed to a version-controlled file (H1, H2: commits `437dd15`, `34d4420`, made after the ten measurement steps and before the runs; G1–S4: commit `c66787b`, made after those runs had started, so the external message is the timestamp that precedes them). The registration texts are in the supplementary material; the repository with their history identifies the authors and is withheld until after review. One amendment to the Qwen3-4B grid was externally timestamped a minute after the runs it describes began; the registration file records which. The registrations E3, v8 and M1–L4 were kept as local files and entered version control only later.

The policy match. At every step the trainer measures, on the first micro-batches of that step (sequences sampled from the pre-step policy), the k_3 estimate of the one-step KL for the aligned-matched polar step and for the Adam step taken from the same state, then rescales all matrix steps by one scalar s. Because the non-matrix branch is not rescaled, the KL is $a + bs^2$ rather than bs^2; a and b are fitted from the KL at $s = 1$ and at $s_1 = \sqrt{\mathrm{KL}_{\mathrm{Adam}}/\mathrm{KL}_{\mathrm{polar}}}$, and s solves $a + bs^2 = \mathrm{KL}_{\mathrm{Adam}}$, clamped to $[1/4, 4]$. The registered rule was the one-shot s_1; two steps of the first run showed that it leaves the KL 37% above AdamW's, and the rule was changed before any outcome, that run restarting from step 0. Over all 85 runs the median final KL relative to AdamW's lies in 0.96–1.04. For Llama the two recorded steps are kept in host memory, which changes no value.

E.8 Quantities built from the input-aligned norm

Proposition 2(c) is a statement about what one step does to the output of one layer; a collapse is the outcome of many steps on many layers. Turning the first into a criterion for the second requires a choice of aggregation, and the choice is not forced by the proposition. Table 39 evaluates ten of them on the 86 runs that log the aligned norm: per-step values on a fixed window and on the pre-collapse window, the cumulative aligned drift at fixed steps and at the collapse step, its growth rate, the anisotropy ratio on its own, two forms that separate the size of a step from the coherence with which steps accumulate, and, as a control that is not an aligned quantity, the trainer–sampler log-probability gap. None separates the stable from the collapsed runs: in every row the largest stable value exceeds the smallest collapsed one. Six of the nine aligned quantities nevertheless order the collapsed runs with $|\rho| \geq 0.85$. The three that do not are the ones that discard the magnitude of the step (the anisotropy ratio alone, and the coherence of the drift) or mix runs of different length (the drift read at each run's own collapse step). The two runs that break the separation were not found by inspection after the fact. Before they were launched we registered, in a file released with the code, the prediction that AdamW at $2.5{\cdot}10^{-6}$ and polar-match at $6{\cdot}10^{-6}$ would fall on the collapsed and the stable side respectively, together with the statement that the separation would have to be narrowed if either turned out otherwise. Both did: in that registered batch two of the three AdamW runs at $2.5{\cdot}10^{-6}$ are stable, and one of the three polar-match runs at $6{\cdot}10^{-6}$ collapses (all runs at these learning rates are counted in Table 27). The wording of Section 5.4 is the result of that test. We read this as evidence that the aligned norm captures a real ordering and that the aggregation, not the proposition, is what fails to yield a threshold; a per-step displacement measured in absolute units and averaged over matrices whose outputs have different scales is not the dimensionless quantity a threshold would need.

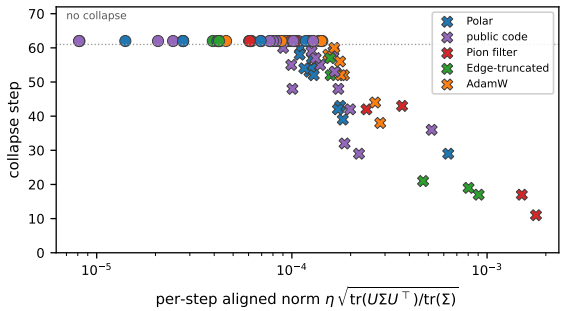


Figure 11: Collapse step against the input-aligned norm of the step, for the 86 Qwen3-0.6B runs that log it (circles at the top: no collapse within 60 steps). Colours are direction families. Repeats of one configuration differ by a median of 6 steps vertically, and change outcome in 4 of 30 groups, all with late collapses (Table 8).

F THE NOISE-CALIBRATED GATE

This appendix specifies the per-matrix gate of Section 3.2 and reports the runs in which it was used. We do not propose it as an optimizer: it is a probe of online direction selection, and its result is a negative one that belongs to the argument of Sections 5.3 and 5.4 — a rule that picks, at each matrix and step, the direction better aligned with the expected gradient moves toward the concentrated direction exactly when the policy starts to destabilize, and collapses earlier than AdamW (Appendix F.5). The noise edge specified here is also the one the edge-truncated direction uses.

We call the gate NCSH (noise-calibrated spectral hybrid) below; the name refers only to this probe. Its analysis is limited to the calibration of the noise edge (Lemma 1 and Corollary 1), which the edge-truncated arm needs.

F.1 NOISE EDGE FROM MICRO-BATCH GRADIENTS

A noise sample comes for free from the first two micro-batches: $\Delta_t = (\hat{g}_{t,1} - \hat{g}_{t,2})/\sqrt{2}$. If the micro-batch gradients are conditionally independent with a common mean, Δ_t has zero mean and the covariance of the noise of a single micro-batch gradient. The noise of G_t has $1/K$ of that covariance, and the bias-corrected momentum $\hat{m}_t = \sum_{i\le t} \omega_{t,i} G_i$ multiplies it by $w_t = \sum_i \omega_{t,i}^2$, for which Lemma 1 gives

$$w_t = \frac{1-\beta_1}{1+\beta_1} \cdot \frac{1+\beta_1^t}{1-\beta_1^t}, \tag{4}$$

much larger than its limit $\frac{1-\beta_1}{1+\beta_1}$ in short runs ($2.07\times$ at $t = 10$ for $\beta_1 = 0.9$). The noise edge e_t of Section 3.2 uses the safety factor $c_{\mathrm{edge}} = 1.05$ and computes $\sigma_{\max}(\Delta_t)$ by $J = 50$ power iterations from a fixed start vector, which return a lower bound. For Gaussian noise, $\sigma_{\max}(\Delta_t)\sqrt{w_t/K}$ has exactly the distribution of the top singular value of the noise in $\hat{m}_t$ (Corollary 1). Unlike the isotropic edge $\tau(\sqrt{d_1} + \sqrt{d_2})$, e_t does not assume i.i.d. entries, which matters because gradient noise in language models concentrates on a few activation directions (Appendix E.2).

F.2 EDGE-TRUNCATED POLAR DIRECTION

Let $\hat{m}_t = \sum_i s_i u_i v_i^\top$ with $s_1 \ge s_2 \ge \cdots$. The truncated direction $N_t = U_{k_t} V_{k_t}^\top$ of Section 3.2 keeps the k_t singular directions above the edge, and $\hat{N}_t = N_t/\sqrt{k_t}$ is its unit version. Since $s_{k_t} > e_t \ge s_{k_t+1}$, N_t is unique even when singular values are repeated, and $e_t = 0$ gives the polar factor. Pion's filter is instead a soft threshold relative to $\|\hat{m}_t\|_F$ (Proposition 1). If $k_t = 0$, or if fewer than two micro-batches are available, no truncated direction exists and the step uses A_t.

