FROM ROTATION TO CONTROL: ADAPTIVE PRECONDITIONING OF VERIFIER-DERIVED UPDATES

Anonymous authors
Paper under double-blind review

ABSTRACT

Adaptive optimizers do more than change step size: coordinate-wise scaling can turn a learning signal before it reaches the parameters. In Qwen2.5-7B LoRA verifier tasks, momentum and decoupled decay contribute almost no rotation, while second-moment scaling turns the update by 59.95° on average across eight tasks. We show that this angle has a simple exact form. If a vector m is scaled by positive coordinate factors a, then $\tan\theta = \mathrm{CV}_w(a)$, where $w_i \propto m_i^2$; rotation is exactly the weighted anisotropy of the scale on coordinates carrying update energy. This identity motivates rotation-budgeted preconditioning, which dynamically chooses the strongest preconditioning exponent whose direction remains within a target angle. During 384-step Qwen training, full adaptive candidates remain rotated by roughly 68°–75°, while a 25° budget adjusts the exponent online and holds the applied update at 25.00°. We then test the geometry on Mistral-7B with conventional code SFT, without a verifier-derived training objective. Second-moment scaling introduces an additional 64.5° turn, the exact law holds to machine precision, and a smooth p sweep again controls the angle. Crucially, fixed $p = 0.4$ drifts by up to roughly 10° across seeds as optimizer state evolves, whereas rotation budgeting remains at 25.000000° for all 1152 training steps. Adaptive preconditioning is therefore a geometric transformation whose magnitude can be measured, predicted, and controlled directly.

1 INTRODUCTION

A learning objective provides a direction, but that direction is not necessarily the update the model receives. Between the gradient and the new parameters sits the optimizer.

The distinction is geometric. Multiplying every coordinate by the same number changes an arrow's length but not its direction. Adaptive optimizers scale coordinates by different amounts. Draw an arrow on graph paper and stretch the horizontal axis more than the vertical one: the arrow is not merely longer; it points somewhere else. Adam's second-moment division performs this high-dimensional operation at every step.

Verifier-based post-training makes the distinction especially visible. A verifier supplies an unusually concrete objective, making it tempting to interpret the objective-derived gradient as "the direction" of learning. In our Qwen2.5-7B LoRA experiments, momentum barely changes that direction, but second-moment scaling turns it by roughly 60°. The same component is largest in all eight valid non-degenerate verifier tasks we test.

What determines the angle? We show that coordinate-wise scaling admits an exact answer. For a vector m and positive scales a, weight coordinate i by the fraction of vector energy it carries, $w_i = m_i^2/\|m\|_2^2$. The angle θ between m and $m \odot a$ then satisfies

$$\tan\theta = \mathrm{CV}_w(a). \tag{1}$$

What matters is not whether the scale factors are large on average. It is how uneven they are on coordinates where the update carries energy. Across more than five thousand Qwen measurements, Equation 1 predicts the observed rotation to roughly one millionth of a degree.

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.

Once the angle is predictable, it can be controlled. We first use an exponent p to vary second-moment preconditioning continuously. We then replace fixed p with a rotation budget: at every step, select the strongest preconditioning whose direction remains within τ degrees of momentum. If full Adam lies inside the budget, use it. Otherwise, reduce p until the update reaches the boundary.

The distinction between these controls is important. Fixed p controls an optimizer parameter; rotation budgeting controls the geometry that parameter produces. During 384-step Qwen training, full adaptive candidates remain rotated by approximately $68°$–$75°$, while a $25°$ budget changes p_t online and keeps the realized update at $25.00°$. In this verifier setting, the budgeted arm finishes above full AdamW on the held-out P1 margin in all three seeds, although we do not interpret three runs as universal optimizer superiority.

The geometric law itself is not specific to verifiers or Qwen. On Mistral-7B with conventional completion-only code SFT, second-moment scaling adds a $64.5°$ reorientation to the current momentum direction, Equation 1 holds to machine precision, and p again gives a smooth angle sweep. More importantly, fixed $p = 0.4$ does not fix the realized geometry: its angle drifts by $-9.197°$, $+9.975°$, and $+0.224°$ across three seeds. Rotation budgeting compensates by moving p_t and stays at $25.000000°$ for all 3×384 steps.

Our contributions are threefold:

- **Exact geometry.** We identify large diagonal-preconditioning rotation in verifier-based LoRA updates and derive the exact identity $\tan\theta = \mathrm{CV}_w(a)$.
- **Direct control.** We turn the identity into rotation-budgeted preconditioning, which controls the realized angle rather than an indirect hyperparameter.
- **Training and external validation.** We show that large adaptive rotation persists over hundreds of Qwen training steps and validate the same law, continuous control, and need for dynamic angle control on Mistral-7B under standard code SFT.

2 Geometry of an Adaptive Update

Figure 1 gives the paper's geometric picture: coordinate-wise scaling turns a learning direction, p controls the strength of that turn, and an angle budget selects p_t online.

