EGGROLL, UNROLLED: UNDERSTANDING AND IMPROVING LOW-RANK EVOLUTION STRATEGIES AT SCALE

Anonymous authors
Paper under double-blind review

ABSTRACT

EGGROLL (Sarkar et al., 2026) makes evolution strategies (ES) practical for LLMs by replacing dense Gaussian weight perturbations with low-rank Gaussian products, often of rank one. This choice is computationally attractive but geometrically severe: Each rank-one perturbation lies in a zero-volume subset of the ambient matrix space, despite having identity covariance. We characterize the EGGROLL update mean field at finite rank and nonzero perturbation radius as a resolvent applied to the gradient of the perturbation-smoothed objective. This transformation can make the mean field nonconservative and reverse the local stability of an optimum. EGGROLL nevertheless recovers the gradient exactly on quadratic objectives at every rank and radius. In finite populations, the additional sampling variance of rank-one perturbations relative to dense Gaussian ES decays inversely with matrix width under a local affine model, and is only 0.098% at width 4096. Finally, we introduce *LOO-ROLL*, a leave-one-out estimator that replaces EGGROLL's two antithetic evaluations per direction by one. At equal evaluation cost, LOO-ROLL halves estimator MSE in transformer blocks. Across fourteen post-training settings up to 14B parameters, matched-time comparisons with EGGROLL yield eleven improvements in individual paired tests and no significant loss. At 1.7B, 8B, and 14B parameters, matched-time gains are 2.9, 14.1, and 7.9 percentage points on GSM8K and 12.2, 8.4, and 8.1 points on MATH-500.

1 INTRODUCTION

Evolution strategies (ES) (Schwefel, 1977; Rechenberg, 1978; Beyer & Schwefel, 2002; Salimans et al., 2017) optimize a scalar fitness using perturbed function evaluations. Their forward-only computation has motivated memory-efficient language-model fine-tuning (Malladi et al., 2023; Gautam et al., 2024; Zhao et al., 2025), while their parallel populations support scalable post-training (Sarkar et al., 2026; Hoy et al., 2026; Xu et al., 2026). A remaining difficulty is moving a large population of independently perturbed parameter settings through the network efficiently. Modern accelerators obtain high throughput when arithmetic reuses data already brought into fast memory (Williams et al., 2009; Dao et al., 2022). Giving each member a dense perturbation requires generating or reading another mn numbers per weight matrix, making memory traffic grow with the population.

EGGROLL (Sarkar et al., 2026) perturbs each weight matrix with a rank-r Gaussian product,

$$E_r = \frac{1}{\sqrt{r}} AB^\top = \frac{1}{\sqrt{r}} \sum_{s=1}^{r} a_s b_s^\top, \qquad a_s \sim \mathcal{N}(0, I_m), \quad b_s \sim \mathcal{N}(0, I_n), \tag{1}$$

which permits the evaluation of very large populations at near-inference throughput. The normalization by $1/\sqrt{r}$ gives $\mathbb{E}[\mathrm{vec}(E_r)\,\mathrm{vec}(E_r)^\top] = I_{mn}$ like in dense Gaussian ES, yet each realization has rank at most r. When $r < \min(m, n)$, the possible realizations occupy zero volume in the ambient matrix space (Absil et al., 2008). For an activation x, the perturbed computation is

$$(W + \sigma E_r)x = Wx + \frac{\sigma}{\sqrt{r}} A(B^\top x). \tag{2}$$

The population shares the base weights W, while each member receives a correction through two thin multiplications with A and $B^\top$. These require $O(r(m+n))$ additional storage and arithmetic rather

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.

than the $O(mn)$ of dense Gaussian ES. Low-rank adapters are already widely used in LLM training (Hu et al., 2022; Dettmers et al., 2023; Zhang et al., 2023; Liu et al., 2024). EGGROLL adopts this representation to estimate an update to network weights without backpropagation.

EGGROLL performs well in LLM post-training even in the seemingly prohibitive regime of $r = 1$. The original analysis provides asymptotic guarantees, but does not identify the field at the finite ranks and radii used in practice. We ask: *(i) what mean field does EGGROLL follow, (ii) how accurately does a finite population estimate it, and (iii) what improvements does our analysis suggest?*

Our answers reveal the mean dynamics of EGGROLL and the sampling accuracy of a finite population. The mean field is a resolvent-filtered gradient, which can be nonconservative and can make a local optimum repelling. The discrepancy vanishes on quadratics, and the first local finite-rank correction is proportional to σ^2/r. The sampling cost is milder: rank-r perturbations add $2(m + n + 1)/r$ to the dense-Gaussian variance factor $mn + 1$. Rank one therefore retains nearly all of the sampling efficiency of dense Gaussian ES for wide matrices. We then introduce *LOO-ROLL*. The released EGGROLL implementation uses two antithetic fitness evaluations per direction. LOO-ROLL uses the other population members as leave-one-out baselines, requiring one evaluation per direction while preserving EGGROLL's expected update. The saving can then be used on additional directions or optimization steps. Across fourteen settings, two model families, and models up to 14B, eleven matched-time gains are individually significant and nine survive Holm-Bonferroni correction (Holm, 1979). The analysis also suggests rank extrapolation to mitigate the bias introduced by finite rank and control variates to mitigate the high variance caused by ES. However, in tested LLM runs, the additional cost of these modifications do not justify their usage, hence we treat them in Appendix E.1.

We summarize our contributions as follows.

- We derive the exact finite-rank EGGROLL population field as a resolvent-filtered gradient and show that it can be nonconservative and locally destabilize an optimum.
- We prove exactness on quadratics, identify the leading finite-rank correction, and show that the surplus finite-population variance relative to dense Gaussian ES decays with matrix width.
- We introduce LOO-ROLL, replacing antithetic pairs with one-evaluation leave-one-out directions while preserving the mean EGGROLL field. Across fourteen LLM post-training settings up to 14B parameters, eleven matched-time gains are individually significant, nine survive Holm–Bonferroni correction, and none significantly favors EGGROLL.

Related work. Random-direction finite differences and simultaneous perturbation are classical zeroth-order methods (Spall, 1992; Nesterov & Spokoiny, 2017; Duchi et al., 2015). MeZO fine-tunes LLMs with inference-level memory, and subsequent work adds variance reduction (Malladi et al., 2023; Gautam et al., 2024). Baselines reduce variance in score-function estimators (Williams, 1992; Wierstra et al., 2014), with REINFORCE leave-one-out applied to LLM response groups (Ahmadian et al., 2024) and curvature-aware zeroth-order tuning (Seung et al., 2026). LOO-ROLL applies this principle to EGGROLL's Gaussian-product population. Scalar product-normal Stein identities are established (Gaunt, 2018b;a). Our matrix operator retains the shared row and column factors and connects their geometry to optimization dynamics. Extended comparisons appear in Appendix A.

2 The finite-rank mean field

We start our inquiry by asking what finite-rank EGGROLL updates follow in expectation. We consider one parameter matrix $W \in \mathbb{R}^{m\times n}$. Appendix E combines independent perturbations across the matrix blocks of a network. Let $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ be the fitness to maximize, and use the Frobenius inner product $\langle A, B\rangle_F = \mathrm{tr}(A^\top B)$. Sampling perturbed weights produces the smoothed objective

$$F_{r,\sigma}(W) = \mathbb{E}[f(W + \sigma E_r)], \tag{3}$$

where $\sigma > 0$ is the perturbation radius. EGGROLL weights each perturbation by its observed fitness. Averaging this update gives the mean field

$$g^f_{r,\sigma}(W) = \frac{1}{\sigma}\mathbb{E}[E_r f(W + \sigma E_r)]. \tag{4}$$

This is the infinite-population update that we will later approximate with a finite population. We write E_∞ for a dense standard Gaussian matrix, corresponding to the $r \to \infty$ limit of the Gaussian-product

law. For a perturbation density p, the *score* is $\nabla \log p$. A dense standard Gaussian matrix E_∞ has score $-E_\infty$. Under the usual regularity conditions, Stein's Gaussian identity (Stein, 1981) gives

$$\frac{1}{\sigma}\mathbb{E}[E_\infty f(W+\sigma E_\infty)] = \mathbb{E}[\nabla f(W+\sigma E_\infty)] = \nabla F_{\infty,\sigma}(W). \tag{5}$$

The same identity does not automatically hold for the Gaussian-product law, since at finite rank, $-E_r$ is not its exact score.

The reported experiments, fitness shaping, and released LLM implementation of Sarkar et al. (2026) use paired antithetic perturbations $+E_r$ and $-E_r$:

$$\widehat{g}_{r,\sigma,N}(W) = \frac{1}{N}\sum_{s=1}^{N} E_r^{(s)} \frac{f(W+\sigma E_r^{(s)}) - f(W-\sigma E_r^{(s)})}{2\sigma}. \tag{6}$$

Here N is the number of independent directions, each evaluated with both signs. The symmetry of E_r makes $\widehat{g}_{r,\sigma,N}$ an unbiased estimator of the mean field. Evaluating both signs on the same prompt batch also isolates the effect of the perturbation. Any fitness component shared by the pair, such as the difficulty of the prompts in that batch, cancels when the two evaluations are subtracted. For a smooth objective, a Taylor expansion around W makes this explicit:

$$\frac{f(W+\sigma E_r) - f(W-\sigma E_r)}{2\sigma} = \langle \nabla f(W), E_r\rangle_F + O(\sigma^2 \|E_r\|_F^3), \tag{7}$$

where the constant and other even-order terms cancel. EGGROLL normally also centers and standardizes fitness values across the population. Our main analysis studies the raw, unstandardized field in Equation (4). Appendix G treats the standardization and quantifies its difference from the raw field.

The definitions above lead to two deterministic questions. First, how does the EGGROLL mean field $g^f_{r,\sigma}$ differ from the gradient $\nabla F_{r,\sigma}$ of the objective smoothed by the same perturbations? Second, when they differ, is $g^f_{r,\sigma}$ still *conservative*, meaning that it is the gradient of some scalar objective? Sampling with E_r smooths f into $F_{r,\sigma}$, while weighting those samples by the dense-Gaussian score produces the mean field. We first characterize the discrepancy between these two fields and then determine when the EGGROLL field remains conservative.

Both $F_{r,\sigma}$ and $g^f_{r,\sigma}$ can be written as convolutions. The smoothed objective $F_{r,\sigma}$ convolves f with the Gaussian-product law, while each component of the mean field $g^f_{r,\sigma}$ convolves f with the same law weighted by the corresponding entry of E_r/σ. Hence, a Fourier frequency analysis lets us compare the smoothed gradient and the EGGROLL mean field. Letting $\Phi_r(T) = \mathbb{E}[e^{i\langle T, E_r\rangle_F}]$ denote the characteristic function of the perturbation law, we have the following result.