F.3 Cross-fitted alignments and the closed-form gate

Cross-fitting. The gate needs the alignment of each candidate direction with the expected gradient $\bar{G}_t$. Scoring $\hat{N}_t$ on G_t is biased upward, because $\hat{N}_t$ is computed from a momentum that contains G_t; Appendix B checks this numerically. Instead, at the start of step t, before the moments are updated, we score the previous unit directions on the fresh gradient, $\hat{a}_t = \langle \hat{N}_{t-1},\, G_t \rangle$ and $\hat{b}_t = \langle \hat{A}_{t-1},\, G_t \rangle$ with $\hat{A}_{t-1} = A_{t-1}/\|A_{t-1}\|_F$. Both estimates are made only when $\hat{N}_{t-1}$ exists, from stored factors and moments (Appendix C). Exponential moving averages with rate $\beta_s = 0.8$, bias-corrected by $1/(1-\beta_s^{n_t})$ with n_t the number of estimates so far, give $\bar{a}_t, \bar{b}_t$. After the moment update we compute $\hat{N}_t$, $\hat{A}_t$ and the exact overlap $\rho_t = \langle \hat{N}_t,\, \hat{A}_t \rangle$.

Closed-form gate. For unit directions $\hat{N}, \hat{A}$ with overlap $\rho \in (-1, 1)$ and projections $a = \langle \hat{N},\, \bar{G} \rangle$, $b = \langle \hat{A},\, \bar{G} \rangle$, let $p = a - \rho b$ and $q = b - \rho a$. The gate is

$$\sigma^\star(a, b, \rho) = \begin{cases} 0, & a \le 0 \text{ and } b \le 0, \\ \dfrac{p}{p+q}, & p > 0 \text{ and } q > 0, \\ 1, & \text{otherwise, if } a > b, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

By elementary geometry in the plane of the two directions, whenever $\max(a, b) > 0$, $\sigma^\star$ is the exact maximizer over $\sigma \in [0, 1]$ of $\cos \angle(\sigma\hat{N} + (1-\sigma)\hat{A},\ \bar{G})$; in the interior case the blend is parallel to the orthogonal projection of $\bar{G}$ onto $\mathrm{span}\{\hat{N}, \hat{A}\}$. When neither direction is positively aligned, every blend has non-positive cosine and the gate returns the Adam direction by convention. The gate depends on (a, b) only through their ratio, so it has no temperature or threshold. At step t we use $\sigma_t = \sigma^\star(\bar{a}_t, \bar{b}_t, \rho_t)$ once $n_t \ge n_{\min} = 2$, and $\sigma_t = 0$ before. The update is $D_t = \sigma_t \hat{N}_t + (1-\sigma_t)\hat{A}_t$, rescaled by Equation (3), and $\sigma_t = 0$ is implemented as $\tilde{D}_t = A_t$ exactly. The implementation also provides two controls with the same $\hat{N}_t$: *fixed-σ* replaces σ_t by a constant, and *shuffled* applies, at every step, a random permutation of the values $\sigma_t^\star$ across the matrices that have a truncated direction.

F.4 Analysis

All statements concern a single matrix; proofs are in Appendix A and numerical checks in Appendix B. We do not claim convergence guarantees. Let $\mathcal{F}_{t-1}$ denote the information available after the parameter update of step $t-1$; it determines θ_t, m_{t-1}, v_{t-1}, $\hat{N}_{t-1}$ and $\hat{A}_{t-1}$. Write $\bar{G}_t = \mathbb{E}[G_t \mid \mathcal{F}_{t-1}]$ for the expected surrogate gradient at θ_t, and $\xi_t = G_t - \bar{G}_t$.

Assumption 1 (Micro-batch noise)**.** Conditionally on $\mathcal{F}_{t-1}$, the micro-batch gradients $\hat{g}_{t,1}, \ldots, \hat{g}_{t,K}$ are independent with common mean $\bar{G}_t$ and common covariance Σ_t (acting on vec). For the edge calibration we additionally assume $\Sigma_t = \Sigma$ over the momentum window.

Lemma 1 (Noise weight of the bias-corrected EMA)**.** *Let $m_0 = 0$, $m_t = \beta m_{t-1} + (1-\beta)G_t$ and $\hat{m}_t = m_t/(1-\beta^t)$. Then $\hat{m}_t = \sum_{i=1}^t \omega_{t,i} G_i$ with $\omega_{t,i} = (1-\beta)\beta^{t-i}/(1-\beta^t)$ and $\sum_i \omega_{t,i} = 1$, and $w_t := \sum_{i=1}^t \omega_{t,i}^2$ is given by Equation (4) with $\beta_1 = \beta$. It decreases from $w_1 = 1$ to $(1-\beta)/(1+\beta)$. If the noise terms ξ_i are uncorrelated with common covariance Σ, the noise of $\hat{m}_t$ has covariance $w_t\Sigma$.*

Corollary 1 (Calibration of the noise edge)**.** *Under Assumption 1 with $\Sigma_t = \Sigma$ and Gaussian micro-batch noise independent across steps, the noise component $\sum_i \omega_{t,i}\xi_i$ of $\hat{m}_t$ has the same distribution as $\sqrt{w_t/K}\,\Delta_t$. Hence $e_t/c_{\mathrm{edge}} = \sigma_{\max}(\Delta_t)\sqrt{w_t/K}$ has exactly the distribution of the top singular value of the momentum noise.*

F.5 Results of the gate probe

The gate is run on Qwen3-0.6B at $\eta = 5{\cdot}10^{-6}$ with RMS matching (Equation (3)) and the protocol of Section 5.1. Because it blends toward the edge-truncated direction, which is close to rank one (Section 5.3), first-order alignment can select a concentrated step. Table 40 compares it with its two

endpoints, the edge-truncated direction ($\sigma_t = 1$) and the Adam direction ($\sigma_t = 0$, i.e. AdamW), and with polar-match. The gate statistics are averaged over matrix-steps before the collapse step. The gate collapses at steps 14/17/18 ($n = 3$), earlier than every AdamW seed and about as early as the edge-truncated direction. In every seed it selects the Adam direction in most matrix-steps during the first ten steps, and moves toward the edge-truncated direction (σ_t between 0.5 and 0.9, $k_t \approx 1$) over the steps in which entropy starts to rise, just before the collapse. First-order alignment with the surrogate gradient thus selects the concentrated step when the policy begins to destabilize, which is consistent with the early collapse of the concentrated directions (Section 5.3) and is why we do not propose the gate as an optimizer.

Table 11: Weight drift at the last step T of each run, from bfloat16 checkpoints (mean over the projection matrices of the first, middle and last layers): operator norm, Frobenius norm and stable rank $\|W_T - W_0\|_F^2 / \|W_T - W_0\|_2^2$, and the operator-norm and Frobenius drift relative to $\|W_0\|_2$ and $\|W_0\|_F$. "steps": optimizer steps taken; collapse-stopped runs end early, and runs that predate the stop rule continue after their collapse step.