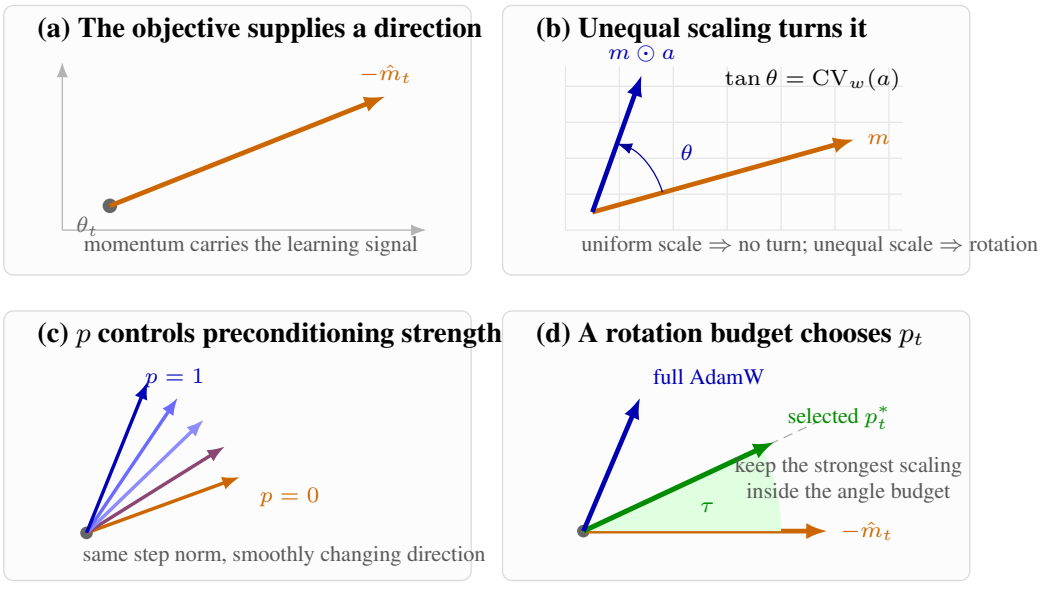

Figure 1: **From adaptive rotation to direct geometric control.** A verifier-derived momentum direction is scaled coordinate-wise by the second-moment state. The exact turn is $\tan\theta = \mathrm{CV}_w(a)$. The exponent p varies the strength of this scaling, and the rotation-budget controller chooses the largest p_t whose direction remains within τ of momentum.

2.1 COORDINATE-WISE SCALING TURNS A DIRECTION

Our primary setting uses a frozen Qwen2.5-7B base model with LoRA adapters. The LoRA parameters form the intervention and training space. Frozen comparisons begin at the same PRE checkpoint; the canonical construction task is `pva244`, and aggregate geometry uses eight valid non-degenerate verifier tasks. Section 5 then tests the same geometry on Mistral-7B with conventional code SFT. The exact verifier objective and RObE construction remain an author-supplied artifact item.

Let g denote an objective gradient and let $\hat{m}$ and $\hat{v}$ be the bias-corrected Adam moments. In the frozen Qwen comparisons, every nonzero direction d receives the same Euclidean parameter budget:

$$\Delta\theta(d) = r\frac{d}{\|d\|_2}, \qquad r = 0.13537724316120148. \tag{2}$$

Thus the comparison changes where the parameters move, not the intended distance traveled.

2.2 ROTATION EQUALS WEIGHTED ANISOTROPY

Consider a nonzero vector m and positive coordinate scales $a_i > 0$. Define the momentum-energy weights and their weighted expectation by

$$w_i = \frac{m_i^2}{\|m\|_2^2}, \qquad \mathbb{E}_w[X] = \sum_i w_i X_i, \qquad \sum_i w_i = 1. \tag{3}$$

The relevant variation in a is not its unweighted spread, but its spread on coordinates where m carries energy. We write

$$\mathrm{CV}_w(a) = \sqrt{\frac{\mathbb{E}_w[a^2]}{\mathbb{E}_w[a]^2} - 1}. \tag{4}$$

Proposition 1 (exact rotation law). Let $\theta = \angle(m, m \odot a)$. Then

$$\cos\theta = \frac{\mathbb{E}_w[a]}{\sqrt{\mathbb{E}_w[a^2]}}, \qquad \tan\theta = \mathrm{CV}_w(a). \tag{5}$$

Proof. The inner product and norm product are

$$m^\top(m \odot a) = \|m\|_2^2\mathbb{E}_w[a], \qquad \|m\|_2\|m \odot a\|_2 = \|m\|_2^2\sqrt{\mathbb{E}_w[a^2]}. \tag{6}$$

Their ratio gives the cosine. Therefore $\tan^2\theta = 1/\cos^2\theta - 1 = \mathbb{E}_w[a^2]/\mathbb{E}_w[a]^2 - 1$. Because $a_i > 0$, $\theta \in [0, \pi/2)$, so the positive square root gives Equation 5. □

The identity makes the geometry concrete. Coordinates where m_i is tiny receive almost no weight. Coordinates carrying most of the update determine the angle. If every scale equals a common constant c, then $\mathrm{CV}_w(a) = 0$ and $\theta = 0$: a scalar learning rate cannot turn the update.

2.3 A CONTINUOUS PRECONDITIONING FAMILY

We control the strength of second-moment scaling with

$$q_t(p) = -\frac{\hat{m}_t}{\hat{v}_t^{p/2} + \epsilon}, \qquad p \in [0, 1]. \tag{7}$$

For frozen interventions we apply the norm-matched displacement $\Delta\theta_p = rq_t(p)/\|q_t(p)\|_2$. At $p = 0$, the denominator is scalar and the direction is momentum; at $p = 1$, it is the usual Adam-style direction. We use $p = 1.2$ only as an extrapolation in the frozen sweep.

The family also has an idealized monotonicity property. When $\epsilon = 0$, let $Y_i = \hat{v}_i^{-1/2}$ on the support of $\hat{m}$. Then $q(p) \propto -\hat{m} \odot Y^p$.

Proposition 2 (idealized monotonicity). For $\epsilon = 0$, the angle $\theta(p) = \angle(\hat{m}, \hat{m} \odot Y^p)$ is nondecreasing in p and is strictly increasing if Y is not constant under w.