Theorem 2.1 (Exact Gaussian-product law)**.** *For $T \in \mathbb{R}^{m\times n}$,*

$$\Phi_r(T) = \det\left(I_m + \frac{TT^\top}{r}\right)^{-r/2}, \qquad \nabla_T \Phi_r(T) = -\Phi_r(T)\left(I_m + \frac{TT^\top}{r}\right)^{-1} T. \tag{8}$$

The proof is given in Appendix B. For the Fourier mode $f_T(W) = e^{i\langle T, W\rangle_F}$, this gives

$$g^{f_T}_{r,\sigma}(W) = i\Phi_r(\sigma T)\left(I_m + \frac{\sigma^2 TT^\top}{r}\right)^{-1} T e^{i\langle T, W\rangle_F}. \tag{9}$$

For the same mode, direct substitution into Equation (3) gives

$$F_{r,\sigma}(W) = \Phi_r(\sigma T) e^{i\langle T, W\rangle_F}, \qquad \nabla F_{r,\sigma}(W) = i\Phi_r(\sigma T) T e^{i\langle T, W\rangle_F}. \tag{10}$$

Comparing this expression with Equation (9), smoothing contributes the common scalar factor $\Phi_r(\sigma T)$, while the score mismatch tilts the gradient direction T to

$$J(T) = \left(I_m + \frac{\sigma^2 TT^\top}{r}\right)^{-1} T. \tag{11}$$

Figure 1 visualizes this transformation on the diagonal subspace $T = \mathrm{diag}(t_1, t_2)$.

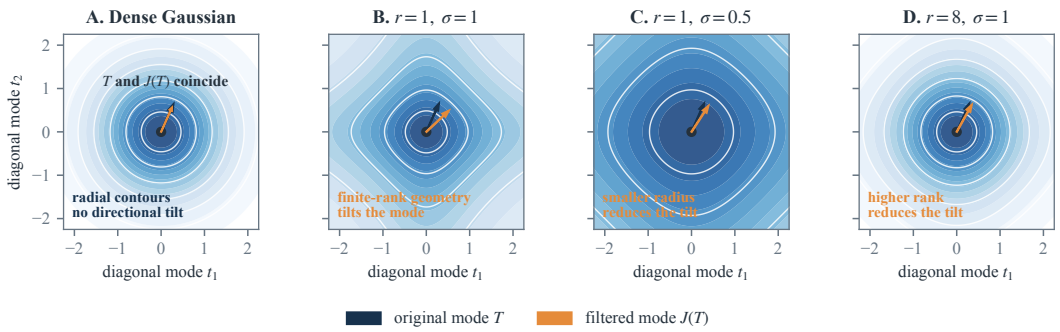


Figure 1: Geometry on the diagonal subspace $T = \mathrm{diag}(t_1, t_2)$. Darker contours denote weaker attenuation. Blue arrow gives the gradient-mode direction T, and orange arrow gives the population-mode direction $J(T)$. **A:** A dense Gaussian has radial contours and does not tilt the mode. **B:** The rank-one product law is nonradial and tilts $J(T)$ away from T. **C:** Decreasing σ reduces attenuation and tilt. **D:** Increasing rank moves the geometry toward the dense-Gaussian limit.

To translate the frequency-wise multiplier back into weight space, define the differential operator

$$(\mathcal{L}V)_{ij} = -\sum_{k=1}^{m}\sum_{\ell=1}^{n} \partial_{i\ell}\partial_{k\ell}V_{kj}, \tag{12}$$

where ∂_{ab} differentiates with respect to W_{ab}. On a mode $Ce^{i\langle T,W\rangle_F}$, $\mathcal{L}$ multiplies C by $TT^\top$. Combining these frequency-wise transformations recovers the relationship between the complete mean field and smoothed gradient.

Theorem 2.2 (The finite-rank resolvent)**.** *If f is locally integrable and of at most polynomial growth,*

$$g^f_{r,\sigma} = \left(I + \frac{\sigma^2}{r}\mathcal{L}\right)^{-1} \nabla F_{r,\sigma}. \tag{13}$$

The equality holds in the sense of tempered distributions. It holds pointwise if $f(W) = \int e^{i\langle T,W\rangle_F}\mu_f(dT)$ with $\int(1+\|T\|_F^3)|\mu_f|(dT) < \infty$.

The inverse $(I + \sigma^2\mathcal{L}/r)^{-1}$ is the *resolvent* of the positive operator $\sigma^2\mathcal{L}/r$ (Bauschke & Combettes, 2020). Thus, the mean field is obtained by spectrally damping the smoothed gradient rather than merely rescaling it. The distributional statement covers polynomially growing objectives such as quadratics, while the stronger Fourier condition gives an ordinary pointwise equality. Appendix B gives the proof and regularity details. If T has singular values s_j, the resolvent scales its corresponding directions by $(1 + \sigma^2 s_j^2/r)^{-1}$. Larger singular values are thus attenuated more strongly, and the unequal factors can rotate the mode direction despite identity covariance.

The resolvent identity also quantifies how far the EGGROLL mean field can depart from the gradient of the correspondingly smoothed objective.

Corollary 2.3 (Exact score-mismatch control)**.** *Let $J_{r,\sigma} = (I + \sigma^2\mathcal{L}/r)^{-1}$. If $\nabla F_{r,\sigma}$ belongs to the domain of $\mathcal{L}$, then*

$$g^f_{r,\sigma} - \nabla F_{r,\sigma} = -\frac{\sigma^2}{r} J_{r,\sigma}\mathcal{L}\nabla F_{r,\sigma}, \qquad \|g^f_{r,\sigma} - \nabla F_{r,\sigma}\|_{L^2} \le \frac{\sigma^2}{r}\|\mathcal{L}\nabla F_{r,\sigma}\|_{L^2}. \tag{14}$$

Thus the departure from the smoothed gradient is controlled by its variation in the matrix-coupled directions measured by $\mathcal{L}$, and vanishes linearly with σ^2/r. The proof appears in Appendix B.

3 When does finite rank change the dynamics?

The resolvent characterizes the EGGROLL mean field, but does not guarantee that it optimizes any scalar objective. We therefore ask which perturbation laws always produce a conservative field,

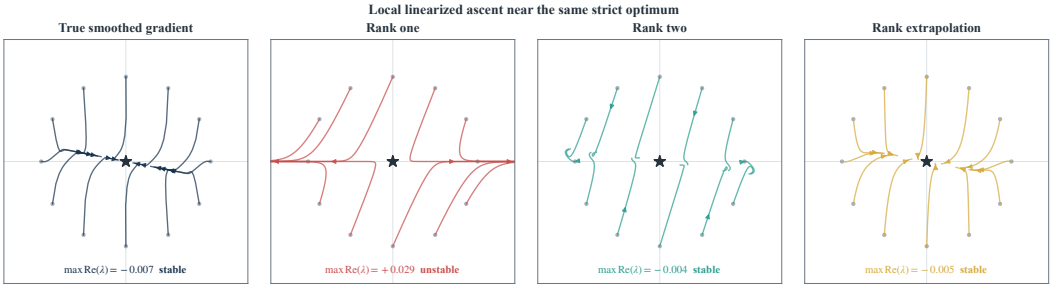


Figure 2: Local linearized dynamics for the cosine objective in Equation (17). Curves solve $\dot{\delta} = A\delta$, where A is each field's Jacobian at the origin, and are shown in the same plane. The origin is a strict local maximum of both f and $F_{1,1}$. Rank-one EGGROLL repels along one direction, while rank two is locally attracting. The rank-extrapolation comparison is developed in Appendix E.1.

equivalently, one with a symmetric Jacobian (Spivak, 1965). Real trigonometric polynomials suffice as a test class because their finite Fourier sums can isolate arbitrary frequencies. Failure on this class rules out a universal guarantee for any broader class containing it.

Theorem 3.1 (Universal conservativity). *Let $d \geq 2$ and let $Z \in \mathbb{R}^d$ be centered with $\mathbb{E}[\|Z\|] < \infty$. The field*

$$x \mapsto \frac{1}{\sigma}\mathbb{E}[Z f(x + \sigma Z)] \tag{15}$$

is conservative for every real trigonometric polynomial f and every $\sigma > 0$ if and only if the law of Z is spherically symmetric.

Spherical symmetry makes the characteristic function $\Phi(t)$ radial, so $\nabla\Phi(t)$ is parallel to the frequency t. Every Fourier mode of the mean field then remains aligned with the corresponding gradient mode. Conversely, if $\nabla\Phi(t)$ is not parallel to t at some frequency, a real trigonometric objective can combine such modes to produce a nonsymmetric Jacobian. The Gaussian-product characteristic function depends on the individual singular values of its matrix argument rather than only on its Frobenius norm. Its law is therefore not spherically symmetric at finite rank, and the EGGROLL mean field need not be conservative. The proof and an explicit cosine example appear in Appendix B.4.

Nonconservativity alone does not necessarily show that the altered field harms optimization. The following construction shows that finite rank can change even the local stability of a strict optimum.

Proposition 3.2 (An optimum can become repelling). *Set $r = \sigma = 1$, $\varepsilon = 0.01$, and*

$$T_1 = \begin{pmatrix} -2 & -2 \\ -2 & -1 \end{pmatrix}, \qquad T_2 = \begin{pmatrix} -2 & 2 \\ 2 & -1 \end{pmatrix}. \tag{16}$$

Define

$$f(W) = \cos\langle T_1, W\rangle_F + \frac{1}{4}\cos\langle T_2, W\rangle_F + \varepsilon \sum_{i,j} \cos W_{ij}. \tag{17}$$

Then zero is a strict local maximum of both f and $F_{1,1}$, but the Jacobian of $g^f_{1,1}$ at zero has a positive eigenvalue $\lambda_+ \approx 0.02925$.

The construction and its trajectories are detailed in Appendix B.4 and Figure 2.

The counterexample establishes that finite rank can change the dynamics, but not which properties of the objective permit this change. Quadratic objectives provide a dividing line. For any quadratic f and any realization of E_r, the antithetic difference equals the directional derivative:

$$\frac{f(W + \sigma E_r) - f(W - \sigma E_r)}{2\sigma} = \langle \nabla f(W), E_r\rangle_F. \tag{18}$$

Multiplying by E_r and taking expectations recovers $\nabla f(W)$ by identity covariance.

Proposition 3.3 (Exactness on quadratic objectives). *For $f(W) = \frac{1}{2}\langle W, \mathcal{H}W\rangle_F + \langle C, W\rangle_F + c$ with self-adjoint $\mathcal{H}$, $g^f_{r,\sigma}(W) = \nabla f(W)$ for every $r \geq 1$ and $\sigma > 0$.*

Quadratic exactness clarifies the source of the instability in Proposition 3.2. The objective and its quadratic Taylor approximation have the same value, gradient, and Hessian at the origin, but EGGROLL evaluates the objective at finite perturbations around that point. At $\sigma = 1$, these evaluations reach regions where the cosine terms are poorly quadratically approximated. The stability reversal therefore comes from higher-order variation over the perturbation scale. For a general smooth objective, the following result identifies the first such departure from quadratic behavior.

Proposition 3.4 (Small-radius expansion). *Fix m, n, r. If $f \in C^5$ and $M_5 = \sup_W \|D^5 f(W)\|_{\mathrm{op}} < \infty$, then, uniformly in W as $\sigma \to 0$,*

$$g^f_{r,\sigma} = \nabla f + \frac{\sigma^2}{2}\nabla\Delta f + \frac{\sigma^2}{r}\mathcal{D}f + R_{r,\sigma}, \quad \sup_W \|R_{r,\sigma}(W)\|_F \le \frac{\sigma^4 M_5}{120}\mathbb{E}[\|E_r\|_F^6], \tag{19}$$

where $\Delta f = \sum_{k,\ell} \partial^2_{k\ell} f$ and $(\mathcal{D}f)_{ij} = \sum_{k,\ell} \partial_{i\ell}\partial_{k\ell}\partial_{kj} f$.