Update	model	seed	steps	collapse step	$\|W_T - W_0\|_2$	$\|W_T - W_0\|_F$	stable rank	$\frac{\|W_T - W_0\|_2}{\|W_0\|_2}$	$\frac{\|W_T - W_0\|_F}{\|W_0\|_F}$
AdamW (1e-6)	0.6B	1	60	–	1.7×10^{-3}	7.8×10^{-3}	26	4.3×10^{-4}	2.0×10^{-4}
AdamW (1e-6)	0.6B	2	60	–	1.9×10^{-3}	7.9×10^{-3}	24	4.6×10^{-4}	2.0×10^{-4}
AdamW (2e-6)	0.6B	0	60	–	5.1×10^{-3}	1.9×10^{-2}	20	1.2×10^{-3}	4.8×10^{-4}
AdamW (2e-6)	0.6B	1	60	–	5.6×10^{-3}	2.0×10^{-2}	17	1.4×10^{-3}	5.1×10^{-4}
AdamW (2e-6)	0.6B	2	60	–	5.5×10^{-3}	2.0×10^{-2}	19	1.3×10^{-3}	5.0×10^{-4}
Fixed gate 0.25 (1e-6)	0.6B	0	60	–	3.7×10^{-3}	8.0×10^{-3}	6	8.9×10^{-4}	2.0×10^{-4}
Fixed gate 0.75 (1e-6)	0.6B	0	60	–	8.7×10^{-3}	1.1×10^{-2}	2	2.0×10^{-3}	2.6×10^{-4}
Gate probe (1e-6)	0.6B	1	60	–	6.2×10^{-3}	8.4×10^{-3}	2	1.5×10^{-3}	2.1×10^{-4}
Muon (public code, κ=0.2)	0.6B	1	60	–	3.8×10^{-4}	4.5×10^{-3}	154	8.7×10^{-5}	1.1×10^{-4}
Muon (public code, κ=0.2)	0.6B	2	60	–	3.9×10^{-4}	4.5×10^{-3}	149	9.0×10^{-5}	1.1×10^{-4}
Per-head Pion (public code, κ=5)	0.6B	0	60	–	6.3×10^{-3}	7.0×10^{-3}	1	1.5×10^{-3}	1.9×10^{-4}
Pion (public code, κ=5)	0.6B	1	60	–	2.0×10^{-3}	2.1×10^{-3}	1	4.9×10^{-4}	5.4×10^{-5}
Pion (public code, κ=5)	0.6B	2	60	–	1.5×10^{-3}	1.5×10^{-3}	1	3.7×10^{-4}	4.1×10^{-5}
Polar-match (1e-6)	0.6B	1	60	–	5.6×10^{-4}	6.8×10^{-3}	157	1.3×10^{-4}	1.7×10^{-4}
Polar-match (1e-6)	0.6B	2	60	–	5.6×10^{-4}	6.8×10^{-3}	155	1.3×10^{-4}	1.7×10^{-4}
Polar-match (2e-6)	0.6B	0	60	–	1.2×10^{-3}	1.8×10^{-2}	204	2.9×10^{-4}	4.3×10^{-4}
Polar-match (2e-6)	0.6B	1	60	–	1.2×10^{-3}	1.8×10^{-2}	212	2.8×10^{-4}	4.3×10^{-4}
Polar-match (2e-6)	0.6B	2	60	–	1.2×10^{-3}	1.7×10^{-2}	211	2.8×10^{-4}	4.3×10^{-4}
Polar-match (3e-6)	0.6B	0	60	–	2.0×10^{-3}	3.0×10^{-2}	243	4.5×10^{-4}	7.4×10^{-4}
Polar-match (3e-6)	0.6B	2	60	–	1.9×10^{-3}	3.0×10^{-2}	247	4.5×10^{-4}	7.5×10^{-4}
Polar-match (5e-6)	0.6B	0	60	–	3.3×10^{-3}	5.6×10^{-2}	288	7.6×10^{-4}	1.4×10^{-3}
Polar-match (5e-6)	0.6B	1	60	–	3.5×10^{-3}	5.8×10^{-2}	280	8.0×10^{-4}	1.4×10^{-3}
Polar-match (5e-6)	0.6B	2	60	–	3.6×10^{-3}	5.9×10^{-2}	276	8.4×10^{-4}	1.5×10^{-3}
AdamW (5e-6)	1.7B	0	60	–	2.0×10^{-2}	9.0×10^{-2}	24	2.7×10^{-3}	1.1×10^{-3}
AdamW (5e-6)	1.7B	1	60	–	2.0×10^{-2}	8.8×10^{-2}	25	2.7×10^{-3}	1.0×10^{-3}
Per-head Pion (public code, κ=5)	1.7B	0	60	–	7.2×10^{-3}	7.9×10^{-3}	1	1.0×10^{-3}	1.3×10^{-4}
Polar-match (1e-6)	1.7B	0	60	–	5.7×10^{-4}	9.2×10^{-3}	289	7.2×10^{-5}	1.1×10^{-4}
Polar-match (1e-5)	1.7B	0	60	–	8.5×10^{-3}	2.0×10^{-1}	525	1.1×10^{-3}	2.3×10^{-3}
Polar-match (1e-5)	1.7B	1	60	–	8.4×10^{-3}	1.9×10^{-1}	522	1.1×10^{-3}	2.3×10^{-3}
Polar-match (5e-6)	1.7B	0	60	–	4.0×10^{-3}	8.7×10^{-2}	465	5.2×10^{-4}	1.0×10^{-3}
Polar-match (5e-6)	1.7B	1	60	–	4.1×10^{-3}	8.7×10^{-2}	457	5.2×10^{-4}	1.0×10^{-3}
AdamW (3e-6)	0.6B	0	60	56	2.0×10^{-2}	4.7×10^{-2}	8	4.8×10^{-3}	1.2×10^{-3}
AdamW (3e-6)	0.6B	1	60	52	2.2×10^{-2}	5.0×10^{-2}	8	5.2×10^{-3}	1.2×10^{-3}
AdamW (3e-6)	0.6B	2	60	56	1.8×10^{-2}	4.4×10^{-2}	9	4.4×10^{-3}	1.1×10^{-3}
AdamW (5e-6)	0.6B	0	60	35	3.6×10^{-2}	7.8×10^{-2}	7	8.6×10^{-3}	2.0×10^{-3}
AdamW (5e-6)	0.6B	1	40	34	2.4×10^{-2}	6.0×10^{-2}	11	5.9×10^{-3}	1.5×10^{-3}
AdamW (5e-6)	0.6B	2	39	32	2.4×10^{-2}	6.1×10^{-2}	9	5.9×10^{-3}	1.5×10^{-3}
Edge-truncated, RMS-matched (1e-6)	0.6B	0	60	57	1.0×10^{-2}	1.2×10^{-2}	2	2.4×10^{-3}	3.0×10^{-4}
Edge-truncated, RMS-matched (1e-6)	0.6B	1	60	51	1.2×10^{-2}	1.5×10^{-2}	2	2.8×10^{-3}	3.7×10^{-4}
Edge-truncated, RMS-matched (1e-6)	0.6B	2	54	35	1.2×10^{-2}	1.6×10^{-2}	2	2.9×10^{-3}	3.9×10^{-4}
Edge-truncated, RMS-matched (5e-6)	0.6B	0	60	16	2.7×10^{-2}	4.1×10^{-2}	3	6.3×10^{-3}	1.0×10^{-3}
Edge-truncated, RMS-matched (5e-6)	0.6B	1	18	11	1.7×10^{-2}	2.5×10^{-2}	2	3.9×10^{-3}	6.1×10^{-4}
Edge-truncated, RMS-matched (5e-6)	0.6B	2	21	16	1.7×10^{-2}	2.5×10^{-2}	3	3.9×10^{-3}	6.3×10^{-4}
Edge-truncated, op.-norm-matched (5e-6)	0.6B	0	36	29	2.2×10^{-2}	2.0×10^{-1}	85	5.3×10^{-3}	5.0×10^{-3}
Edge-truncated, op.-norm-matched (5e-6)	0.6B	1	29	21	2.1×10^{-2}	2.1×10^{-1}	109	5.0×10^{-3}	5.3×10^{-3}
Fixed gate 0.5 (1e-6)	0.6B	0	60	59	8.4×10^{-3}	1.2×10^{-2}	2	1.9×10^{-3}	3.0×10^{-4}
Gate probe (1e-6)	0.6B	0	60	45	1.1×10^{-2}	1.6×10^{-2}	2	2.6×10^{-3}	3.9×10^{-4}
Gate probe (1e-6)	0.6B	2	60	57	9.0×10^{-3}	1.2×10^{-2}	2	2.1×10^{-3}	2.9×10^{-4}
Gate probe (5e-6)	0.6B	0	24	14	1.7×10^{-2}	3.1×10^{-2}	5	3.9×10^{-3}	7.8×10^{-4}
Gate probe (5e-6)	0.6B	1	24	17	1.7×10^{-2}	3.2×10^{-2}	6	3.9×10^{-3}	8.0×10^{-4}
Gate probe (5e-6)	0.6B	2	24	18	1.6×10^{-2}	3.2×10^{-2}	5	3.7×10^{-3}	7.9×10^{-4}
Gate, op.-norm-matched (5e-6)	0.6B	0	33	27	2.0×10^{-2}	1.4×10^{-1}	59	4.7×10^{-3}	3.5×10^{-3}
Muon (public code, κ=5)	0.6B	1	50	40	9.4×10^{-3}	1.7×10^{-1}	336	2.2×10^{-3}	4.3×10^{-3}
Muon (public code, κ=5)	0.6B	2	53	43	9.9×10^{-3}	1.8×10^{-1}	326	2.3×10^{-3}	4.4×10^{-3}
Pion filter, RMS-matched (1e-6)	0.6B	0	54	36	1.7×10^{-2}	1.8×10^{-2}	1	4.0×10^{-3}	4.6×10^{-4}
Pion filter, RMS-matched (1e-6)	0.6B	1	43	34	1.5×10^{-2}	1.6×10^{-2}	1	3.5×10^{-3}	4.0×10^{-4}
Pion filter, RMS-matched (1e-6)	0.6B	2	60	50	1.6×10^{-2}	1.7×10^{-2}	1	3.7×10^{-3}	4.3×10^{-4}
Pion filter, RMS-matched (5e-6)	0.6B	0	60	17	5.0×10^{-2}	5.4×10^{-2}	1	1.2×10^{-2}	1.3×10^{-3}
Pion filter, RMS-matched (5e-6)	0.6B	1	15	11	3.1×10^{-2}	3.3×10^{-2}	1	7.4×10^{-3}	8.2×10^{-4}
Pion filter, RMS-matched (5e-6)	0.6B	2	15	11	2.8×10^{-2}	3.0×10^{-2}	1	6.5×10^{-3}	7.4×10^{-4}
Pion filter, op.-norm-matched (5e-6)	0.6B	0	60	19	3.1×10^{-2}	3.2×10^{-2}	1	7.3×10^{-3}	8.1×10^{-4}
Pion filter, op.-norm-matched (5e-6)	0.6B	1	43	36	2.3×10^{-2}	2.5×10^{-2}	1	5.5×10^{-3}	6.4×10^{-4}
Polar, op.-norm-matched (5e-6)	0.6B	0	24	17	2.1×10^{-2}	3.7×10^{-1}	338	5.1×10^{-3}	9.4×10^{-3}
Polar, op.-norm-matched (5e-6)	0.6B	1	15	11	1.6×10^{-2}	3.0×10^{-1}	363	3.9×10^{-3}	7.6×10^{-3}
Polar-match (1e-5)	0.6B	0	46	39	1.0×10^{-2}	1.4×10^{-1}	247	2.4×10^{-3}	3.6×10^{-3}
Polar-match (1e-5)	0.6B	1	50	43	1.1×10^{-2}	1.5×10^{-1}	243	2.4×10^{-3}	3.7×10^{-3}
Polar-match (1e-5)	0.6B	2	46	39	9.8×10^{-3}	1.4×10^{-1}	247	2.3×10^{-3}	3.5×10^{-3}
Polar-match (7e-6)	0.6B	0	60	57	7.9×10^{-3}	1.1×10^{-1}	234	1.8×10^{-3}	2.7×10^{-3}
Polar-match (7e-6)	0.6B	1	60	54	8.6×10^{-3}	1.2×10^{-1}	225	2.0×10^{-3}	2.9×10^{-3}
Polar-match (7e-6)	0.6B	2	60	53	8.5×10^{-3}	1.2×10^{-1}	233	2.0×10^{-3}	2.9×10^{-3}
Shuffled gate (1e-6)	0.6B	0	60	54	9.8×10^{-3}	1.3×10^{-2}	2	2.3×10^{-3}	3.1×10^{-4}
AdamW (1e-5)	1.7B	0	48	41	7.1×10^{-2}	2.0×10^{-1}	12	9.7×10^{-3}	2.4×10^{-3}
AdamW (1e-5)	1.7B	1	58	44	1.1×10^{-1}	2.5×10^{-1}	8	1.4×10^{-2}	3.0×10^{-3}
AdamW (2e-5)	1.7B	0	57	43	1.6×10^{-1}	4.5×10^{-1}	8	2.2×10^{-2}	5.3×10^{-3}
AdamW (2e-5)	1.7B	1	31	18	1.3×10^{-1}	3.4×10^{-1}	10	1.7×10^{-2}	4.0×10^{-3}
Polar-match (2e-5)	1.7B	0	44	37	2.4×10^{-2}	4.1×10^{-1}	382	3.1×10^{-3}	4.9×10^{-3}
Polar-match (2e-5)	1.7B	1	51	34	3.8×10^{-2}	5.0×10^{-1}	276	4.7×10^{-3}	5.9×10^{-3}