Proof sketch. Let $\phi(p) = \log \mathbb{E}_w[Y^p]$. Equation 5 gives $\log\cos\theta(p) = \phi(p) - \frac{1}{2}\phi(2p)$, hence

$$\frac{d}{dp}\log\cos\theta(p) = \phi'(p) - \phi'(2p) \le 0, \tag{8}$$

because ϕ is convex. The inequality is strict when Y varies under w. This proposition is only for $\epsilon = 0$; our implemented $\epsilon > 0$ sweeps are empirical controls, not consequences of this theorem.

2.4 From preconditioning strength to a rotation budget

Full AdamW accepts as much diagonal preconditioning as its optimizer state specifies. Our controller asks a geometric question instead: how much of that preconditioning can be retained before the update turns more than a chosen angle? If the full candidate lies inside the budget, we keep it. Otherwise, we reduce p until the direction reaches the boundary (Figure 1).

At step t, let $\theta_t(p) = \angle(-\hat{m}_t, q_t(p))$. For budget τ, choose

$$p_t^* = \max\{p \in [0,1] : \theta_t(p) \le \tau\}. \tag{9}$$

With $a_t(p) = (\hat{v}_t^{p/2} + \epsilon)^{-1}$, Proposition 1 makes the constraint equivalent to

$$\mathrm{CV}_{w_t}(a_t(p)) \le \tan\tau. \tag{10}$$

We preserve the full-AdamW adaptive-step norm while changing its direction:

$$\Delta_t^{\mathrm{RB}} = \|\Delta_t^{\mathrm{AdamW}}\|_2 \frac{q_t(p_t^*)}{\|q_t(p_t^*)\|_2}. \tag{11}$$

The verifier, learning rate, and decoupled weight-decay term are unchanged. The primary budget is $\tau = 25°$; a $20°$ run tests sensitivity. The controller adds no learned parameters.

3 The Turn Is Large in Verifier-Based Post-Training

We first ask where an adaptive update turns in the Qwen verifier setting. On the canonical `pva244` task, replacing raw descent $-g$ with the first moment $-\hat{m}$ changes the direction by only $0.02°$. Dividing by the coordinate-wise second-moment scale then adds $58.33°$. Decoupled AdamW weight decay contributes only another $0.03°$ (Figure 2a).

The same decomposition holds across eight valid non-degenerate verifier tasks. Second-moment scaling is the largest component in $8/8$ cases and accounts for 0.9996714 of the summed sequential rotation. Its mean angle is $59.9549°$, with median $60.7709°$, bootstrap 95% CI $[57.8075°, 61.4086°]$, and range $53.1128°$–$62.2057°$. By comparison, raw gradient to momentum averages $0.00742°$ (maximum $0.01978°$), while decay averages $0.01170°$ (maximum $0.03426°$). Across every task, almost the entire turn appears when the first moment is divided coordinate by coordinate by the second-moment scale.

The exact law matches the measurements. We evaluate Proposition 1 on 441 frozen $(\text{task}, p, \epsilon)$ points, 63 mechanism-control points, and 4608 real training steps. The maximum absolute prediction errors are 1.2074×10^{-6}, 1.4788×10^{-6}, and 1.708×10^{-6} degrees, respectively (Figure 2d). Across more than five thousand measurements, weighted scale anisotropy predicts the observed rotation to roughly one millionth of a degree.

p provides continuous control. On the canonical task, increasing p moves the angle through $0.0198°$, $11.8267°$, $23.5299°$, $35.1328°$, $46.6955°$, and $58.3153°$ for $p = 0, 0.2, \ldots, 1.0$; the $p = 1.2$ extrapolation reaches $69.8340°$. Across eight tasks, the $p = 0$ to $p = 1$ change has the same sign in $8/8$ cases, with mean $+59.9475°$ and bootstrap 95% CI $[57.798°, 61.388°]$. Across 63 norm-matched points, the maximum relative norm mismatch is 1.10×10^{-7}. Changing p therefore acts as a smooth geometric dial under a fixed parameter budget; Figure 2c compares this curve with the Mistral validation.

The stabilizer ϵ supplies a narrower empirical control: at $p = 1$, the tested values move the angle over $53.7617°$–$58.3351°$. This supports the same explanation—flatter coordinate scales produce less rotation—without extending Proposition 2 beyond its stated $\epsilon = 0$ scope.