The proof is given in Appendices B.5 and E.3. The gradient-Laplacian term is the ordinary cost of dense Gaussian smoothing. The additional third-derivative term is specific to finite rank and need not be a gradient field. Both vanish on quadratics. The expansion concerns the complete displacement σE_r, whose root-mean-square Frobenius norm is $\sigma\sqrt{mn}$, so its accuracy depends on dimension and derivative bounds as well as radius. Proofs and nonasymptotic bounds under standard smoothness appear in Appendix B.5. Conventional convergence consequences are collected in Appendix E.2.

4 The Sampling Cost of Finite Rank

The preceding sections determine the mean update averaged over the perturbation law. An actual EGGROLL step replaces this expectation with an average over finitely many directions. We now ask how accurately that population estimates its mean. To isolate this sampling variability, we condition on any randomness in fitness evaluation and use the local affine model $f(W+\sigma E) = f(W)+\sigma\langle G, E\rangle_F$, where $G = \nabla f(W)$. One antithetic direction then gives $X = E_r\langle E_r, G\rangle_F$, with $\mathbb{E}[X] = G$. Because X is unbiased, its total variance, the trace of its covariance, equals $\mathbb{E}[\|X - G\|_F^2]$.

Theorem 4.1 (Finite-population variance). *For independent $X_1, \dots, X_N$ and $\widehat{G}_N = N^{-1}\sum_s X_s$,*

$$\mathbb{E}[\|\widehat{G}_N - G\|_F^2] = \frac{\kappa_r}{N}\|G\|_F^2, \qquad \kappa_r := mn + 1 + \frac{2(m+n+1)}{r}. \tag{20}$$

The full covariance and proof appear in Appendix B.6. Dense Gaussian directions share the factor $mn+1$, while finite rank adds $2(m+n+1)/r$. For a square matrix of width w, the relative variance increase over dense Gaussian ES is therefore $2(2w+1)/[r(w^2+1)] = O(1/(rw))$. To give an LLM-scale example, at $w = 4096$, as in Qwen3-8B, the rank-one increase is only 0.098%. Rank one can therefore estimate its mean field almost as efficiently as dense Gaussian ES estimates its own.

A complementary question is how the mean field itself changes with rank. For $\|u\|_{\infty,p} := \sup_W \|u(W)\|_F/(1+\|W\|_F)^p$, the kernel expansion in Appendix B.7 gives a uniform statement.

Theorem 4.2 (Fixed-radius rank expansion). *If $|f(W)| \le C_f(1+\|W\|_F^p)$, then for every fixed $\sigma > 0$ there exist $R_{\sigma,p} < \infty$ and r_0 such that, for $r \ge r_0$,*

$$\left\| g^f_{r,\sigma} - g^f_{\infty,\sigma} - \frac{1}{r}\mathcal{B}_\sigma f \right\|_{\infty,p} \le \frac{R_{\sigma,p}}{r^2}. \tag{21}$$

If $\mathcal{B}_\sigma f$ is nonzero, the $1/r$ rate is tight in this norm.

The coefficient $\mathcal{B}_\sigma f$ identifies the leading finite-rank error, motivating Richardson extrapolation using ranks r and αr. However, each estimate retains the dense-Gaussian variance coefficient $mn+1$. Extrapolation therefore improves the bias order while requiring more evaluations and amplifying sampling noise. Appendices B.7 and E.1 gives the proof and develops this tradeoff.

5 EGGROLL leads to LOO-ROLL

The preceding analysis identifies finite-rank bias and the large directional variance already present in dense Gaussian ES. The antithetic implementation adds a third cost: two evaluations per direction.

LOO-ROLL reduces this cost by using the rest of the population to construct a leave-one-out baseline. At iteration t, sample $N \geq 2$ independent directions and write $F_{t,s} = f(W_t + \sigma E_{t,s})$. Define

$$\widehat{g}_t^{\mathrm{LOO}} = \frac{1}{N\sigma} \sum_{s=1}^{N} E_{t,s}(F_{t,s} - \overline{F}_{t,-s}), \qquad \overline{F}_{t,-s} = \frac{1}{N-1} \sum_{j \neq s} F_{t,j}. \tag{22}$$

Since $F_{t,s} - \overline{F}_{t,-s} = N(F_{t,s} - \overline{F}_t)/(N-1)$, the baseline requires only the population mean and no additional evaluations.

Proposition 5.1 (Unbiasedness of LOO-ROLL)**.** *Let $\mathcal{F}_t$ denote the sigma-field generated by the optimization history before iteration t. Conditional on $\mathcal{F}_t$, suppose that $N \geq 2$ and $E_{t,1}, \ldots, E_{t,N}$ are independent, centered, and have distribution $\mathrm{Law}(E_r)$. Assume that $F_{t,s}$ and $\|E_{t,s}\|_F F_{t,s}$ are conditionally integrable. Then,*

$$\mathbb{E}\left[\widehat{g}_t^{\mathrm{LOO}} \mid \mathcal{F}_t\right] = g_{r,\sigma}^f(W_t). \tag{23}$$

This unbiasedness statement holds before population score standardization. The practical implementation retains EGGROLL's standardization, whose effect on the expected update is treated in Appendix G. To compare variances, we use the same affine model as in Section 4. Write $\widehat{g}^{\mathrm{LOO}}(W; N)$ and $\widehat{g}^{\mathrm{anti}}(W; N)$ for the estimators formed from N directions, and let $d = mn$.

Proposition 5.2 (Variance of LOO-ROLL in the affine model)**.** *Suppose $f(W + \sigma E) = f(W) + \sigma\langle G, E\rangle_F$ and write $d = mn$, with κ_r as defined in Equation (20). For $N \geq 2$ independent directions, the LOO-ROLL estimator satisfies*

$$\mathbb{E}\left[\|\widehat{g}^{\mathrm{LOO}}(W; N) - G\|_F^2\right] = \left[\frac{\kappa_r}{N} + \frac{d+1}{N(N-1)}\right] \|G\|_F^2. \tag{24}$$

For $G \neq 0$, comparing the variance of $2N$ LOO-ROLL directions with N antithetic directions at equal evaluation cost gives

$$\frac{\mathbb{E}[\|\widehat{g}^{\mathrm{LOO}}(W; 2N) - G\|_F^2]}{\mathbb{E}[\|\widehat{g}^{\mathrm{anti}}(W; N) - G\|_F^2]} = \frac{1}{2} + \frac{d+1}{2\kappa_r(2N-1)}. \tag{25}$$

Proofs of both propositions are given in Appendix C. Antithetic sampling and baseline subtraction are both established ways to reduce the variance of score-function estimators (Salimans et al., 2017; Williams, 1992; Wierstra et al., 2014). EGGROLL spends two evaluations on each independent direction and subtracts their fitnesses, canceling reward components shared by the pair. LOO-ROLL instead removes the population reward offset with a leave-one-out baseline, allowing every evaluation to use a new independent direction. At fixed N, estimating the baseline adds only an order-N^{-2} term to the order-N^{-1} directional variance. With $2N$ evaluations, LOO-ROLL uses twice as many independent directions and its variance approaches one half of EGGROLL's.

LOO-ROLL preserves EGGROLL's computational structure. Each evaluation uses a low-rank perturbation and the same shared base weights and batched forward computations. The implementation retains seed-based reconstruction, layerwise updates, and weight synchronization. The baseline requires neither additional model-sized state nor communication rounds. Evaluations previously spent on antithetic partners can therefore fund more directions or further updates.

6 Experiments

We begin by testing the theoretical predictions before examining their consequences for LLM post-training and LOO-ROLL. Complete protocols, additional figures, and numerical aggregates appear in Appendices D and F.

Sampling variance and stability follow the predictions. Monte Carlo variance agrees to within 0.9% at every tested rank with the exact finite-population prediction $\kappa_r\|G\|_F^2/N$ from Equation (20) and Table 16. The nonquadratic counterexample reproduces the predicted Jacobian eigenvalue 0.02925 at rank one and -0.00444 at rank two (Figure 2). Cubic and Fourier objectives recover the σ^2/r and fixed-radius $1/r$ corrections. On a quadratic objective, the finite-population formula predicts both the one-step change and the 25-step trajectories from the objective, step size, rank,

and population size (Figure 5). At step size 0.3 and $N = 16$, expected loss grows at rank one but decreases at rank eight and under dense Gaussian ES. Here, the instability arises entirely from sampling variance, since the mean fields agree. Appendix F.4 gives the full experiments.

Transformer blocks recover the finite-rank surcharge. We perturb 16×16 attention and MLP sub-blocks of Qwen3-0.6B, evaluating next-token prediction loss on a fixed minibatch. With 8,192 directions and three perturbation seeds, attention-block MSE ratios relative to dense Gaussian ES are $1.264, 1.127, 1.089, 1.028$ at ranks $1, 2, 4, 8$, close to the affine predictions $1.257, 1.128, 1.064, 1.032$ (Figure 3). Across ranks, the cosine similarity between the Monte Carlo estimate of the EGGROLL mean field and the backpropagated block gradient remains between 0.980 and 0.984. Thus the experiment resolves the predicted rank-dependent variance while finding little rank dependence in the mean direction at the tested radii. In a separate audit at $N = 128$, the mean updates obtained with and without population score standardization have cosine similarity 0.99970 at rank one (Table 17).

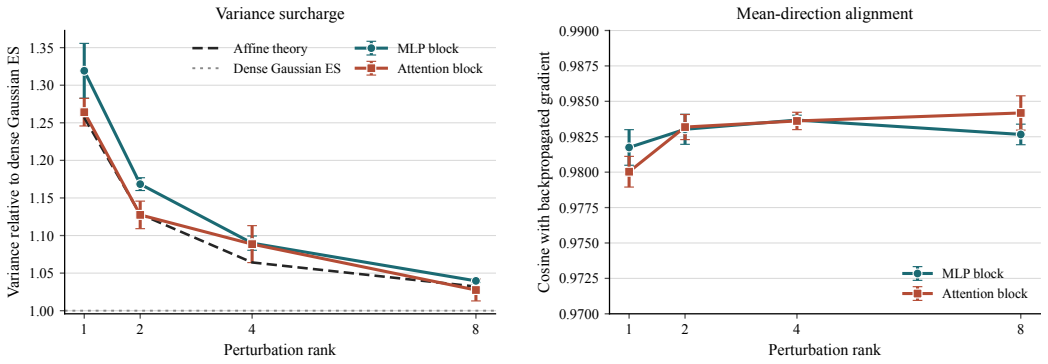


Figure 3: Finite-rank audit of two Qwen3-0.6B blocks. **Left:** single-direction MSE relative to dense Gaussian ES. The dashed curve is the exact affine prediction $[257 + 66/r]/257$ for a 16×16 block. **Right:** the mean of 8,192 antithetic directions remains aligned with the backpropagated block gradient. Points average four radii within each seed, and bars are standard errors over three seeds.

Increasing rank does not yield a reproducible reward advantage. Across Countdown and GSM8K comparisons, rank-eight minus rank-one intervals contain zero, while rank eight increases iteration time by 12–25% (Table 4). Raising σ from 10^{-3} to 4×10^{-3} reduces mean reward by 70–100% in the 0.6B and 1.7B studies (Figure 4 and Table 11). These results show that performance is substantially more sensitive to radius than rank in these settings.