Table 12: MATH (levels 3–5 for training, MATH-500 for evaluation), Qwen3-0.6B, 60 steps.

Arm	η	n	Δacc (step 50)	peak	collapse step	aligned norm	predicted	observed
AdamW	$2{\cdot}10^{-6}$	1	5.2	5.2	none	$1.28{\times}10^{-4}$	?	S
AdamW	$3{\cdot}10^{-6}$	2	6.5 ± 0.9	6.5	none	$1.92{\times}10^{-4}$	?/?	S/S
AdamW	$5{\cdot}10^{-6}$	3	-40.3 ± 2.0	6.6	37/36/33	$3.16{\times}10^{-4}$	C/C/C	C/C/C
Polar-match	$5{\cdot}10^{-6}$	3	5.0 ± 0.8	6.3	none	$1.17{\times}10^{-4}$	S/S/S	S/S/S
Polar-match	$7{\cdot}10^{-6}$	2	2.7 ± 0.3	4.3	57/none	$1.64{\times}10^{-4}$	?/?	C/S
Polar-match	10^{-5}	1	-8.6	6.6	49	$2.33{\times}10^{-4}$	C	C
Edge-truncated	$2{\cdot}10^{-6}$	1	-24.8	4.6	24	$5.29{\times}10^{-4}$	C	C
Edge-truncated	$5{\cdot}10^{-6}$	3	-39.9 ± 2.8	0.0	24/17/27	$8.40{\times}10^{-4}$	C/C/C	C/C/C

Table 13: GSM8K, Qwen3-0.6B, 200 steps (evaluation every 20 steps).

Arm	η	n	Δacc (final step)	peak	collapse step	aligned norm	predicted	observed
AdamW	$2{\cdot}10^{-6}$	2	-60.0 ± 2.8	11.8	86/83	$1.45{\times}10^{-4}$	–	C/C
Polar-match	$5{\cdot}10^{-6}$	2	-52.3 ± 6.1	10.5	77/65	$1.18{\times}10^{-4}$	–	C/C
Polar-match	$6{\cdot}10^{-6}$	1	-52.6	10.8	78	$1.43{\times}10^{-4}$	–	C

Table 14: The per-matrix aligned-norm comparison beyond Qwen3-0.6B (GSM8K, 60 steps; registered as AL1–AL3). "Aligned-matched": every matrix step rescaled to the input-aligned norm of its own Adam step, with every parameter at the stated learning rate; the AdamW and RMS-matched rows are the runs of the main scans at the same size and learning rate.