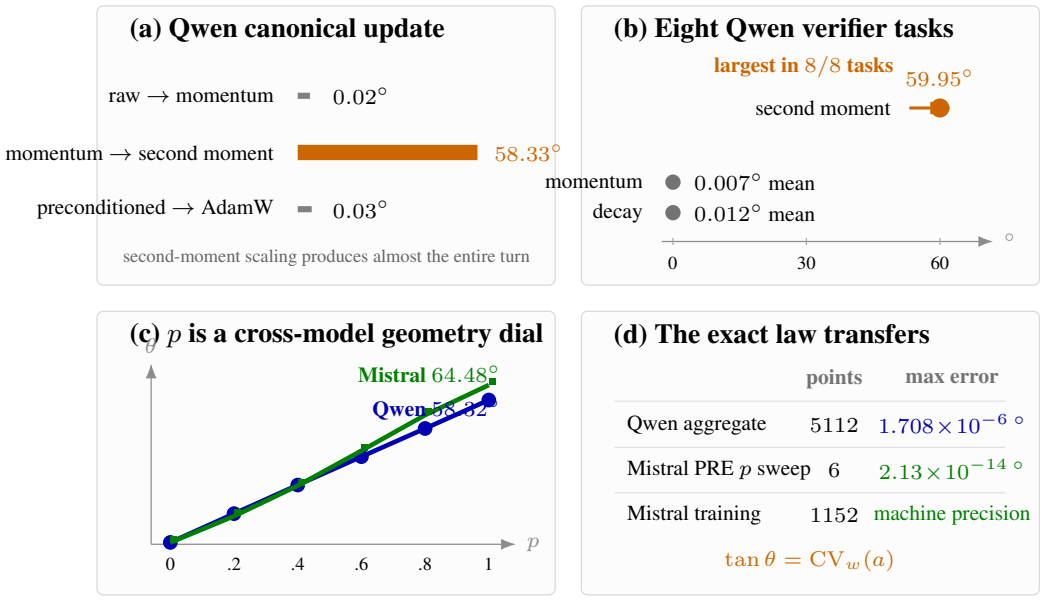

Figure 2: **The same exact geometry appears under two models and objectives.** (a–b) In the Qwen verifier setting, second-moment scaling produces almost the entire adaptive turn and is largest in all eight tasks. (c) The preconditioning exponent gives a smooth angle sweep on both Qwen and Mistral code SFT. (d) Weighted anisotropy predicts the measured rotation to numerical precision in both settings.

4 Controlling the Geometry During Qwen Training

The frozen analyses show that preconditioning can turn a verifier-derived update. We next ask whether the effect survives repeated optimization and whether the angle can be controlled online. We train four arms for 384 steps from the same PRE checkpoint: momentum geometry (P0, $p = 0$), fixed partial preconditioning (P04, $p = 0.4$), full AdamW ($p = 1$), and rotation-budgeted preconditioning (ROT_BUDGET, $\tau = 25°$). Each arm uses three seeds, for 12 primary runs. Within a seed, data, verifier, schedule, LoRA setup, evaluation plan, learning rate, and step-norm convention are shared.

Large candidate rotations persist. At every step we evaluate the full Adam-style candidate $q_t(1)$ on the current optimizer state, even when another direction is applied. Its mean rotation remains large: $75.195°$ along P0 trajectories, $73.355°$ along P04, $67.798°$ along AdamW, and $73.627°$ along ROT_BUDGET. The approximately $60°$ frozen-checkpoint effect therefore does not disappear during training; the candidate turn grows to roughly $68°$–$75°$ in these runs.

The controller enforces the requested geometry. The mean applied rotations are $0.000°$ for P0, $34.097°$ for P04, $67.798°$ for AdamW, and $25.000°$ for ROT_BUDGET (Figure 3a). The controller does not secretly reduce to a fixed arm. Across the three primary seeds, p_t ranges from approximately 0.238 to 0.424, with per-seed medians between 0.241 and 0.311; the fraction of steps at either $p_t = 0$ or $p_t = 1$ is zero. Tightening the budget to $20°$ shifts the range down to approximately 0.211–0.342, with median near 0.232, as intended.

Held-out P1 endpoints. The PRE margin is 8.10438. After 384 steps, the three-seed mean endpoints are 15.075 for ROT_BUDGET, 14.634 for P04, 12.760 for P0, and 10.061 for AdamW, corresponding to mean gains of $+6.971$, $+6.529$, $+4.656$, and $+1.957$. Every reduced-rotation arm finishes above full AdamW in all three seeds. For ROT_BUDGET, the paired within-seed advantages over AdamW are $+1.415$, $+12.237$, and $+1.390$ (mean $+5.014$).

These are descriptive results from three seeds, not a broad optimizer benchmark. In particular, ROT_BUDGET exceeds fixed P04 by only 0.442 in mean endpoint, so the runs do not establish a