Having validated the mean-field and variance predictions, we now evaluate LOO-ROLL in complete LLM post-training runs and ask whether its saved evaluations translate into better performance at equal wall time. We compare Qwen3 models from 0.6B to 14B (Yang et al., 2025) and SmolLM2-1.7B (Allal et al., 2025) over five paired training seeds per setting. Both methods start from the same pretrained checkpoint and retain the same learning rate, rank, radius, prompt batches, and score processing. Equal-iteration runs use the same direction count. Equal-evaluation runs double LOO-ROLL's directions. Equal-time runs continue LOO-ROLL to the paired EGGROLL runtime. Timing includes model loading and checkpoint writing, and accumulates all resumed training.

GSM8K (Cobbe et al., 2021) uses all 1,319 test questions and MATH-500 uses all 500 problems. MATH training excludes MATH-500. Countdown and next-token prediction use 256 shared validation examples. Checkpoints are selected by update count or stopping budget, without choosing the best test score. Table 8 provides exact hardware, direction counts, batches, updates, and measured runtimes.

At equal evaluation cost, update MSE is approximately halved. Across two transformer blocks, two radii, and three seeds, using the same number of directions increases MSE by at most 5.2%, falling to 0.8% at $N = 128$. Doubling the directions at equal evaluation cost reduces MSE by 49.0–50.1% across $N = 16, 32, 64, 128$ (Table 1), agreeing with Equation (25).

Additional updates yield gains across scale and task. Tables 2 and 6 report paired mean changes and sample standard deviations. At matched time, GSM8K accuracy rises from $38.10\% \pm 2.03$ to $63.00\% \pm 0.80$ at 0.6B, from $65.87\% \pm 0.78$ to $80.00\% \pm 0.78$ at 8B, and from $71.22\% \pm 0.67$ to

Table 1: Signed percentage change in LOO-ROLL MSE relative to antithetic EGGROLL. Positive values mean higher MSE. The equal-direction comparison uses N directions for both estimators, while the equal-evaluation comparison uses $2N$ LOO-ROLL directions against N antithetic directions. Values average two transformer blocks, two radii, and three seeds.

Antithetic directions N	Equal directions	Equal evaluations
16	$+5.2\%$	$\mathbf{-49.1\%}$
32	$+2.5\%$	$\mathbf{-49.0\%}$
64	$+1.2\%$	$\mathbf{-50.1\%}$
128	$+0.8\%$	$\mathbf{-49.5\%}$

Table 2: LOO-ROLL minus antithetic EGGROLL reward. GSM8K and MATH-500 results use the full test sets, while Countdown uses 256 shared examples. Values are mean $\pm$ standard deviation of five paired training-seed differences, in percentage points. Bold marks positive mean improvements.

Model	Task	Equal iterations	Equal evaluations	Equal wall time
SmolLM2-1.7B	Countdown	$-0.20\% \pm 0.22$	$-0.08\% \pm 0.34$	$-0.10\% \pm 0.46$
SmolLM2-1.7B	GSM8K	$\mathbf{+0.50\% \pm 0.64}$	$\mathbf{+0.23\% \pm 0.74}$	$\mathbf{+4.18\% \pm 1.54}$
Qwen3-0.6B	Countdown	$-4.09\% \pm 3.97$	$-1.05\% \pm 1.49$	$-0.62\% \pm 3.46$
Qwen3-0.6B	GSM8K	$\mathbf{+5.35\% \pm 1.76}$	$-4.82\% \pm 4.87$	$\mathbf{+24.90\% \pm 2.38}$
Qwen3-1.7B	Countdown	$-3.89\% \pm 1.22$	$-0.31\% \pm 1.56$	$\mathbf{+1.68\% \pm 1.03}$
Qwen3-1.7B	GSM8K	$0.00\% \pm 1.01$	$-0.26\% \pm 0.81$	$\mathbf{+2.85\% \pm 0.83}$
Qwen3-1.7B	MATH-500	$\mathbf{+0.56\% \pm 1.17}$	$-0.08\% \pm 1.25$	$\mathbf{+12.20\% \pm 1.56}$
Qwen3-8B	GSM8K	$\mathbf{+0.85\% \pm 0.66}$	$-1.08\% \pm 1.21$	$\mathbf{+14.13\% \pm 0.88}$
Qwen3-8B	MATH-500	$\mathbf{+0.24\% \pm 1.36}$	$\mathbf{+0.60\% \pm 1.65}$	$\mathbf{+8.36\% \pm 1.68}$
Qwen3-14B	GSM8K	$\mathbf{+0.41\% \pm 0.89}$	$-0.08\% \pm 0.79$	$\mathbf{+7.85\% \pm 1.35}$
Qwen3-14B	MATH-500	$\mathbf{+1.12\% \pm 0.70}$	$\mathbf{+0.60\% \pm 0.47}$	$\mathbf{+8.08\% \pm 1.23}$

$79.08\% \pm 0.92$ at 14B. On SmolLM2-1.7B, accuracy rises from $18.54\% \pm 0.51$ to $22.73\% \pm 1.15$. MATH-500 gains are 12.20, 8.36, and 8.08 percentage points at 1.7B, 8B, and 14B. At 14B, accuracy rises from $44.68\% \pm 0.72$ to $52.76\% \pm 1.13$, with every seed improving. LOO-ROLL completes 522.8 updates on average against EGGROLL's 200 in the same approximately 29-hour budget.

Next-token prediction loss in Table 6 also falls at equal time, by 1.20% and 2.16% relative to the EGGROLL baselines at 1.7B and 8B. The corresponding differences at equal evaluations are small, so the improvement follows from the additional updates that the saved computation permits. Across the fourteen primary settings, eleven gains are significant in individual paired tests, nine survive Holm correction, and none significantly favors EGGROLL (Table 10). The average paired time ratio is approximately 1.001. At equal evaluations, thirteen differences are unresolved, with a smaller 0.60-point gain on 14B MATH-500 ($p = 0.0459$). The improvements use the EGGROLL learning rate and perturbation radius, without separate tuning for LOO-ROLL. At a fixed direction count, the gap narrows as the population grows. The practical advantage comes from spending the saved evaluations on more directions or updates.

7 Conclusion

We showed that finite-rank EGGROLL follows a resolvent-filtered smoothed gradient that can be nonconservative and destabilize an optimum, yet is exact on quadratics and adds little sampling variance for wide matrices. We proposed LOO-ROLL, which replaces antithetic evaluations with leave-one-out baselines while retaining EGGROLL's low-rank, batched, forward-only computation. At equal evaluation cost it approximately halves update MSE, and across fourteen LLM post-training settings with up to 14B parameters, eleven matched-time gains are individually significant and none significantly favors EGGROLL. The main theory is limited to the unstandardized field and a local affine variance model, and the experiments to two model families and three tasks. Future work should treat adaptive score processing and dynamic allocation of rank, radius, and population.

AI USE STATEMENT

Generative AI tools assisted with literature discovery, experimental orchestration, and language and formatting revisions. The authors independently checked all citations. The authors made all final scientific and editorial decisions and take responsibility for the complete contents of the submission.

ETHICS STATEMENT

This work studies optimization methods using publicly available pretrained models and benchmark datasets. It involves no human subjects, private data, or deployment, and raises no ethical concerns beyond those generally associated with language-model research.

REPRODUCIBILITY STATEMENT

All theorem assumptions are stated in the main text, and complete proofs are provided in the appendix. The supplementary material includes the complete training and evaluation code, experiment configurations, and analysis scripts. LOO-ROLL is implemented as a 37-line patch to the EGGROLL vLLM code at commit `bcc215e8784f5f44d24985145c0a71e74283cf1f`, comprising 15 insertions and 22 deletions in one file. The algorithmic patch is provided separately from the shared model-compatibility changes so that the modification can be inspected and reproduced directly. Reward-based runs use rank one, $\sigma = 10^{-3}$, and learning rate 2×10^{-4}. Next-token prediction uses 32 directions and $\sigma = 10^{-2}$ at 0.6B and 1.7B or 3×10^{-3} at 8B, with learning rate 4×10^{-4} only for the 8B setting. Equal-evaluation comparisons double the LOO-ROLL direction count. Experiments use one NVIDIA H100 80GB GPU and 14 CPU cores, except 14B MATH training, which uses two H100 GPUs and 28 CPU cores for both methods. Paired comparisons begin from the same pretrained checkpoint and share seeds, prompt batches, learning rates, ranks, perturbation radii, and score processing. Timing includes model loading, updates, and checkpoint writing, with elapsed time accumulated across resumed runs. Checkpoints are fixed by update count or stopping budget rather than test performance. GSM8K and MATH-500 use their complete 1,319- and 500-example test sets, with MATH-500 excluded from training. Countdown and next-token prediction use 256 shared validation examples. Table 8 reports direction counts, batch sizes, update budgets, and measured runtimes for every primary comparison.

REFERENCES

P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

Arash Ahmadian, Chris Cremer, Matthias Gallé, Marzieh Fadaee, Julia Kreutzer, Olivier Pietquin, Ahmet Üstün, and Sara Hooker. Back to basics: Revisiting REINFORCE-style optimization for learning from human feedback in LLMs. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 12248–12267, 2024.

Loubna Ben Allal, Anton Lozhkov, Elie Bakouch, Gabriel Martín Blázquez, Guilherme Penedo, Lewis Tunstall, Andrés Marafioti, Hynek Kydlíček, Agustín Piqueres Lajarín, Vaibhav Srivastav, et al. SmolLM2: When smol goes big–data-centric training of a small language model. *arXiv preprint arXiv:2502.02737*, 2025.

Krishnakumar Balasubramanian and Saeed Ghadimi. Zeroth-order nonconvex stochastic optimization: Handling constraints, high dimensionality, and saddle points. *Foundations of Computational Mathematics*, 22(1):35–76, 2022.

Heinz H Bauschke and Patrick L Combettes. Correction to: convex analysis and monotone operator theory in Hilbert spaces. In *Convex analysis and monotone operator theory in Hilbert spaces*, pp. C1–C4. Springer, 2020.

Hans-Georg Beyer and Hans-Paul Schwefel. Evolution strategies–a comprehensive introduction. *Natural computing*, 1(1):3–52, 2002.

Krzysztof Choromanski, Mark Rowland, Vikas Sindhwani, Richard Turner, and Adrian Weller. Structured evolution with compact architectures for scalable policy optimization. In *International Conference on Machine Learning*, pp. 970–978. PMLR, 2018.

Krzysztof M Choromanski, Aldo Pacchiano, Jack Parker-Holder, Yunhao Tang, and Vikas Sindhwani. From complexity to simplicity: Adaptive ES-active subspaces for blackbox optimization. *Advances in Neural Information Processing Systems*, 32, 2019.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Tri Dao, Dan Fu, Stefano Ermon, Atri Rudra, and Christopher Ré. Flashattention: Fast and memory-efficient exact attention with IO-awareness. *Advances in neural information processing systems*, 35:16344–16359, 2022.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. QLoRA: Efficient finetuning of quantized LLMs. *Advances in neural information processing systems*, 36:10088–10115, 2023.

John C Duchi, Michael I Jordan, Martin J Wainwright, and Andre Wibisono. Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, 61(5):2788–2806, 2015.

Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 1999.

Robert Gaunt. Products of normal, beta and gamma random variables: Stein operators and distributional theory. *Brazilian Journal of Probability and Statistics*, 32(2):437–466, April 2018a. doi: 10.1214/16-BJPS349.

Robert E. Gaunt. On Stein's method for products of normal random variables and zero bias couplings. *Bernoulli*, 23(4B):3311–3345, 2018b. ISSN 13507265. URL http://www.jstor.org/stable/26491760.

Tanmay Gautam, Youngsuk Park, Hao Zhou, Parameswaran Raman, and Wooseok Ha. Variance-reduced zeroth-order methods for fine-tuning language models. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of

Proceedings of Machine Learning Research, pp. 15180–15208. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/gautam24a.html.

Nikolaus Hansen. The CMA evolution strategy: A tutorial. *arXiv preprint arXiv:1604.00772*, 2016.

Nicholas J Higham. *Functions of matrices: theory and computation*. SIAM, 2008.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pp. 65–70, 1979.

William Hoy, Binxu Wang, and Xu Pan. Matching accuracy, different geometry: Evolution strategies vs GRPO in LLM post-training. In *Third Conference on Language Modeling*, 2026. URL https://openreview.net/forum?id=HsyvUQdeGS.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Leon Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139, 1918.

Jake Levi, Seth Nabarro, and Mark van der Wilk. On the scalability of forward gradients, evolution strategies, and control variates. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=s6g8yZimHE.

Shih-Yang Liu, Chien-Yi Wang, Hongxu Yin, Pavlo Molchanov, Yu-Chiang Frank Wang, Kwang-Ting Cheng, and Min-Hung Chen. DoRA: Weight-decomposed low-rank adaptation. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 32100–32121. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/liu24bn.html.

Niru Maheswaranathan, Luke Metz, George Tucker, Dami Choi, and Jascha Sohl-Dickstein. Guided evolutionary strategies: Augmenting random search with surrogate gradients. In *International Conference on Machine Learning*, pp. 4264–4273. PMLR, 2019.

Sadhika Malladi, Tianyu Gao, Eshaan Nichani, Alex Damian, Jason D Lee, Danqi Chen, and Sanjeev Arora. Fine-tuning language models with just forward passes. *Advances in Neural Information Processing Systems*, 36:53038–53075, 2023.

Robb J Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons, 2009.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017.

Ingo Rechenberg. Evolutionsstrategien. In *Simulationsmethoden in der Medizin und Biologie: Workshop, Hannover, 29. Sept.–1. Okt. 1977*, pp. 83–114. Springer, 1978.

Tim Salimans, Jonathan Ho, Xi Chen, Szymon Sidor, and Ilya Sutskever. Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864*, 2017.

Bidipta Sarkar, Mattie Fellows, Juan Agustin Duque, Alistair Letcher, Antonio León Villares, Anya Sims, Clarisse Wibault, Dmitry Samsonov, Dylan Cope, Jarek Luca Liesen, Kang Li, Lukas Seier, Theo Wolf, Uljad Berdica, Valentin Mohl, Alexander David Goldie, Aaron Courville, Karin Sevegnani, Shimon Whiteson, and Jakob Nicolaus Foerster. Evolution strategies at the hyperscale. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=bfVJ4GsHrO.

Hans-Paul Schwefel. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie: mit einer vergleichenden Einführung in die Hill-Climbing-und Zufallsstrategie*, volume 57. Springer, 1977.

Hyunseok Seung, Jaewoo Lee, and Hyunsuk Ko. Low-rank curvature for zeroth-order optimization in LLM fine-tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 25235–25242, 2026.

James C Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE transactions on automatic control*, 37(3):332–341, 1992.

Michael Spivak. *Calculus on manifolds: a modern approach to classical theorems of advanced calculus*. Westview Press, 1965.

Charles M Stein. Estimation of the mean of a multivariate normal distribution. *The annals of Statistics*, pp. 1135–1151, 1981.

Yunhao Tang, Krzysztof Choromanski, and Alp Kucukelbir. Variance reduction for evolution strategies via structured control variates. In *International Conference on Artificial Intelligence and Statistics*, pp. 646–656. PMLR, 2020.

Sharan Vaswani, Francis Bach, and Mark Schmidt. Fast and faster convergence of SGD for over-parameterized models and an accelerated perceptron. In *The 22nd international conference on artificial intelligence and statistics*, pp. 1195–1204. PMLR, 2019.

Jun-Kun Wang, Xiaoyun Li, and Ping Li. Zeroth order optimization by a mixture of evolution strategies, 2020. URL https://openreview.net/forum?id=SklE_CNFPr.

Zhongruo Wang, Krishnakumar Balasubramanian, Shiqian Ma, and Meisam Razaviyayn. Zeroth-order algorithms for nonconvex–strongly-concave minimax problems with improved complexities. *Journal of Global Optimization*, 87(2):709–740, 2023.

Daan Wierstra, Tom Schaul, Tobias Glasmachers, Yi Sun, Jan Peters, and Jürgen Schmidhuber. Natural evolution strategies. *The Journal of Machine Learning Research*, 15(1):949–980, 2014.

Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256, 1992.

Samuel Williams, Andrew Waterman, and David Patterson. Roofline: an insightful visual performance model for multicore architectures. *Communications of the ACM*, 52(4):65–76, 2009.

Yinggan Xu, Kajetan Schweighofer, Risto Miikkulainen, and Xin Qiu. Quantized evolution strategies: High-precision fine-tuning of quantized LLMs at low-precision cost. *arXiv preprint arXiv:2602.03120*, 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.

Qingru Zhang, Minshuo Chen, Alexander Bukharin, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. Adaptive budget allocation for parameter-efficient fine-tuning. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=lq62uWRJjiY.

Yan Zhang, Yi Zhou, Kaiyi Ji, Yi Shen, and Michael M Zavlanos. Boosting one-point derivative-free online optimization via residual feedback. *IEEE Transactions on Automatic Control*, 69(9): 6309–6316, 2024.

Yanjun Zhao, Sizhe Dang, Haishan Ye, Guang Dai, Yi Qian, and Ivor Tsang. Second-order fine-tuning without pain for LLMs: A Hessian informed zeroth-order optimizer. In *International Conference on Learning Representations*, volume 2025, pp. 43496–43520, 2025.

APPENDIX TABLE OF CONTENTS

A Additional related work **16**

B Mean-field and finite-rank theory **16**

B.1 Proof of Theorem 2.1 16

B.2 Proof of Theorem 2.2 17

B.3 Proof of Corollary 2.3 18

B.4 Proofs and examples for conservativity 18

B.4.1 Proof of Theorem 3.1 18

B.4.2 An explicit nonconservative mode 19

B.4.3 Proof of Proposition 3.2 20

B.5 Proofs and bounds near quadratic objectives 20

B.5.1 Proof of Proposition 3.3 20

B.5.2 Proof of Proposition 3.4 21

B.5.3 Nonasymptotic bounds under standard smoothness 21

B.6 Complete covariance calculation 21

B.7 Fixed-radius rank expansion 23

B.8 Proof of Theorem 4.2 24

C Proofs for LOO-ROLL **25**

C.1 Proof of Proposition 5.1 25

C.2 Proof of Proposition 5.2 25

D Additional LLM comparisons **25**

E Network extensions and further theoretical results **27**

E.1 Secondary corrections suggested by the analysis 28

E.1.1 Rank extrapolation 28

E.1.2 Variance reduction with a reference point 29

E.2 Convergence consequences of the field bounds 29

E.2.1 Finite-population convergence 30

E.3 Fourth moments of the Gaussian-product law 32

F Complete empirical results **33**

F.1 Training budgets 33

F.2 Paired statistical comparisons 33

F.3 Rank comparisons and transformer audits 33

F.4 Controlled numerical experiments 36

G Population centering and standardization **37**

G.1 Transformer-block audit 38

A Additional related work

Random-direction finite differences and simultaneous perturbation are classical gradient-free methods (Spall, 1992; Nesterov & Spokoiny, 2017; Duchi et al., 2015). Their nonconvex analyses make the dependence on dimension and smoothing radius explicit (Balasubramanian & Ghadimi, 2022). MeZO shows that an in-place two-forward-pass implementation can fine-tune large language models with inference-level memory, and subsequent work adds stochastic variance reduction (Malladi et al., 2023; Gautam et al., 2024). Residual feedback uses a stored function value to obtain a one-query estimator for online zeroth-order optimization (Zhang et al., 2024). These methods use dense or coordinatewise perturbations. EGGROLL instead constructs each matrix perturbation as a low-rank Gaussian product and batches the resulting population during inference, giving rise to the finite-rank population field analyzed here.

Baselines have long been used to reduce the variance of score-function and policy-gradient estimators without changing their expectation (Williams, 1992). In REINFORCE leave-one-out, the reward of each sampled response is compared with the average reward of the other responses generated from the same prompt (Ahmadian et al., 2024). LOREN also uses REINFORCE leave-one-out in curvature-aware zeroth-order LLM tuning (Seung et al., 2026). LOO-ROLL specializes the baseline to EGGROLL's Gaussian-product population and evaluates the resulting alternative to its antithetic training estimator.

The original EGGROLL analysis proves a fixed-rank, high-dimensional vanishing-radius guarantee and convergence of its Gaussian-score update to dense Gaussian ES at rate $O(1/r)$ under regularity assumptions (Sarkar et al., 2026). Our results concern the complementary regime of finite dimension, finite radius, and finite population. Recent work compares full Gaussian ES with gradient-based LLM post-training and studies its behavior in flat, linear, and quadratic geometries (Hoy et al., 2026; Levi et al., 2026).

Orthogonal and structured perturbations can reduce estimator variance or the cost of applying a direction, while structured control variates use additional problem information (Choromanski et al., 2018; Tang et al., 2020). Guided and active-subspace ES concentrate exploration in learned subspaces (Maheswaranathan et al., 2019; Choromanski et al., 2019). Natural ES and CMA-ES adapt a Gaussian search law (Wierstra et al., 2014; Hansen, 2016). EGGROLL makes a different choice, where the Gaussian-product law is fixed by the desired matrix computation. We analyze the mean field and sampling error that follow from that choice.

Stein operators for scalar products of independent normal variables, and for sums and broader products built from such variables, are well established (Gaunt, 2018b;a). These identities act on one scalar product-normal variable. The entries of $AB^\top$ instead share row and column factors, so the matrix law couples different coordinates. Our operator identity keeps this dependence, produces a matrix differential operator acting on the population field, and connects its resolvent to finite-rank optimization dynamics.

Wang et al. (2020) show that convergence can fail for structured zeroth-order perturbations and proposed mixing them with isotropic Gaussian directions. Their use of isotropy concerns an ambient Gaussian component. EGGROLL already has identity covariance, yet its law is not spherically symmetric in the vectorized matrix space. Our conservativity theorem isolates this stronger distributional requirement. Exact spherical symmetry cannot be achieved by a nondegenerate law supported entirely on rank-deficient matrices when $m, n \geq 2$, since a uniformly rotated nonzero matrix is full rank almost surely after reshaping. Occasional dense directions could supply an ambient isotropic component, though those evaluations would no longer have the purely low-rank cost of EGGROLL.