Model	Arm	η	collapsed	collapse steps
Qwen3-1.7B	AdamW	10^{-5}	3 of 3	41/44/41
Qwen3-1.7B	Polar, RMS-matched	10^{-5}	0 of 3	–/–/–
Qwen3-1.7B	Polar, aligned-matched	10^{-5}	3 of 3	17/13/45
Qwen3-1.7B	Polar, aligned-matched	$5{\cdot}10^{-6}$	1 of 3	41/–/–
Qwen3-4B	AdamW	10^{-5}	1 of 3	–/39/–
Qwen3-4B	Polar, RMS-matched	10^{-5}	0 of 3	–/–/–
Qwen3-4B	Polar, aligned-matched	10^{-5}	3 of 3	32/35/22

Table 15: The flat polar direction with its matrix steps matched per matrix to the input-aligned norm of its own Adam step, every parameter at AdamW's learning rate, against AdamW in the same setting: runs that collapse (with the KL penalty: that first exceed a per-token KL of 1), and in parentheses the median collapse step ("–": fewer than half collapse); every run's step is in Table 16. "KL": penalty $\beta = 0.04$; "clip": gradient clip 1.0; "KL-matched": the matched step times 0.88, a constant derived from the one-step KL of nine saved states (the runs themselves need 0.62–0.89, see below); "policy-matched": the matrix steps rescaled at every step by one scalar so that their measured one-step KL equals that of the Adam step from the same state; "Muon mom.": the polar factor of Muon's own momentum ($\beta = 0.95$, no bias correction, Nesterov input) instead of the Adam momentum. Every row is a registered test (Appendix E.7); only at Qwen3-4B, $5{\cdot}10^{-6}$ (six runs) does the aligned-matched flat direction collapse less often than AdamW. The policy-matched scale is fitted at every step from two KL measurements; the rule and its one registered change are in Appendix E.7. Llama rows use, for both arms alike, the rule registered for that model: the first of reward below half its step-11–20 mean and the absorbing state (reward ≤ 0.02 with $\geq 90\%$ truncated responses for five steps), so a collapse can precede step 11.

Model	steps	η	aligned-matched	KL-matched	policy-matched	Muon mom.	AdamW
Llama 1B	60	$2{\cdot}10^{-5}$	3/3 (17)		3/3 (29)		3/3 (21)
0.6B	60	$2{\cdot}10^{-6}$	2/5 (–)		0/3 (–)		0/10 (–)
0.6B	60	$2.5{\cdot}10^{-6}$	5/5 (45)	4/5 (54)	8/32 (–)	2/3 (52)	5/35 (–)
0.6B	60	$3{\cdot}10^{-6}$	3/3 (42)	3/3 (54)	26/32 (55.5)	3/3 (40)	26/35 (57)
0.6B	200	$2{\cdot}10^{-6}$	3/3 (59)				2/2 (84.5)
0.6B	200, KL	$5{\cdot}10^{-6}$	3/3 (48)		3/3 (62)		3/3 (66)
0.6B	200, KL, clip	$5{\cdot}10^{-6}$	3/3 (50)		3/3 (62)		3/3 (63)
0.6B, MATH	60	$3{\cdot}10^{-6}$	3/3 (52)		1/3 (–)		0/3 (–)
0.6B, MATH	60	$5{\cdot}10^{-6}$	3/3 (36)		3/3 (38)		3/3 (36)
0.6B, frozen	60	$3{\cdot}10^{-6}$	4/4 (45.5)		3/4 (56)		4/4 (54)
0.6B, frozen	60	$5{\cdot}10^{-6}$	4/4 (26)		4/4 (33)		4/4 (36)
1.7B	60	$5{\cdot}10^{-6}$	1/3 (–)				0/2 (–)
1.7B	60	$1{\cdot}10^{-5}$	3/3 (17)		2/3 (58)		3/3 (41)
4B	60	$5{\cdot}10^{-6}$	0/3 (–)				2/3 (56)
4B	60	$1{\cdot}10^{-5}$	3/3 (32)				3/5 (39)

Table 16: Table 15 run by run: the collapse step of every run, in seed order ("–": no collapse within the horizon).

Model	steps	η	polar, matched	KL-matched	policy-matched	Muon mom., matched	AdamW
Llama 1B	60	$2{\cdot}10^{-5}$	3/3 (6/17/19)		3/3 (29/36/21)		3/3 (30/21/21)
0.6B	60	$2{\cdot}10^{-6}$	2/5 (–/–/–/49/58)		0/3 (–/–/–)		0/10 (–/–/–/–/–/–/–/–/–/–)
0.6B	60	$2.5{\cdot}10^{-6}$	5/5 (50/41/59/45/43)	4/5 (–/55/54/45/53)	8/32 (53/–/–/52/58/–/–/–/–/–/–/–/–/–/55/–/–/–/60/–/–/56/–/–/–/–/60/57/–/–/–/–)	2/3 (–/52/50)	5/35 (–/–/–/–/58/–/58/–/–/–/–/–/–/–/–/–/–/–/–/–/57/–/–/–/–/–/–/–/59/60/–/–/–/–)
0.6B	60	$3{\cdot}10^{-6}$	3/3 (47/42/40)	3/3 (54/58/43)	26/32 (–/58/60/44/44/60/51/52/51/–/58/56/60/–/49/60/57/50/52/60/52/–/57/51/52/–/51/49/48/55/–/49)	3/3 (40/39/41)	26/35 (56/60/52/52/56/–/51/55/58/–/56/53/–/56/58/–/56/54/59/59/53/57/58/–/57/–/–/–/53/51/54/59/51/58/–)
0.6B	200	$2{\cdot}10^{-6}$	3/3 (72/56/59)				2/2 (86/83)
0.6B	200, KL	$5{\cdot}10^{-6}$	3/3 (51/44/48)		3/3 (62/58/64)		3/3 (62/66/73)
0.6B	200, KL, clip	$5{\cdot}10^{-6}$	3/3 (50/45/50)		3/3 (63/61/62)		3/3 (58/63/67)
0.6B, MATH	60	$3{\cdot}10^{-6}$	3/3 (51/53/52)		1/3 (–/–/53)		0/3 (–/–/–)
0.6B, MATH	60	$5{\cdot}10^{-6}$	3/3 (36/30/38)		3/3 (42/38/38)		3/3 (37/36/33)
0.6B, frozen	60	$3{\cdot}10^{-6}$	4/4 (46/47/45/35)		3/4 (–/51/55/57)		4/4 (60/54/51/54)
0.6B, frozen	60	$5{\cdot}10^{-6}$	4/4 (31/27/25/20)		4/4 (39/33/33/29)		4/4 (40/34/38/34)
1.7B	60	$5{\cdot}10^{-6}$	1/3 (41/–/–)				0/2 (–/–)
1.7B	60	$1{\cdot}10^{-5}$	3/3 (17/13/45)		2/3 (28/–/58)		3/3 (41/44/41)
4B	60	$5{\cdot}10^{-6}$	0/3 (–/–/–)				2/3 (56/41/–)
4B	60	$1{\cdot}10^{-5}$	3/3 (32/35/22)				3/5 (–/39/–/21/29)

Table 17: Residual-scale test (RS1, RS2) and the policy-matched arm (KP) at the boundary cell, Qwen3-0.6B, GSM8K, 60 steps, $\eta = 2.5{\cdot}10^{-6}$: runs that collapse and when. Rows other than AdamW are the polar direction matched per matrix to the aligned norm of its own Adam step, times the constant shown; the last row is rescaled at every step to AdamW's measured one-step KL. All rows registered (Appendix E.7).

Arm at $2.5{\cdot}10^{-6}$	collapsed	collapse steps
AdamW	5/35	–/–/–/–/58/–/58/–/–/–/–/–/–/–/–/–/–/–/–/–/–/57/–/–/–/–/–/–/–/–/59/60/–/–/–/–
polar, aligned-matched $\times 1.0$	5/5	50/41/59/45/43
polar, aligned-matched $\times 0.88$ (KL-matched)	4/5	–/55/54/45/53
polar, aligned-matched $\times 0.71$	0/3	–/–/–
polar, aligned-matched $\times 0.50$	0/3	–/–/–
polar, policy-matched (every step)	8/32	53/–/–/52/58/–/–/–/–/–/–/–/–/–/55/–/–/–/60/–/–/56/–/–/–/–/60/57/–/–/–/–

Table 18: Learning-rate scan on Llama-3.2-1B-Instruct (GSM8K, 60 steps, the protocol of Section 5.1). "Registered rule": first step after step 20 with the training reward below half of its mean over steps 11–20, fixed before the runs because this model spends its first ten steps learning the answer format; "absorbing state": first of five consecutive steps with reward ≤ 0.02 and $\geq 90\%$ truncated responses. The grid up to 10^{-5} is registered (F1, F2, which fail: no threshold is bracketed); the rows at $2{\cdot}10^{-5}$ and $4{\cdot}10^{-5}$ were added after that outcome and are exploratory. Their AdamW runs use micro-batches of 4 instead of 8: the first executions ran out of memory once every response was truncated at 1024 tokens, after showing the same collapses.