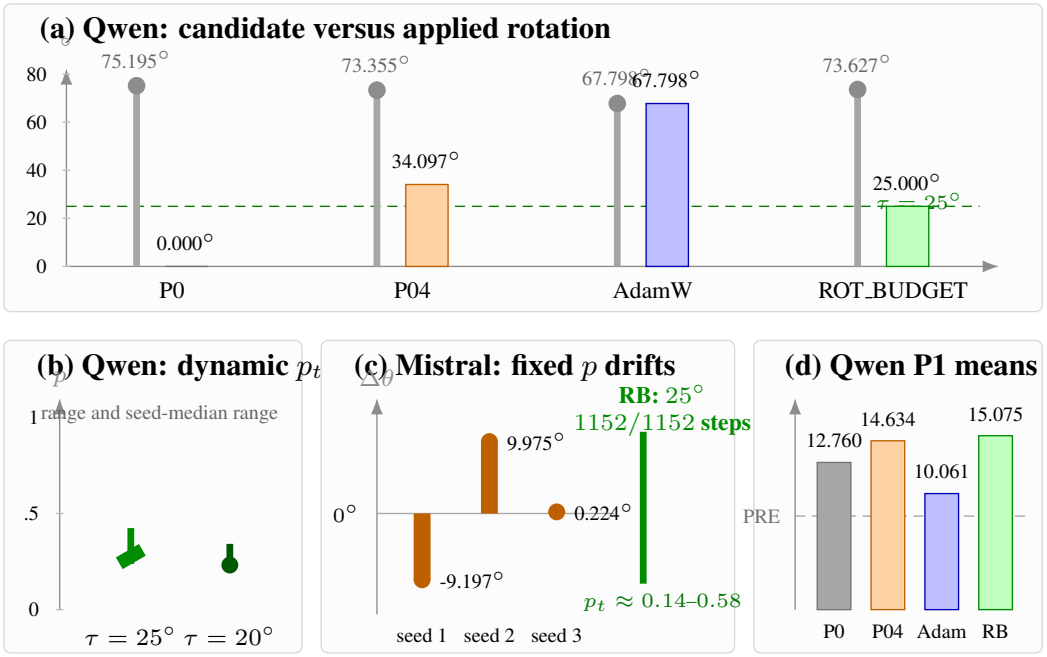

Figure 3: **Fixed preconditioning strength does not imply fixed geometry.** (a–b) In Qwen training, full Adam candidates remain near 68°–75°, while rotation budgeting dynamically selects p_t and applies a 25° update. (c) On Mistral, fixed $p = 0.4$ has seed-dependent angle drift, whereas the budgeted update is exactly 25° for all 1152 steps by moving p_t. (d) In the Qwen verifier setting, reduced-rotation arms have higher three-seed mean P1 endpoints than full AdamW; no Mistral performance claim is made.

meaningful advantage of dynamic budgeting over a well-chosen fixed exponent. They do show that the large adaptive turn persists, can be constrained directly, and that reduced-rotation arms finish above full AdamW on the measured held-out P1 margin in these runs.

5 The Geometry Transfers Beyond the Verifier Setting

The Qwen experiments show that rotation budgeting controls adaptive geometry in the verifier setting. The law itself is more general: it applies to any positive diagonal scaling. We therefore test it with a different architecture and a conventional training objective. The external setting uses Mistral-7B-Instruct-v0.2 with completion-only token cross-entropy on MBPP-style code tasks, not RObE or a verifier-derived training objective. LoRA has rank 8 and $\alpha = 16$ on the query, key, value, and output projections. We use 48 training tasks, 16 held-out tasks, and a shared PRE state after 300 standard AdamW warm-up steps; three seeds then run for 384 steps.

Large additional adaptive rotation. At PRE, the current raw gradient and momentum already differ by $83.185°$, so the Qwen component ordering does not transfer and we do not claim that second-moment scaling dominates the full decomposition. The positive cross-model result is narrower: scaling momentum by the second-moment state adds another $64.475°$ turn, while decoupled decay adds $0.078°$; raw descent and the final AdamW direction differ by $87.003°$.

The exact law remains unchanged. The weighted anisotropy is $\mathrm{CV}_w(a_1) = 2.0942$, predicting $\arctan \mathrm{CV}_w(a_1) = 64.475°$. Across $p \in \{0, .2, .4, .6, .8, 1\}$, the measured angles are $0°$, $10.7694°$, $23.2401°$, $37.5697°$, $52.1950°$, and $64.4750°$, with maximum prediction error 2.13×10^{-14} degrees. Every point uses the same step norm, 0.0140354. Thus p again provides a smooth geometry dial despite the different architecture, objective, and optimizer state (Figure 2c–d).

Fixed p is not fixed geometry. During training, fixed P04 begins near 23° but its realized angle changes as the moments evolve. The net drift is -9.197°, $+9.975^\circ$, and $+0.224^\circ$ across the three seeds; per-step standard deviation reaches 11.422° and 9.224° in the two drifting seeds. In contrast, ROT_BUDGET holds the applied rotation at 25.000000° for all $3 \times 384 = 1152$ steps, with zero standard deviation, by moving p_t over approximately 0.14–0.58 (Figure 3c). Fixed p fixes the knob. Rotation budgeting fixes what the knob actually does.

6 Related Work

Verifier-guided learning. Verifiable feedback supplies objective signals for mathematical reasoning and code generation (Shao et al., 2024; DeepSeek-AI, 2025; Yu et al., 2025; Setlur et al., 2025; Chen et al., 2026). This literature primarily studies how verifier rewards are constructed and optimized. We instead hold the learning setting fixed and study the geometry imposed between the verifier-derived signal and the realized update.

Adaptive and geometric preconditioning. Adam scales first moments coordinate-wise, while AdamW separates this operation from weight decay (Kingma & Ba, 2015; Loshchilov & Hutter, 2019). Sign-based views explain why adaptive normalization can become sign-like (Balles & Hennig, 2018); convergence and generalization analyses likewise distinguish adaptive scaling from scalar-step gradient descent (Wilson et al., 2017; Reddi et al., 2018). Matrix and geometry-aware optimizers make the choice of preconditioning geometry explicit (Gupta et al., 2018; Vyas et al., 2024; Zhang & Pilanci, 2024). We focus on a different object: the realized angle between a learning signal and its diagonally preconditioned update. We give that angle an exact scalar characterization and use it as an online control variable across two models and objectives.

Objectives and realized model changes. Optimization and implementation choices can separate an intended objective from the observed model response (Engstrom et al., 2020; Liu et al., 2025; Razin et al., 2025; Gopalan et al., 2026). Our analysis isolates one such transformation—directional rotation—while the mechanism controls show why it should not be treated as a complete account of task outcome.

7 Discussion

A gradient or momentum direction and the adaptive update received by the model are not interchangeable. In the Qwen verifier setting, second-moment scaling turns momentum by roughly 60° across tasks. In Mistral code SFT, it adds a 64.5° turn even though momentum is already far from the current raw gradient. The exact component decomposition is setting-dependent, but strong diagonal-preconditioning rotation appears in both.

Proposition 1 makes the angle interpretable. The turn is not an opaque consequence of optimizer complexity: it is the weighted coefficient of variation of the coordinate scales, measured where the current vector carries energy. This explains why scalar normalization cannot rotate an update and why large scale variation on negligible coordinates contributes little.