B Mean-field and finite-rank theory

B.1 Proof of Theorem 2.1

Proof of Theorem 2.1. For a pair $(a, b) \in \mathbb{R}^m \times \mathbb{R}^n$, condition on a:

$$\mathbb{E}_b\left[e^{i\langle T,\, ab^\top\rangle_F/\sqrt{r}}\right] = \mathbb{E}_b\left[e^{i\,\mathrm{tr}(T^\top ab^\top)/\sqrt{r}}\right] = \mathbb{E}_b\left[e^{ib^\top T^\top a/\sqrt{r}}\right]. \tag{26}$$

Because we have conditioned on a, $c := T^\top a \in \mathbb{R}^n$ is fixed. Furthermore, since $b \sim \mathcal{N}(0, I_n)$, $b^\top c$ is a one-dimensional Gaussian, i.e.,

$$b^\top c \sim \mathcal{N}(0, \|c\|^2). \tag{27}$$

Hence,

$$\mathbb{E}_b\big[e^{ib^\top T^\top a/\sqrt{r}}\big] = \mathbb{E}_b\big[e^{ib^\top c/\sqrt{r}}\big] = \exp\Big(-\frac{1}{2r}\|c\|^2\Big) = \exp\Big(-\frac{1}{2r}\|T^\top a\|^2\Big), \tag{28}$$

by the Gaussian characteristic function. We have thus obtained the expectation conditional on a. To have the full expectation, we now evaluate

$$\mathbb{E}_a\Big[\exp\Big(-\frac{1}{2r}\|T^\top a\|^2\Big)\Big] = \mathbb{E}_a\Big[\exp\Big(-\frac{1}{2r}a^\top TT^\top a\Big)\Big]. \tag{29}$$

By the Gaussian quadratic-form identity (Muirhead, 2009), this is

$$\det\Big(I_m + \frac{TT^\top}{r}\Big)^{-1/2}. \tag{30}$$

This means that one rank-one term contributes

$$\mathbb{E}\big[e^{i\langle T,\, ab^\top\rangle_F/\sqrt{r}}\big] = \det\Big(I_m + \frac{TT^\top}{r}\Big)^{-1/2}. \tag{31}$$

Now,

$$E_r = \frac{1}{\sqrt{r}}\sum_{s=1}^{r} a_s b_s^\top, \tag{32}$$

so

$$\Phi_r(T) = \mathbb{E}\big[e^{i\langle T,\, E_r\rangle_F}\big] = \prod_{s=1}^{r}\mathbb{E}\big[e^{i\langle T,\, a_s b_s^\top\rangle_F/\sqrt{r}}\big] = \det\Big(I_m + \frac{TT^\top}{r}\Big)^{-r/2}, \tag{33}$$

since all r pairs (a_s, b_s) are independent. We finish by computing the gradient of the resulting determinant. Write $S = I_m + TT^\top/r$. Since $\log\Phi_r(T) = -(r/2)\log\det S$, its differential is

$$d\log\Phi_r(T) = -\frac{r}{2}\operatorname{tr}(S^{-1}dS), \qquad dS = \frac{(dT)T^\top + T(dT)^\top}{r}. \tag{34}$$

Substituting dS cancels the factor r and gives

$$d\log\Phi_r(T) = -\frac{1}{2}\operatorname{tr}\left(S^{-1}(dT)T^\top + S^{-1}T(dT)^\top\right). \tag{35}$$

The matrix S^{-1} is symmetric. Cyclicity of the trace and $\operatorname{tr}(A(dT)^\top) = \operatorname{tr}(A^\top dT)$ therefore make the two terms equal, so

$$d\log\Phi_r(T) = -\operatorname{tr}\left(T^\top S^{-1}dT\right) = \langle -S^{-1}T,\, dT\rangle_F. \tag{36}$$

By the defining identity $dh(T) = \langle\nabla_T h(T),\, dT\rangle_F$ for the Frobenius gradient, $\nabla_T\log\Phi_r(T) = -S^{-1}T$. Finally, $\nabla_T\Phi_r(T) = \Phi_r(T)\nabla_T\log\Phi_r(T)$, which proves Equation (8). □

B.2 Proof of Theorem 2.2

When both fields are square-integrable and $g^f_{r,\sigma}$ belongs to the domain of $\mathcal{L}$, taking the L^2 inner product of Equation (13) with $g^f_{r,\sigma}$ gives the energy identity

$$\big\langle g^f_{r,\sigma}, \nabla F_{r,\sigma}\big\rangle_{L^2} = \|g^f_{r,\sigma}\|^2_{L^2} + \frac{\sigma^2}{r}\|\mathcal{L}^{1/2}g^f_{r,\sigma}\|^2_{L^2}. \tag{37}$$

Proof of Theorem 2.2. For a tempered distribution f, averaging its translations against the perturbation law gives

$$\widehat{F}_{r,\sigma}(T) = \Phi_r(\sigma T)\widehat{f}(T), \qquad \widehat{g}^f_{r,\sigma}(T) = i\Phi_r(\sigma T)\left(I_m + \frac{\sigma^2 TT^\top}{r}\right)^{-1} T\widehat{f}(T) \tag{38}$$

in the distributional sense. The Fourier multiplier of $I + (\sigma^2/r)\mathcal{L}$ is left multiplication by $I_m + \sigma^2 TT^\top/r$. Applying it to the second expression gives $iT\widehat{F}_{r,\sigma}(T)$, the Fourier transform of $\nabla F_{r,\sigma}$. This proves Equation (13) as an identity of tempered distributions.

Under the stated Fourier moment condition, substituting the Fourier representation of f into Equation (4), interchanging expectation and integration, and using Equation (8) gives the classical formula

$$g^f_{r,\sigma}(W) = \int i\Phi_r(\sigma T)\left(I_m + \frac{\sigma^2 TT^\top}{r}\right)^{-1} T e^{i\langle T,\, W\rangle_F}\, \mu_f(dT), \tag{39}$$

whereas

$$\nabla F_{r,\sigma}(W) = \int i\Phi_r(\sigma T) T e^{i\langle T,\, W\rangle_F}\, \mu_f(dT). \tag{40}$$

The moment condition justifies these operations and makes the identity pointwise. At each frequency T, the multiplier $I_m + \sigma^2 TT^\top/r$ is symmetric with eigenvalues at least one, making its inverse positive, contractive, and firmly nonexpansive. Parseval's identity equates L^2 inner products of fields in parameter space with the corresponding inner products of their Fourier coefficients (Folland, 1999). Applying this identity to $(I + (\sigma^2/r)\mathcal{L})g = \nabla F$ gives Equation (37). □

B.3 Proof of Corollary 2.3

Proof of Corollary 2.3. Set $\alpha = \sigma^2/r$ and write $J = J_{r,\sigma} = (I + \alpha\mathcal{L})^{-1}$. On the domain of $\mathcal{L}$, the resolvent identity gives

$$J - I = -\alpha J\mathcal{L}. \tag{41}$$

To see this directly, take any matrix vector field h for which $\mathcal{L}h$ is defined. Since J is the inverse of $I + \alpha\mathcal{L}$,

$$J(h + \alpha\mathcal{L}h) = h. \tag{42}$$

Linearity of J then gives $Jh - h = -\alpha J\mathcal{L}h$. Applying this equality with $h = \nabla F_{r,\sigma}$ and using $g^f_{r,\sigma} = J\nabla F_{r,\sigma}$ yields

$$g^f_{r,\sigma} - \nabla F_{r,\sigma} = -\frac{\sigma^2}{r} J_{r,\sigma}\mathcal{L}\nabla F_{r,\sigma}, \tag{43}$$

which proves the exact formula. The Fourier calculation in Theorem 2.2 also shows that J multiplies every eigendirection at every frequency by a number in $[0, 1]$. Parseval's identity therefore gives $\|Jh\|_{L^2} \le \|h\|_{L^2}$ for every square-integrable field h. Taking L^2 norms in the exact formula and applying this inequality to $h = \mathcal{L}\nabla F_{r,\sigma}$ proves Equation (14). □

B.4 Proofs and examples for conservativity

B.4.1 Proof of Theorem 3.1

For the proof, write $\Phi(t) = \mathbb{E}[e^{it^\top Z}]$ and define

$$(\mathcal{T}_{Z,\sigma} f)(x) = \frac{1}{\sigma}\mathbb{E}[Z f(x + \sigma Z)]. \tag{44}$$

We prove the equivalence of the following statements:

(i) for every $\sigma > 0$ and every real trigonometric polynomial f, the field $\mathcal{T}_{Z,\sigma} f$ is conservative;
(ii) for every $t \in \mathbb{R}^d$, $\nabla\Phi(t) = \lambda(t)t$ for some complex scalar $\lambda(t)$;
(iii) Φ is radial;
(iv) the law of Z is spherically symmetric.

Proof of Theorem 3.1. We first prove the equivalence of (i) and (ii). Fix $t \in \mathbb{R}^d$ and consider the complex Fourier mode $f_t(x) = e^{it^\top x}$. The real and imaginary parts of this mode are $\cos(t^\top x)$ and $\sin(t^\top x)$, so applying $\mathcal{T}_{Z,\sigma}$ to f_t records its action on both real modes. From the definition of Φ,

$$\nabla\Phi(\sigma t) = i\mathbb{E}\left[Z e^{i\sigma t^\top Z}\right]. \tag{45}$$

Consequently,

$$(\mathcal{T}_{Z,\sigma} f_t)(x) = \frac{e^{it^\top x}}{\sigma} \mathbb{E}\left[Z e^{i\sigma t^\top Z}\right] = \frac{1}{i\sigma} \nabla\Phi(\sigma t) e^{it^\top x}. \tag{46}$$

A vector field of the form $ce^{it^\top x}$ is a gradient field if and only if its coefficient $c \in \mathbb{C}^d$ is parallel to t. Indeed, if $c = \gamma t$, then

$$ce^{it^\top x} = \nabla\left(\frac{\gamma}{i} e^{it^\top x}\right). \tag{47}$$

Conversely, the Jacobian of $ce^{it^\top x}$ is $ict^\top e^{it^\top x}$. Any continuously differentiable gradient field has a symmetric Jacobian because its Jacobian is the Hessian of its potential. For $t \neq 0$, symmetry of the present Jacobian is equivalent to $ct^\top = tc^\top$. Choose an index j for which $t_j \neq 0$. Comparing the (i, j) entries gives $c_i t_j = t_i c_j$ for every i, and hence $c = (c_j/t_j)t$. Thus the Jacobian is symmetric only when c and t are parallel. Equation (46) is therefore conservative precisely when $\nabla\Phi(\sigma t)$ is parallel to t. Since $\sigma > 0$ and t are arbitrary, this condition is equivalent to (ii). Linearity then extends the conclusion from individual sine and cosine modes to every trigonometric polynomial, proving (i)⇔(ii). The zero-frequency mode causes no exception because $\nabla\Phi(0) = i\mathbb{E}[Z] = 0$.

We next prove (ii)⇔(iii). Suppose (ii) holds. If $v^\top t = 0$, then v is tangent at t to the sphere of radius $\|t\|$, and

$$v^\top \nabla\Phi(t) = \lambda(t) v^\top t = 0. \tag{48}$$

Thus Φ has zero directional derivative along every tangent direction to each sphere. Spheres are connected when $d \geq 2$, so Φ is constant on every sphere and hence depends only on $\|t\|$. This is precisely radiality. Conversely, if Φ is radial, then it is constant in every direction tangent to a sphere. The gradient is therefore orthogonal to every tangent direction and must be parallel to the radial direction t, which gives (ii).