Arm	η	grid	collapsed	registered rule	absorbing state
AdamW	$2{\cdot}10^{-6}$	registered	0 of 2	–/–	–/–
AdamW	$5{\cdot}10^{-6}$	registered	0 of 2	–/–	–/–
AdamW	$1{\cdot}10^{-5}$	registered	0 of 2	–/–	–/–
AdamW	$2{\cdot}10^{-5}$	exploratory	3 of 3	30/21/21	36/36/–
AdamW	$4{\cdot}10^{-5}$	exploratory	2 of 2	–/–	3/4
Polar-match	$2{\cdot}10^{-6}$	registered	0 of 2	–/–	–/–
Polar-match	$5{\cdot}10^{-6}$	registered	0 of 2	–/–	–/–
Polar-match	$1{\cdot}10^{-5}$	registered	0 of 2	–/–	–/–
Polar-match	$2{\cdot}10^{-5}$	exploratory	0 of 3	–/–/–	–/–/–
Polar-match	$4{\cdot}10^{-5}$	exploratory	1 of 2	–/54	–/–

Table 19: The main figure under a realistic recipe: 200 steps, KL penalty $\beta = 0.04$ and a global gradient clip of 1.0, three seeds per arm (registered as RC1, RC2). "Without clipping": the runs of Figure 6. "Steps with the clip active": share of steps whose pre-clip gradient norm exceeds 1.0.

Arm ($\eta = 5{\cdot}10^{-6}$)	first step with KL > 1	without clipping	steps with the clip active
AdamW	58/63/67	62/66/73	29%
Polar, RMS-matched	144/149/129	146/135/147	20%
Polar, aligned-matched	50/45/50	51/44/48	19%

Table 20: Gradient clipping at $\eta = 5{\cdot}10^{-6}$ on Qwen3-0.6B (GSM8K, 60 steps, registered as GC1 and GC2). "Steps with the clip active": the fraction of steps whose pre-clip global gradient norm exceeds the clip. Without clipping, AdamW collapses in 3 of 3 runs at steps 35/34/32 and polar-match in none of 3.

Arm	clip	n	steps with the clip active	collapsed	collapse steps
AdamW	1	2	4%	2 of 2	47/35
Polar-match	1	2	0%	0 of 2	–/–
AdamW	0.15	2	40%	2 of 2	40/33
Polar-match	0.15	2	2%	0 of 2	–/–

Table 21: Held-out directions $US^pV^\top$ of the bias-corrected momentum (RMS-matched, GSM8K, Qwen3-0.6B, 60 steps). Columns as in Table 12; "predicted" uses the registered aligned-norm band.

Arm	η	n	Δacc (final step)	peak	collapse step	aligned norm	predicted	observed
$US^{0.5}V^\top$	$1.2{\cdot}10^{-6}$	1	10.8	11.1	59	$1.25{\times}10^{-4}$	S	C
$US^{0.5}V^\top$	$2.4{\cdot}10^{-6}$	3	-63.9 ± 0.1	9.7	38/34/35	$2.43{\times}10^{-4}$	C/C/C	C/C/C
$US^{0.5}V^\top$	$4{\cdot}10^{-6}$	1	-64.8	5.7	28	$3.91{\times}10^{-4}$	C	C
$US^{0.25}V^\top$	$4{\cdot}10^{-6}$	1	0.3	9.9	50	$1.88{\times}10^{-4}$	?	C
$US^{0.25}V^\top$	$6{\cdot}10^{-6}$	2	-62.5 ± 1.4	9.4	29/30	$2.80{\times}10^{-4}$	C/C	C/C

Table 22: GSM8K, Qwen3-0.6B, 60 steps, with a per-token k3 KL penalty ($\beta = 0.04$) to the frozen initial policy (bfloat16). Columns as in Table 12.

Arm	η	n	Δacc	KL at step 50	max KL	first step with KL > 1
AdamW	$5{\cdot}10^{-6}$	2	7.9 ± 0.1	0.034	0.12	none/none
AdamW	10^{-5}	3	1.8 ± 1.0	1.301	9.04	29/33/33
AdamW	$2{\cdot}10^{-5}$	1	1.1	3.264	7.61	19
Polar-match	$5{\cdot}10^{-6}$	2	6.9 ± 1.1	0.014	0.02	none/none
Polar-match	10^{-5}	3	7.9 ± 0.4	0.039	0.07	none/none/none
Polar-match	$2{\cdot}10^{-5}$	2	0.7 ± 1.3	5.230	8.56	39/34

Table 23: GSM8K, Qwen3-0.6B, 200 steps with the KL penalty of Table 22. "first step with KL > 1": first step at which the per-token KL to the initial policy exceeds 1, per run.

Arm	η	n	Δacc	KL at step 50	max KL	first step with KL > 1
AdamW	$1.7{\cdot}10^{-6}$	4	6.9 ± 1.0	0.006	0.40	none/none/none/none
AdamW	$5{\cdot}10^{-6}$	3	6.0 ± 0.8	0.034	7.80	62/66/73
Polar-match	$5{\cdot}10^{-6}$	3	5.4 ± 0.9	0.014	7.92	146/135/147
Polar, aligned-matched	$5{\cdot}10^{-6}$	3	-2.6 ± 10.5	1.649	17.20	51/44/48
Polar-match	$1.5{\cdot}10^{-5}$	3	-36.0 ± 22.3	2.496	60.04	50/48/48

Table 24: Exploratory (not registered) runs of the held-out direction $US^{0.5}V^\top$, added after the registered control collapsed, to locate its threshold. Columns as in Table 12.

Arm	η	n	Δacc (final step)	peak	collapse step	aligned norm	predicted	observed
$US^{0.5}V^\top$	$8{\cdot}10^{-7}$	2	9.2 ± 0.2	9.8	none	$8.28{\times}10^{-5}$	S/S	S/S
$US^{0.5}V^\top$	$1.2{\cdot}10^{-6}$	1	10.1	10.1	57	$1.22{\times}10^{-4}$	S	C

Table 25: Qwen3-4B: the registered cross-scale prediction, run by run. The half-probability aligned norm is the one extrapolated from Qwen3-0.6B and Qwen3-1.7B, $1.86{\cdot}10^{-3}$, not refitted; "band" is the registered factor-two interval around it, inside which a run counts as undecided. GSM8K is nearly saturated for this model (92.2% accuracy before training), so these runs test stability rather than accuracy gains.

Direction	η	a/η	aligned norm	band	predicted	outcome
AdamW	$2{\cdot}10^{-6}$	153	$0.31{\cdot}10^{-3}$	below	stable	stable
AdamW	10^{-5}	173	$1.73{\cdot}10^{-3}$	in band	stable	stable
AdamW	$2{\cdot}10^{-5}$	188	$3.75{\cdot}10^{-3}$	above	collapse	step 20
AdamW	$5{\cdot}10^{-5}$	161	$8.06{\cdot}10^{-3}$	above	collapse	step 11
polar	$5{\cdot}10^{-6}$	38	$0.19{\cdot}10^{-3}$	below	stable	stable
polar	10^{-5}	38	$0.38{\cdot}10^{-3}$	below	stable	stable
polar	$2{\cdot}10^{-5}$	38	$0.77{\cdot}10^{-3}$	below	stable	step 26
polar	$8{\cdot}10^{-5}$	50	$3.99{\cdot}10^{-3}$	above	collapse	step 25
polar	$2{\cdot}10^{-4}$	44	$8.72{\cdot}10^{-3}$	above	collapse	step 11

Table 26: Held-out log-loss per direction family (fit on all other families; lower is better) and, below, the half-probability aligned norm of each direction from the fit with a shared slope and direction offsets (Qwen3-0.6B, GSM8K, 60 steps).