The cross-model training results also separate hyperparameter control from geometry control. A fixed exponent specifies how the second moment enters the denominator, but the resulting angle changes as $\hat{m}_t$ and $\hat{v}_t$ evolve. Mistral makes the distinction explicit: fixed $p = 0.4$ drifts differently across seeds, whereas rotation budgeting changes p_t to keep the realized direction at 25°. Rotation budgeting is therefore not merely automatic tuning of a preferred fixed exponent; it controls a different object.

Angle is not a universal utility score. One-step mechanism controls show that coordinate correspondence also matters: different second-moment states can produce outcomes that are not ordered by rotation magnitude. Likewise, the Qwen P1 endpoints support a controlled study in that setting, not a claim that smaller angles or rotation budgeting universally improve optimization. We make no Mistral downstream-performance claim.

Limitations. The experiments cover two 7B model families and LoRA parameterizations, eight valid Qwen verifier tasks, one code-SFT construction, and three training seeds per setting. Proposi-

tion 2 formally covers only $\epsilon = 0$, whereas implemented sweeps use $\epsilon > 0$. The supplied Mistral summaries establish geometric control but do not include a downstream benefit claim. Broader conclusions require additional models, parameterizations, objectives, and training regimes.

8 Conclusion

A learning objective supplies a direction, but adaptive optimization decides how that direction is realized in parameter space. Across verifier-based Qwen updates and conventional Mistral code SFT, second-moment scaling introduces a large turn. That turn is exactly governed by the weighted anisotropy of the coordinate scaling. Because the angle is predictable, it can be bounded directly even as optimizer state evolves. Optimizer geometry is not merely a by-product of the learning rule; it is part of the learning signal the model ultimately receives.

AI use statement

We used generative AI tools to improve the clarity and language of the manuscript and to assist in developing portions of the experimental infrastructure code. Large language models are also the subject of our experiments; their use for model generation, training, and evaluation is part of the research methodology rather than authorship assistance. The authors take responsibility for the final manuscript, code, analyses, and conclusions.

Ethics statement

This work studies the geometry of adaptive optimization in parameter-efficient post-training of pre-trained language models on verifier-scored and code-SFT tasks. The main ethical consideration is dual use: methods that make learning signals more measurable and controllable can support auditing and reproducibility, but may also facilitate adaptation of capable models for harmful applications. We limit our claims to the controlled models, tasks, and metrics studied, and we do not interpret verifier margins, construction objectives, or rotation control as evidence of model safety or general correctness. The experiments inherit the biases, failure modes, and licensing conditions of the base models, data, and verifiers. Any release of code, configurations, checkpoints, or evaluation artifacts should therefore respect applicable licenses, exclude sensitive information and credentials, and document intended use and known limitations. The authors take responsibility for responsible artifact release and for compliance with the relevant data and model terms.

Reproducibility statement

At submission, we will release the code and experimental artifacts needed to reproduce the paper's results as anonymous supplementary material. The release will include experiment and training configurations, verifier-objective specifications, model and checkpoint identifiers, LoRA and optimizer settings, task lists, random seeds, minimally processed measurements, and scripts that regenerate the reported tables and figures. We will also provide environment information and instructions linking each result to its generating configuration and record. For any model or data asset that cannot be redistributed because of licensing restrictions, we will provide its exact identifier and acquisition instructions.

References

Lukas Balles and Philipp Hennig. Dissecting Adam: The sign, magnitude and variance of stochastic gradients. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 404–413, 2018. URL https://proceedings.mlr.press/v80/balles18a.html.

Guanzheng Chen, Michael Qizhe Shieh, and Lidong Bing. LongRLVR: Long-context reinforcement learning requires verifiable context rewards. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2603.02146.

DeepSeek-AI. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning, 2025. URL https://arxiv.org/abs/2501.12948.

Logan Engstrom, Andrew Ilyas, Shibani Santurkar, Dimitris Tsipras, Firdaus Janoos, Larry Rudolph, and Aleksander Madry. Implementation matters in deep RL: A case study on PPO and TRPO. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=r1etN1rtPB.

Aditya Gopalan, Sayak Ray Chowdhury, and Debangshu Banerjee. Why DPO is a misspecified estimator and how to fix it. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/8f6c674a7edf57912f1a5830f4a66e3f-Abstract-Conference.html.

Vineet Gupta, Tomer Koren, and Yoram Singer. Shampoo: Preconditioned stochastic tensor optimization. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1842–1850, 2018. URL https://proceedings.mlr.press/v80/gupta18a.html.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. URL https://arxiv.org/abs/1412.6980.

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. In *Conference on Language Modeling*, 2025. URL https://openreview.net/forum?id=5PAF7PAY2Y.

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

Noam Razin, Sadhika Malladi, Adithya Bhaskar, Danqi Chen, Sanjeev Arora, and Boris Hanin. Unintentional unalignment: Likelihood displacement in direct preference optimization. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2410.08847.

Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar. On the convergence of Adam and beyond. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=ryQu7f-RZ.

Amrith Setlur, Chirag Nagpal, Adam Fisch, Xinyang Geng, Jacob Eisenstein, Rishabh Agarwal, Alekh Agarwal, Jonathan Berant, and Aviral Kumar. Rewarding progress: Scaling automated process verifiers for LLM reasoning. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/98711dea460bdefe0e651ca23ec98ba2-Abstract-Conference.html.

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, 2024. URL https://arxiv.org/abs/2402.03300.

Nikhil Vyas, Depen Morwani, Rosie Zhao, Mujin Kwun, Itai Shapira, David Brandfonbrener, Lucas Janson, and Sham Kakade. SOAP: Improving and stabilizing Shampoo using Adam, 2024. URL https://arxiv.org/abs/2409.11321.