Finally, (iii)⇔(iv) is the standard characteristic-function characterization of spherical symmetry (Muirhead, 2009). If the law of Z is spherically symmetric, then QZ and Z have the same law for every orthogonal matrix Q, and hence $\Phi(Q^\top t) = \Phi(t)$. Therefore Φ depends only on $\|t\|$. Conversely, if Φ is radial, the characteristic function of QZ is

$$\mathbb{E}\left[e^{it^\top QZ}\right] = \Phi(Q^\top t) = \Phi(t). \tag{49}$$

Uniqueness of characteristic functions then implies that QZ and Z have the same law, so the law is spherically symmetric. The assumption $d \geq 2$ is necessary because in one dimension every continuous scalar field is conservative, regardless of whether the perturbation law is symmetric. □

B.4.2 An explicit nonconservative mode

For the Gaussian-product law, spherical symmetry fails because different singular directions are attenuated by different amounts. A single diagonal Fourier mode is enough to display the resulting nonconservativity.

Corollary B.1 (A bounded smooth nonconservative example)**.** *Let $m = n = 2$, $T = \mathrm{diag}(1, 2)$, and $f(W) = \cos\langle T, W\rangle_F$. Then*

$$g^f_{r,\sigma}(W) = -\Phi_r(\sigma T)\left(I + \frac{\sigma^2 T T^\top}{r}\right)^{-1} T \sin\langle T, W\rangle_F. \tag{50}$$

For every finite r and $\sigma > 0$, its Jacobian is nonsymmetric at $W = 0$ and at every W with $\cos\langle T, W\rangle_F \neq 0$. Hence, the field is not conservative.

Proof. Recall that $J(T) = (I + \sigma^2 TT^\top/r)^{-1} T$. After vectorization, differentiating Equation (50) gives

$$Dg^f_{r,\sigma}(W) = -\Phi_r(\sigma T) \cos\langle T, W\rangle_F \, \mathrm{vec}(J(T))\, \mathrm{vec}(T)^\top. \tag{51}$$

A nonzero rank-one matrix $uv^\top$ is symmetric only if u and v are parallel, since its column and row spaces are spanned by u and v, respectively, and these spaces coincide for a symmetric matrix. Here,

$$J(T) = \mathrm{diag}\left(\frac{1}{1 + \sigma^2/r}, \frac{2}{1 + 4\sigma^2/r}\right), \tag{52}$$

which is not a scalar multiple of $T = \mathrm{diag}(1, 2)$ for finite r and $\sigma > 0$. The displayed Jacobian is therefore nonsymmetric whenever $\cos\langle T, W\rangle_F \neq 0$. □

B.4.3 Proof of Proposition 3.2

Proof of Proposition 3.2. Let $t_q = \mathrm{vec}(T_q)$. The Hessian of f at zero is

$$-t_1 t_1^\top - \frac{1}{4} t_2 t_2^\top - \varepsilon I_4 \prec 0. \tag{53}$$

Smoothing multiplies the first two cosine coefficients by $\Phi_1(T_1) = \Phi_1(T_2) = 1/(3\sqrt{2})$ and every coordinate cosine by $\Phi_1(e_{ij}) = 1/\sqrt{2}$:

$$\nabla^2 F_{1,1}(0) = -\frac{1}{3\sqrt{2}} \left(t_1 t_1^\top + \frac{1}{4} t_2 t_2^\top \right) - \frac{\varepsilon}{\sqrt{2}} I_4 \prec 0. \tag{54}$$

Hence, the origin remains a strict local maximum after smoothing.

We next compute the Jacobian of the population field. For a cosine mode with frequency T, Equation (50) gives the contribution

$$-\Phi_1(T)\, \mathrm{vec}(J(T))\, \mathrm{vec}(T)^\top \tag{55}$$

at the origin. We have

$$J(T_1) = \begin{pmatrix} 0 & -1/3 \\ -1/3 & 1/6 \end{pmatrix}, \qquad J(T_2) = \begin{pmatrix} 0 & 1/3 \\ 1/3 & 1/6 \end{pmatrix}. \tag{56}$$

Therefore, the contribution of the first two cosine modes is

$$A_0 = -\frac{1}{3\sqrt{2}} \left(\mathrm{vec}(J(T_1)) t_1^\top + \frac{1}{4} \mathrm{vec}(J(T_2)) t_2^\top \right). \tag{57}$$

Substituting the displayed matrices and expanding the determinant gives

$$\det(\lambda I_4 - A_0) = \lambda^2 \left(\lambda^2 + \frac{35\sqrt{2}}{144} \lambda - \frac{1}{81} \right). \tag{58}$$

The positive root of the quadratic factor is

$$\lambda_+(A_0) = \frac{\sqrt{1737} - 35}{144\sqrt{2}}. \tag{59}$$

For a coordinate mode $T = e_{ij}$, we have $J(e_{ij}) = e_{ij}/2$ and $\Phi_1(e_{ij}) = 1/\sqrt{2}$. The corresponding Jacobian contribution is $-\varepsilon\, \mathrm{vec}(e_{ij})\, \mathrm{vec}(e_{ij})^\top/(2\sqrt{2})$. Summing over the four coordinates gives $-\varepsilon I_4/(2\sqrt{2})$. This scalar shift subtracts $\varepsilon/(2\sqrt{2})$ from every eigenvalue of A_0, so the larger eigenvalue of the full Jacobian is

$$\frac{\sqrt{1737} - 35}{144\sqrt{2}} - \frac{\varepsilon}{2\sqrt{2}}, \tag{60}$$

which is positive at $\varepsilon = 0.01$. □

B.5 Proofs and bounds near quadratic objectives

B.5.1 Proof of Proposition 3.3

Proof of Proposition 3.3. Applying the antithetic identity gives

$$g^f_{r,\sigma}(W) = \mathbb{E}\left[E_r \langle \nabla f(W),\, E_r \rangle_F \right] = \nabla f(W), \tag{61}$$

where the last equality follows from $\mathbb{E}[\mathrm{vec}(E_r)\, \mathrm{vec}(E_r)^\top] = I$. □

B.5.2 Proof of Proposition 3.4

Proof of Proposition 3.4. Taylor expansion of the antithetic update gives

$$g^f_{r,\sigma}(W) = \nabla f(W) + \frac{\sigma^2}{6}\mathbb{E}\left[E_r D^3 f(W)[E_r, E_r, E_r]\right] + R_{r,\sigma}(W), \tag{62}$$

where identity covariance was used for the first term. The order-σ^2 expectation is

$$\mathbb{E}\left[E_r D^3 f(W)[E_r, E_r, E_r]\right] = 3\nabla\Delta f(W) + \frac{6}{r}\mathcal{D}f(W). \tag{63}$$

This identity follows by writing $E_r = r^{-1/2}\sum_{q=1}^r a_q b_q^\top$ and applying the Gaussian fourth-moment formula (Isserlis, 1918) separately to the independent row and column factors. The terms that also occur for a dense Gaussian matrix give $3\nabla\Delta f$. The remaining product-law contribution gives $6\mathcal{D}f/r$. Substitution into Equation (62) yields the two order-σ^2 terms in Equation (19). Appendix E.3 provides a more in-depth version of this calculation, including the provenance of the coefficient $1/r$.

Finally, let $M_5 = \sup_W \|D^5 f(W)\|_{\mathrm{op}}$. Taylor's theorem bounds the matrix-valued remainder in Equation (62) by

$$\|R_{r,\sigma}(W)\|_F \le \frac{\sigma^4 M_5}{120}\mathbb{E}[\|E_r\|_F^6]. \tag{64}$$

□

B.5.3 Nonasymptotic bounds under standard smoothness

For $d = mn$, the exact fourth moment of the perturbation norm is

$$\mu_4(r) := \mathbb{E}[\|E_r\|_F^4] = d(d+2) + \frac{2d(m+n+1)}{r}. \tag{65}$$

Lemma B.2 (Field error under standard smoothness)**.** *Suppose h is L-smooth. Then, for every W,*

$$\|g^h_{r,\sigma}(W) - \nabla h(W)\|_F \le \frac{L\sigma}{2}\mu_4(r)^{3/4}. \tag{66}$$

If, in addition, the Hessian is M-Lipschitz as an operator, meaning that $\|\nabla^2 h(W) - \nabla^2 h(V)\|_{\mathrm{op}} \le M\|W - V\|_F$ for all W, V, then

$$\|g^h_{r,\sigma}(W) - \nabla h(W)\|_F \le \frac{M\sigma^2}{6}\mu_4(r). \tag{67}$$

Proof. For any matrix E, L-smoothness gives

$$\left|\frac{h(W+\sigma E) - h(W-\sigma E)}{2\sigma} - \langle \nabla h(W),\, E\rangle_F\right| \le \frac{L\sigma}{2}\|E\|_F^2. \tag{68}$$

Multiplying by $E = E_r$, taking expectations, and using identity covariance yields Equation (66), since $\mathbb{E}[\|E_r\|_F^3] \le \mu_4(r)^{3/4}$. If the Hessian is M-Lipschitz, the second-order Taylor remainders at $W \pm \sigma E$ are bounded by $M\sigma^3\|E\|_F^3/6$. The quadratic terms cancel in the antithetic difference, leaving

$$\left|\frac{h(W+\sigma E) - h(W-\sigma E)}{2\sigma} - \langle \nabla h(W),\, E\rangle_F\right| \le \frac{M\sigma^2}{6}\|E\|_F^3. \tag{69}$$

Averaging after multiplication by E_r gives Equation (67). □

B.6 Complete covariance calculation

The total-variance formula in Theorem 4.1 follows from the complete entrywise covariance below.

Theorem B.3 (Exact finite-rank covariance)**.** *For all indices $i, k \in [m]$ and $j, \ell \in [n]$,*

$$\mathrm{Cov}(X_{ij}, X_{k\ell}) = \|G\|_F^2\delta_{ik}\delta_{j\ell} + G_{ij}G_{k\ell} + \frac{2}{r}\left[\delta_{ik}(G^\top G)_{j\ell} + \delta_{j\ell}(GG^\top)_{ik} + G_{i\ell}G_{kj}\right]. \tag{70}$$

Table 3: Rank-one variance increase over dense Gaussian ES for square attention-output matrices at Qwen3 hidden widths.