Held-out family	AdamW	Polar	Edge-truncated	Pion filter
Learning rate only	0.66	0.89	0.68	0.54
Operator norm	0.92	0.61	0.55	0.55
Geometric mean of RMS and operator norm	0.80	0.58	0.54	0.48
Input-aligned norm	0.66	0.48	0.48	0.48
Base rate	0.70	0.69	0.69	0.70
Half-probability aligned norm (per-direction offsets)	2.02×10^{-4}	1.43×10^{-4}	1.60×10^{-4}	2.02×10^{-4}

Table 27: Per-direction aligned norm per unit learning rate a/η (mean over runs, with its coefficient of variation), the learning rate $\eta_{1/2}$ at which the pooled logistic fit of Figure 7b predicts collapse probability $1/2$, and the observed number of collapsed runs within 60 steps at each learning rate. Counts are over the runs that log the aligned norm (the fit's sample), which is a subset of the runs of Table 1 and does not include the seeds added later for Table 15.

Direction	runs	a/η (CV)	predicted $\eta_{1/2}$	collapsed runs / runs, by learning rate
Polar	23	24 (0.9%)	$6.3{\cdot}10^{-6}$	10^{-6}: 0/1, $2{\cdot}10^{-6}$: 0/3, $3{\cdot}10^{-6}$: 0/2, $5{\cdot}10^{-6}$: 0/2, $6{\cdot}10^{-6}$: 3/6, $7{\cdot}10^{-6}$: 5/6, 10^{-5}: 3/3
AdamW	16	72 (2.1%)	$2.1{\cdot}10^{-6}$	10^{-6}: 0/1, $2{\cdot}10^{-6}$: 0/2, $2.5{\cdot}10^{-6}$: 1/6, $3{\cdot}10^{-6}$: 4/5, $5{\cdot}10^{-6}$: 2/2
Edge-truncated	16	206 (14.2%)	$7.2{\cdot}10^{-7}$	$3{\cdot}10^{-7}$: 0/3, $5{\cdot}10^{-7}$: 0/1, $7{\cdot}10^{-7}$: 3/6, 10^{-6}: 3/4, $5{\cdot}10^{-6}$: 2/2
Pion filter	14	341 (1.9%)	$4.4{\cdot}10^{-7}$	$3{\cdot}10^{-7}$: 0/3, $4{\cdot}10^{-7}$: 0/2, $5{\cdot}10^{-7}$: 0/3, $7{\cdot}10^{-7}$: 3/3, 10^{-6}: 1/1, $5{\cdot}10^{-6}$: 2/2

Table 28: Pooled logistic fits over all runs, in the logarithm of one step size (steps 1–10). Errors: runs on the wrong side of probability $1/2$. Leave one direction family out: fit on the other families and score the held-out runs; confident runs are held-out runs predicted with probability below 0.1 or above 0.9.

	all runs		leave one direction family out		
Step size (steps 1–10)	log-loss	errors	log-loss	confident errors	confident runs
Learning rate only	0.59	32/86	0.68	1	12
Operator norm	0.58	29/86	0.66	1	10
Geometric mean of RMS and operator norm	0.58	28/86	0.72	5	12
Input-aligned norm	0.52	22/86	0.71	4	18

Table 29: The runs behind the best-rate comparison (BB) and the public-code residuals (AP1): Qwen3-0.6B, GSM8K, 60 steps, one run per seed, mean GSM8K test accuracy (%) over all seeds of the arm, collapsed runs included. Registered (Appendix E.7).

Arm	seeds	collapsed	accuracy at step 50	at step 60
AdamW, $2{\cdot}10^{-6}$	8	0	74.6	75.6
polar-match, $5{\cdot}10^{-6}$	8	0	75.7	75.7
public Pion, $\kappa = 5$, 10^{-6}	8	0	72.3	72.9
public per-head Pion, $\kappa = 5$, 10^{-6}	8	3	73.1	69.9

Table 30: Registered cells reported in the text: the practitioner recipe (PR1; warmup over 10 steps, cosine decay to $0.1\times$, KL penalty 0.04, clip 1.0, 200 steps), public Pion at two further rates (PB1) and the Qwen3-1.7B bracket (B17). "KL onset": runs whose per-token KL to the initial policy exceeds 1. Final accuracy: last evaluation, mean over all runs of the cell, collapsed runs included.

Cell	runs	collapsed (steps)	KL onset	final accuracy (s.e.)
AdamW, $2{\cdot}10^{-6}$, practitioner recipe (200 steps)	8	0 (–/–/–/–/–/–/–/–)	0	71.6 (0.2)
polar-match, $5{\cdot}10^{-6}$, practitioner recipe (200 steps)	8	0 (–/–/–/–/–/–/–/–)	0	72.3 (0.3)
public Pion, $\kappa = 5$, $1.5{\cdot}10^{-6}$	3	0 (–/–/–)	0	73.5 (0.2)
public Pion, $\kappa = 5$, $2{\cdot}10^{-6}$	3	1 (–/–/56)	0	69.3 (2.9)
Qwen3-1.7B, AdamW, $7{\cdot}10^{-6}$	5	3 (49/–/–/35/32)	0	39.7 (19.7)
Qwen3-1.7B, polar-match, $1.4{\cdot}10^{-5}$	5	3 (–/54/39/36/–)	0	41.7 (18.6)

Table 31: Accuracy gain over 60 steps at a matched aligned norm (Qwen3-0.6B, GSM8K, stable runs with aligned norm in $[1.3{\cdot}10^{-4}, 2.0{\cdot}10^{-4}]$; mean and standard error over runs). Only runs that stay stable have an end-of-training accuracy, so this sample is conditioned on stability, and each direction's largest aligned norm here is the one closest to its own threshold.

Direction	runs	aligned norm	accuracy gain (points)
AdamW	7	$1.73{\cdot}10^{-4}$	$+11.09 \pm 0.60$
Pion filter	5	$1.56{\cdot}10^{-4}$	$+9.52 \pm 0.50$
edge-truncated	3	$1.62{\cdot}10^{-4}$	$+9.35 \pm 0.40$
polar	4	$1.48{\cdot}10^{-4}$	$+9.25 \pm 0.96$

Table 32: Cox proportional-hazards fits of the time to collapse on one step size (Qwen3-0.6B, GSM8K, 69 runs, 32 collapses, the rest right-censored). Hazard ratio per decade of the step size; p from the likelihood-ratio test against the null; concordance is Harrell's C.

Step size (steps 1–10)	hazard ratio per decade	p	concordance
Learning rate only	5	$3{\cdot}10^{-4}$	0.70
Operator norm	42	$6{\cdot}10^{-8}$	0.72
Geometric mean of RMS and operator norm	275	$2{\cdot}10^{-10}$	0.81
Input-aligned norm	35870	$2{\cdot}10^{-16}$	0.85

Table 33: Per state: the factor spanned by the directions' one-step KL at an equal step size, and the KL of each direction relative to the Adam step at the same learning rate (geometric mean over the four rates).

KL at equal step spans a factor of	seed 0, step 2	seed 0, step 20	seed 0r, step 40	seed 1, step 2	seed 1, step 20	seed 1, step 40	seed 2, step 2	seed 2, step 20	seed 2, step 40
RMS (= learning rate)	135.3	337.6	765.5	48.5	157.3	224.2	140.1	129.3	138.3
operator norm	10.1	19.6	31.9	15.6	11.1	11.4	10.8	9.2	30.0
input-aligned norm	2.6	3.4	3.6	4.5	2.6	2.1	2.6	2.5	6.4
KL relative to Adam at the same learning rate									
momentum	8.34	13.82	25.53	4.56	7.15	9.64	8.53	7.17	5.03
flat, Adam basis	0.07	0.04	0.03	0.09	0.05	0.05	0.07	0.06	0.04
flat, momentum basis	0.14	0.09	0.08	0.18	0.10	0.11	0.15	0.12	0.07
polar (NS5)	0.14	0.09	0.08	0.18	0.10	0.11	0.14	0.12	0.07
$US^{0.5}V^\top$	2.21	1.86	3.73	1.69	1.81	2.17	2.22	1.91	1.41
Pion filter	8.81	13.39	18.99	3.66	7.00	11.10	9.12	7.48	5.10
rank one	8.81	13.39	18.99	3.66	7.00	11.10	9.12	7.48	5.10

Table 34: As Table 33, for Llama-3.2-1B-Instruct.