Ashia C. Wilson, Rebecca Roelofs, Mitchell Stern, Nathan Srebro, and Benjamin Recht. The marginal value of adaptive gradient methods in machine learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://arxiv.org/abs/1705.08292.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Juncai Liu, Lingjun Liu, Xin Liu, Haibin Lin, Zhiqi Lin, Bole Ma, Guangming Sheng, Yuxuan Tong, Chi Zhang, Mofan Zhang, Ru Zhang, Wang Zhang, Hang Zhu, Jinhua Zhu, Jiaze Chen, Jiangjie Chen, Chengyi Wang, Hongli Yu, Yuxuan Song, Xiangpeng Wei, Hao Zhou, Jingjing Liu, Wei-Ying Ma, Ya-Qin Zhang, Lin Yan, Yonghui Wu, and Mingxuan Wang. DAPO: An open-source LLM reinforcement learning system at scale. In

Advances in Neural Information Processing Systems, volume 38, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/a4277440d50f1f15d2cb4c14f7e0c0d2-Abstract-Conference.html.

Fangzhao Zhang and Mert Pilanci. Riemannian preconditioned LoRA for fine-tuning foundation models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 59641–59669, 2024. URL https://proceedings.mlr.press/v235/zhang24ax.html.

A IDEALIZED MONOTONICITY PROOF

For Proposition 2, let $Y_i = \hat{v}_i^{-1/2} > 0$ on the support of $\hat{m}$, let $w_i = \hat{m}_i^2/\|\hat{m}\|_2^2$, and define $\phi(p) = \log \mathbb{E}_w[Y^p]$. From Proposition 1,

$$\cos\theta(p) = \frac{\mathbb{E}_w[Y^p]}{\sqrt{\mathbb{E}_w[Y^{2p}]}}, \qquad \log\cos\theta(p) = \phi(p) - \tfrac{1}{2}\phi(2p). \tag{12}$$

The log-moment function ϕ is convex, so ϕ' is non-decreasing. Hence

$$\frac{d}{dp}\log\cos\theta(p) = \phi'(p) - \phi'(2p) \le 0. \tag{13}$$

Thus cosine is non-increasing and the angle is non-decreasing on $[0, \pi/2)$. Strict convexity holds when Y is non-constant under w, yielding strict increase for $p > 0$. This result assumes $\epsilon = 0$ and is not presented as a theorem for the implemented $\epsilon > 0$ denominator.

B CROSS-TASK COMPONENT AND IDENTITY AUDITS

Table 1: Aggregate component localization across eight valid non-degenerate verifier tasks.

Component	Mean angle	Maximum / range	Rotation share
Raw gradient → momentum	$0.00742°$	max $0.01978°$	$\approx 1.77 \times 10^{-14}$
Second-moment scaling	$59.9549°$	$53.1128°$–$62.2057°$	0.9996714
Preconditioned → decay	$0.01170°$	max $0.03426°$	0.0001601

Second-moment scaling is largest in $8/8$ tasks. Its median is $60.7709°$ and its bootstrap 95% confidence interval is $[57.8075°, 61.4086°]$.

Table 2: Numerical validation of $\tan\theta = \mathrm{CV}_w(a)$.

Measurement family	Points	Maximum absolute angle error
Frozen $(\mathrm{task}, p, \epsilon)$ grid	441	1.2074×10^{-6} degrees
Mechanism controls	63	1.4788×10^{-6} degrees
Real training steps	4608	1.7080×10^{-6} degrees
Mistral PRE p sweep	6	2.13×10^{-14} degrees
Mistral training steps	1152	approximately machine precision

C FROZEN PRECONDITIONING CONTROLS

Table 3: Canonical dense p sweep. Every direction uses Equation 2.

p	Rotation	Construction gain
0.0	$0.0198°$	$+3.2110$
0.2	$11.8267°$	$+2.9611$
0.4	$23.5299°$	$+2.3044$
0.6	$35.1328°$	$+2.3309$
0.8	$46.6955°$	$+1.7028$
1.0	$58.3153°$	$+1.1236$
1.2 (extrapolation)	$69.8340°$	$+0.4849$

Across eight tasks, the $p = 0$ to $p = 1$ increase has the same sign in $8/8$ cases, with mean $+59.9475°$ and bootstrap 95% CI $[57.798°, 61.388°]$. The 63-point norm audit has maximum absolute discrepancy 1.49×10^{-8} and maximum relative discrepancy 1.10×10^{-7}. At $p = 1$, the tested ϵ values span $53.7617°$–$58.3351°$.

Table 4: Mean full-candidate and applied rotations over 384 training steps.

Arm	Full q_1 candidate	Applied update
P0	75.195°	0.000°
P04	73.355°	34.097°
AdamW	67.798°	67.798°
ROT_BUDGET ($\tau = 25^\circ$)	73.627°	25.000°

D Training Geometry and Endpoints

For $\tau = 25^\circ$, p_t ranges approximately from 0.238 to 0.424, with per-seed medians between 0.241 and 0.311; no step uses $p_t = 0$ or $p_t = 1$. For the $\tau = 20^\circ$ sensitivity run, p_t ranges approximately from 0.211 to 0.342, with median approximately 0.232.

Table 5: Held-out P1 endpoints after 384 steps (three-seed means).

Setting	Final margin	Gain from PRE (8.10438)
P0	12.760	+4.656
P04	14.634	+6.529
AdamW	10.061	+1.957
ROT_BUDGET	15.075	+6.971

Table 6: Paired within-seed endpoint differences relative to AdamW.