KL at equal step spans a factor of	seed 0, step 2	seed 0, step 20	seed 1, step 2	seed 1, step 20	seed 2, step 2	seed 2, step 20
RMS (= learning rate)	39.8	454.4	31.7	108.9	43.5	5688.3
operator norm	6.4	1247.8	7.1	42.7	5.9	1208.7
input-aligned norm	2.1	49.7	2.3	4.9	2.5	21.8
KL relative to Adam at the same learning rate						
momentum	2.61	2.40	2.57	2.40	2.52	6.79
flat, Adam basis	0.07	0.01	0.08	0.03	0.06	0.00
flat, momentum basis	0.10	0.01	0.13	0.04	0.10	0.00
polar (NS5)	0.11	0.01	0.14	0.04	0.11	0.00
$US^{0.5}V^\top$	1.05	0.64	1.17	0.61	1.15	0.19
Pion filter	2.34	1.25	2.15	2.79	2.20	5.51
rank one	2.34	1.25	2.15	2.79	2.20	5.51

Table 35: As Table 33, for Qwen3-1.7B (AdamW at $5{\cdot}10^{-6}$).

KL at equal step spans a factor of	seed 0, step 2	seed 0, step 20	seed 1, step 2	seed 1, step 20	seed 2, step 2	seed 2, step 20
RMS (= learning rate)	104.2	161.6	280.4	318.0	122.1	622.7
operator norm	13.8	14.4	10.1	20.7	17.3	56.5
input-aligned norm	3.6	3.9	2.8	3.8	4.2	8.4
KL relative to Adam at the same learning rate						
momentum	6.51	6.64	10.72	12.99	7.48	17.59
flat, Adam basis	0.06	0.04	0.04	0.04	0.06	0.03
flat, momentum basis	0.13	0.09	0.11	0.09	0.13	0.07
polar (NS5)	0.13	0.09	0.12	0.09	0.14	0.06
$US^{0.5}V^{\top}$	2.05	1.85	2.91	2.29	2.24	3.09
Pion filter	6.19	5.26	11.92	12.28	6.51	16.55
rank one	6.19	5.26	11.92	12.28	6.51	16.55

Table 36: Aligned-norm matching matrix by matrix at AdamW's learning rate, against the references of the main scan (Qwen3-0.6B, GSM8K, 60 steps). Aligned norm: steps 1–10, probe matrices; gain over stable runs.

Arm	η	aligned norm	collapsed	collapse steps	gain (points)
AdamW (reference)	$2{\cdot}10^{-6}$	$1.48{\cdot}10^{-4}$	0 of 2	–/–	$+10.4$
AdamW (reference)	$3{\cdot}10^{-6}$	$2.15{\cdot}10^{-4}$	4 of 5	60/52/52/56/–	$+11.1$
polar, RMS-matched (reference)	$6{\cdot}10^{-6}$	$1.43{\cdot}10^{-4}$	3 of 6	–/60/–/54/58/–	$+9.8$
polar, aligned-matched	$2{\cdot}10^{-6}$	$1.44{\cdot}10^{-4}$	0 of 3	–/–/–	$+11.0$
polar, aligned-matched	$3{\cdot}10^{-6}$	$2.11{\cdot}10^{-4}$	3 of 3	47/42/40	–
Pion filter, aligned-matched	$2{\cdot}10^{-6}$	$1.56{\cdot}10^{-4}$	2 of 3	53/–/56	$+11.2$

Table 37: Where the account stops. "Registered" marks a prediction or control written down, with its refutation rule, before the runs or states it concerns (Appendix E.7).

Where	What happens	
Qwen3-4B	The threshold extrapolated from the smaller models fails: the constants differ fourfold, the thresholds do not	registered, Table 25
200 steps, no KL penalty	Every run collapses, and the aligned norm does not order the collapse times	registered, Table 13
An arbitrary rotation	A Haar-random basis at the same aligned norm changes the policy 743.1–4112.2 times less: the norm bounds gradient-derived steps only	registered, Appendix E.5
Llama-3.2-1B	One step: holds at step 2, in one of three states at step 20. Training: the ordering appears only above the registered grid	registered, Tables 18 and 34
Public Pion filters, predicted	Threshold placed too high, by 1.4–$2.8\times$ (per-head) and 3.6–$11\times$ (whole-matrix)	registered, Section 5.4
Qwen3-1.7B, matched per matrix	The flat direction also collapses at $5{\cdot}10^{-6}$, where AdamW does not: a residual, in AdamW's favour	registered, Table 14
Held out on AdamW; MATH	Log-loss 0.66 against a base rate of 0.70; on MATH learning-rate thresholds predict as well	Table 26, Section 5.6
A threshold, not an ordering	No aggregation of the aligned norm separates stable from collapsed runs	Table 39
Accuracy line	Public Pion ($\kappa = 5$) gains $+2.3$ points more than the line gives at its displacement, which lies below the fitted range; polar-match at $5{\cdot}10^{-6}$ sits $+1.9$ above it with new seeds	registered, Figure 8

Table 39: Ten run-level quantities built from the input-aligned norm, on the 86 Qwen3-0.6B runs that log it. "ratio" is the smallest collapsed value divided by the largest stable one; a value above 1 would mean the two outcomes separate. ρ is the Spearman correlation between the quantity and the collapse step over the collapsed runs.

quantity	largest stable	smallest collapsed	ratio	separates	ρ
per-step aligned norm, steps 1–10	2.2×10^{-4}	4.3×10^{-5}	0.20	no	-0.63
per-step aligned norm, before collapse	1.4×10^{-4}	9.0×10^{-5}	0.63	no	-0.82
aligned drift at the collapse step	5.0×10^{-3}	3.8×10^{-3}	0.75	no	-0.57
aligned drift at step 10	1.7×10^{-3}	3.9×10^{-4}	0.23	no	-0.62
aligned drift at step 20	2.3×10^{-3}	7.9×10^{-4}	0.34	no	-0.69
aligned drift growth, steps 5–15	1.1×10^{-3}	4.4×10^{-4}	0.40	no	-0.67
anisotropy ratio alone	16.36	1.17	0.07	no	-0.08
per-step norm $\times$ accumulation coherence	1.7×10^{-4}	3.9×10^{-5}	0.23	no	-0.62
aligned drift coherence, steps 1–10	0.91	0.63	0.69	no	0.40
log-probability gap at step 10 (control)	1.3×10^{-2}	1.0×10^{-2}	0.83	no	-0.49

Table 40: Gate probe on Qwen3-0.6B at $\eta = 5{\cdot}10^{-6}$, RMS matched. Δacc and collapse step as in Table 1; σ_t and k_t: mean gate value and number of kept singular directions (including matrix-steps with $k_t = 0$); $k_t = 0$ and $\sigma_t = 0$: percentage of matrix-steps in which no singular value exceeds the edge, and in which the gate selects the Adam direction.

Arm	n	Δacc	collapse step	σ_t	k_t	$k_t = 0$ (%)	$\sigma_t = 0$ (%)
Gate (NCSH)	3	-63.8 ± 0.0	14/17/18	0.30	0.6	46	64
Edge-truncated polar	3	-62.8 ± 1.5	16/11/16	–	0.7	39	–
AdamW	3	-63.8 ± 0.0	35/34/32	–	–	–	–
Polar-match	3	11.4 ± 0.5	none	–	–	–	–