Arm	Seed 1	Seed 2	Seed 3	Mean
P0 − AdamW	+3.248	+2.177	+2.672	+2.699
P04 − AdamW	+1.167	+11.277	+1.273	+4.572
ROT_BUDGET − AdamW	+1.415	+12.237	+1.390	+5.014

The manuscript snapshot does not contain checkpoint-level evaluation values. Those raw records are required before replacing the endpoint summary in Figure 3 with training curves.

E Mistral External-Validation Details

The external setting uses Mistral-7B-Instruct-v0.2 and completion-only token cross-entropy on MBPP-style code tasks. It does not use RObE or a verifier-derived training objective. LoRA uses rank 8, $\alpha = 16$, and query/key/value/output projections. The split contains 48 training tasks and 16 held-out tasks. All three seeds share a PRE state produced by 300 standard AdamW warm-up steps, followed by 384 measured training steps.

At $p = 1$, $\mathrm{CV}_w(a_1) = 2.0942$ and $\arctan \mathrm{CV}_w(a_1) = 64.475^\circ$. The matched step norm is 0.0140354 at every PRE sweep point.

Across the $3 \times 384 = 1152$ budgeted steps, the applied-angle standard deviation is zero and p_t ranges approximately from 0.14 to 0.58. These measurements validate geometry control; no Mistral downstream-performance claim is made.

F One-Step Mechanism Controls

The cross-task scale rotates more than the true current scale, yet the outcomes are not ordered by angle. Rotation magnitude and coordinate correspondence are distinct properties.

G Random and Structured Direction Controls

The Holm-adjusted family-wise value across the three planned metrics is 0.011673.

Table 7: Mistral PRE component angles. The Qwen-specific dominance claim is not applied here.

Comparison	Angle
Raw gradient → momentum	83.185°
Momentum → Adam	64.475°
Adam → AdamW	0.078°
Raw gradient → AdamW	87.003°

Table 8: Mistral PRE p sweep.

p	Measured angle	Maximum identity error over sweep
0.0	0.0000°	
0.2	10.7694°	
0.4	23.2401°	
0.6	37.5697°	2.13×10^{-14} degrees
0.8	52.1950°	
1.0	64.4750°	

H Cross-Host Aggregate Reproduction

The canonical preconditioning rotation is 58.3337° on the original host and 58.3153° on a second host, a difference of approximately 0.018°. The norm radius 0.13537724316120148 reproduces exactly. Detailed hardware manifests and high-energy versus numerical-floor coordinate diagnostics remain artifact requirements.

I Reproducibility Records Required for Submission

The authors must supply the records needed to reconstruct every reported quantity:

1. the exact verifier objectives, `pva244` construction, RObE algorithm, task identifiers, and inclusion rule;
2. model and PRE-checkpoint identifiers, LoRA configuration, trainable parameters, precision, source revision, and random states;
3. Adam/AdamW state, hyperparameters, clipping, decay convention, and the numerical search used for p_t^*;
4. all frozen-grid, identity-validation, mechanism-control, and random-null measurements with generation recipes and seeds;
5. complete configurations for all 12 primary Qwen runs and the $\tau = 20^\circ$ sensitivity run, together with per-step p_t, angle, and checkpoint-level P1 records;
6. the Mistral task split, warm-up and training configurations, all three fixed-P04 and ROT_BUDGET step records, and the matched-norm audit;
7. exact metric definitions, reduction units, bootstrap procedures, and artifact paths mapping each table and figure to its generating record; and
8. author-approved AI-use, ethics, and reproducibility statements.

Table 9: Mistral training-time geometry.

Quantity	Seed 1	Seed 2	Seed 3
Fixed-P04 net angle drift	-9.197°	$+9.975^\circ$	$+0.224^\circ$
Reported per-step SD	11.422°	9.224°	not supplied
ROT_BUDGET applied angle	25.000000° at every step		

Table 10: Selected aggregate mechanism controls across nine tasks. All directions are norm matched.

Control	Rotation	CV_w	Construction	Held-out P1	KL / logit ℓ_2
Momentum	0.000°	0	0.50579	8.17130	0.00401 / 263.76
Isotropic v	0.000°	0	0.50579	8.17130	0.00401 / 263.76
Permuted v	35.327°	0.71138	0.48800	8.15592	0.00361 / 244.34
Cross-task v	80.638°	19.04645	0.39390	8.09255	0.00287 / 190.23
True current v	59.773°	1.72874	0.27299	8.10727	0.00485 / 191.14

Table 11: RObE against 256 norm-matched isotropic random directions.

Metric	RObE	Random mean	Rank	Empirical p
Token KL	0.01389	0.003032	$1/257$	0.003891
Logit ℓ_2	703.71	189.744	$1/257$	0.003891
Construction gain	$+3.369$	$+0.350$	$1/257$	0.003891

Table 12: Distance from RObE in standardized token-KL, logit-ℓ_2, and construction-gain space.

Control	Distance	Control	Distance
Raw gradient	0.7805	Momentum	0.7805
q_0	0.7805	$q_{0.2}$	0.8483
$q_{0.4}$	1.9962	$q_{0.6}$	2.6865
PCA	2.8073	$q_{0.8}$	3.2213
$\mathrm{sign}(g)$	3.7705	q_1	3.8538
Coordinate-permuted RObE	4.0363	RMSProp	4.0794
Tensor-shuffled RObE	4.5241	$\mathrm{sign}(\hat{m})$	4.5585
Sign-randomized RObE	4.7671		