Model	Width w	Relative increase
Qwen3-0.6B	1024	0.391%
Qwen3-1.7B	2048	0.195%
Qwen3-4B	2560	0.156%
Qwen3-8B	4096	0.098%
Qwen3-14B	5120	0.078%

Define the variance factor

$$\kappa_r := mn + 1 + \frac{2(m+n+1)}{r}. \tag{71}$$

Then

$$\mathbb{E}[\|X - G\|_F^2] = \kappa_r \|G\|_F^2. \tag{72}$$

For independent copies $X_1, \dots, X_N$ and their average $\widehat{G}_N = \frac{1}{N}\sum_{s=1}^N X_s$,

$$\mathbb{E}[\|\widehat{G}_N - G\|_F^2] = \frac{1}{N}\mathbb{E}[\|X - G\|_F^2] = \frac{\kappa_r}{N}\|G\|_F^2. \tag{73}$$

Proof. Substitute $E_r = r^{-1/2}\sum_{q=1}^r a_q b_q^\top$ into $\mathbb{E}[X_{ij}X_{k\ell}]$ and apply the Gaussian fourth-moment identity to the independent factors a_q and b_q. Summing the resulting fourth-moment terms against the entries of G gives

$$\mathbb{E}[X_{ij}X_{k\ell}] = \|G\|_F^2\delta_{ik}\delta_{j\ell} + 2G_{ij}G_{k\ell} + \frac{2}{r}\left[\delta_{ik}(G^\top G)_{j\ell} + \delta_{j\ell}(GG^\top)_{ik} + G_{i\ell}G_{kj}\right]. \tag{74}$$

Since $\mathbb{E}[X] = G$, subtracting $G_{ij}G_{k\ell}$ proves Equation (70). Taking $i = k$, $j = \ell$, and summing yields

$$\operatorname{tr}\operatorname{Cov}(X) = mn\|G\|_F^2 + \|G\|_F^2 + \frac{2}{r}\left[m\|G\|_F^2 + n\|G\|_F^2 + \|G\|_F^2\right], \tag{75}$$

which is Equation (72). Independence gives the factor $1/N$ in Equation (20). □

Dense Gaussian directions and Gaussian-product directions share the variance factor $mn + 1$. Finite rank adds only the product-law correction $2(m+n+1)/r$. The relative variance increase of finite-rank EGGROLL over dense Gaussian ES in the affine model is therefore

$$\rho_{m,n,r} = \frac{2(m+n+1)}{r(mn+1)}. \tag{76}$$

For a square matrix of width w, the rank-one relative increase is $2(2w+1)/(w^2+1) = O(1/w)$. Table 3 evaluates this quantity at the hidden widths of the Qwen3 models used in this work (Yang et al., 2025). These widths describe, for example, their square attention-output matrices. The 16×16 matrices in the transformer audit are extracted sub-blocks chosen to make the finite-rank surcharge measurable, not complete model weight matrices.

A conclusion to draw is that a rank-one law can be singular with respect to a dense Gaussian law while having nearly the same affine-model variance. Singularity describes the support of the perturbation law, whereas estimator variance depends on a particular combination of its fourth moments. The rank-dependent contribution grows as $m + n$, while the shared dense-Gaussian contribution grows as mn.

For a network, independence confines the additional finite-rank variance to within-block terms. Appendix E gives the exact formula and shows how each block is weighted by its local gradient energy.

B.7 FIXED-RADIUS RANK EXPANSION

This appendix proves the fixed-radius expansion stated in Theorem 4.2. We hold σ fixed and expand the finite-rank mean field around its dense-Gaussian limit as $r \to \infty$, identifying both the leading $1/r$ correction and the $O(r^{-2})$ remainder.

For a fixed frequency T, expanding the exact characteristic function as $r \to \infty$ using the matrix-logarithm series (Higham, 2008) gives

$$\log \Phi_r(T) = -\frac{1}{2}\|T\|_F^2 + \frac{1}{4r}\operatorname{tr}\left((TT^\top)^2\right) - \frac{1}{6r^2}\operatorname{tr}\left((TT^\top)^3\right) + O(r^{-3}). \tag{77}$$

The leading term is the logarithm of the characteristic function of a dense Gaussian and depends only on $\|T\|_F$. The first finite-rank correction also depends on how this norm is distributed across singular directions through $\operatorname{tr}((TT^\top)^2)$. Thus, the same anisotropy that produced the resolvent in Section 2 appears as the leading $1/r$ error at fixed radius.

The frequency-wise expansion must be strengthened before it can be integrated against a general objective. We therefore expand the score-weighted perturbation law itself and control its remainder in a polynomially weighted integral norm. This yields a uniform field bound for objectives of polynomial growth without requiring a Fourier density for f. For a matrix displacement Z, define the matrix-valued measure

$$\nu_{r,\sigma}(dZ) := \frac{Z}{\sigma^2}\,\mathbb{P}(\sigma E_r \in dZ). \tag{78}$$

The population field can then be written as

$$g^f_{r,\sigma}(W) = \int f(W+Z)\,\nu_{r,\sigma}(dZ). \tag{79}$$

Thus, an expansion of $\nu_{r,\sigma}$ immediately becomes an expansion of the update field. For these kernels, write $\widehat{\nu}(T) = \int e^{i\langle T,\,Z\rangle_F}\nu(dZ)$, with a positive sign because the field evaluates $f(W+Z)$. The transform of the update measure is then the multiplier derived in Section 2:

$$\widehat{\nu}_{r,\sigma}(T) = i\Phi_r(\sigma T)\left(I_m + \frac{\sigma^2}{r}TT^\top\right)^{-1} T. \tag{80}$$

Recall from Section 2 that $\Phi_r(\sigma T) = \det(I_m + \sigma^2 TT^\top/r)^{-r/2}$. We expand both rank-dependent factors in powers of $1/r$:

$$\begin{aligned}\Phi_r(\sigma T) &= e^{-\sigma^2\|T\|_F^2/2}\left[1 + \frac{\sigma^4}{4r}\operatorname{tr}\left((TT^\top)^2\right) + O(r^{-2})\right],\\ \left(I_m + \frac{\sigma^2}{r}TT^\top\right)^{-1} &= I_m - \frac{\sigma^2}{r}TT^\top + O(r^{-2}).\end{aligned} \tag{81}$$

Consequently,

$$\widehat{\nu}_{r,\sigma}(T) = ie^{-\sigma^2\|T\|_F^2/2}T + \frac{1}{r}\widehat{\beta}_\sigma(T) + O(r^{-2}), \tag{82}$$

where the matrix-valued correction kernel β_σ is defined by

$$\widehat{\beta}_\sigma(T) = ie^{-\sigma^2\|T\|_F^2/2}\left[\frac{\sigma^4}{4}\operatorname{tr}\left((TT^\top)^2\right)T - \sigma^2 TT^\top T\right], \tag{83}$$

and its operation on the objective is denoted by

$$\mathcal{B}_\sigma f(W) := \int f(W+Z)\,\beta_\sigma(dZ). \tag{84}$$

The two terms in Equation (83) have distinct origins. The term containing $\operatorname{tr}((TT^\top)^2)T$ records the difference between finite-rank and dense-Gaussian smoothing, while the term containing $TT^\top T$ records the score mismatch identified in Section 2. The main text states the resulting uniform rank expansion. The constants may depend on m, n, p, σ, and C_f.

B.8 Proof of Theorem 4.2

Proof of Theorem 4.2. We expand the update kernel in the reciprocal rank $t = 1/r$ and bound the remainder before integrating against f. First set $\sigma = 1$ and write $D = mn$. For $t > 0$, define

$$\phi_t(T) = \det(I_m + tTT^\top)^{-1/(2t)}, \qquad A_t(T) = (I_m + tTT^\top)^{-1}, \qquad m_t(T) = i\phi_t(T)A_t(T)T. \tag{85}$$

At $t = 0$, set $\phi_0(T) = e^{-\|T\|_F^2/2}$ and $A_0(T) = I_m$. Then $m_{1/r} = \widehat{\nu}_{r,1}$, $m_0 = \widehat{\nu}_{\infty,1}$, and $\partial_t m_0 = \widehat{\beta}_1$. These functions are smooth at $t = 0$ because

$$\ell_t(T) := \log \phi_t(T) = -\frac{1}{2}\int_0^1 \operatorname{tr}\left[TT^\top(I_m + utTT^\top)^{-1}\right] du. \tag{86}$$

In particular, $\partial_t \ell_0 = \operatorname{tr}((TT^\top)^2)/4$ and $\partial_t A_0 = -TT^\top$, recovering the two terms in Equation (83).

To control the remainder, let α be a multi-index for the D entries of T, so that ∂_T^α has total derivative order $|\alpha|$. For every fixed order k, there are constants C_k and an integer K_k, independent of $0 \le t \le 1$, such that

$$\|\partial_T^\alpha \partial_t^2 m_t(T)\|_F \le C_k(1 + \|T\|_F)^{K_k}\phi_t(T), \qquad |\alpha| \le k. \tag{87}$$

To verify this bound, differentiate Equation (86) under the integral. Every derivative of an inverse matrix is obtained from $\partial B^{-1} = -B^{-1}(\partial B)B^{-1}$. All inverse factors here have operator norm at most one, while derivatives of $TT^\top$ are polynomial in T. Thus every fixed mixed derivative of ℓ_t and A_t has a polynomial bound, uniformly for $0 \le t \le 1$. Repeated differentiation of e^{ℓ_t} leaves the factor ϕ_t multiplied by products of these bounded derivatives. Applying the product rule to $m_t = i\phi_t A_t T$ proves Equation (87).

The determinant supplies a common decay bound. For $0 < t \le t_0 \le 1$,

$$\phi_t(T) \le (1 + t\|T\|_F^2)^{-1/(2t)} \le (1 + t_0\|T\|_F^2)^{-1/(2t_0)}. \tag{88}$$

The first inequality follows by expanding the product of $1 + ts_j^2$, where s_j are the singular values of T. The second follows because $\log(1 + ta)/t$ decreases with t for $a \ge 0$. The bound also holds at $t = 0$ by continuity. Choose t_0 so that $1/t_0 > K_k + D/2$. The right-hand side of Equation (87), with Equation (88), is then square-integrable in T, uniformly over $0 \le t \le t_0$.

Taylor's formula with an integral remainder therefore gives

$$\partial_T^\alpha(m_t - m_0 - t\partial_t m_0) = t^2\int_0^1 (1-v)\partial_T^\alpha \partial_t^2 m_{vt}\, dv, \qquad \|\partial_T^\alpha(m_t - m_0 - t\partial_t m_0)\|_{L^2} \le C_k' t^2. \tag{89}$$

For general fixed $\sigma > 0$, the update multiplier is $\sigma^{-1}m_t(\sigma T)$. Rescaling T and taking $t = 1/r$ therefore proves, for $r \ge r_0 := \lceil 1/t_0 \rceil$,

$$\left\| \partial_T^\alpha \left(\widehat{\nu}_{r,\sigma} - \widehat{\nu}_{\infty,\sigma} - \frac{1}{r}\widehat{\beta}_\sigma \right) \right\|_{L^2} \le \frac{C_{\sigma,k}}{r^2}, \qquad |\alpha| \le k. \tag{90}$$

Finally choose an integer $q > D/2$ and take $k = p + q$. Let $\rho_{r,\sigma}$ be the inverse transform of the remainder, using the positive-sign kernel convention of Equation (80). Cauchy–Schwarz and Plancherel's identity (Folland, 1999) imply

$$\int (1 + \|Z\|_F)^p \|\rho_{r,\sigma}(Z)\|_F\, dZ \le C_{D,p,q} \sum_{|\alpha| \le p+q} \|\partial_T^\alpha \widehat{\rho}_{r,\sigma}\|_{L^2} \le \frac{C_{\sigma,p}}{r^2}. \tag{91}$$

Indeed, apply Cauchy–Schwarz with the square-integrable weight $(1 + \|Z\|_F)^{-q}$, then bound the weighted L^2 norm by the finitely many monomials $Z^\alpha \rho_{r,\sigma}$ of degree at most $p + q$. Plancherel converts those monomials into Fourier derivatives. The correction kernel β_σ has a Schwartz density because its transform is a polynomial times a Gaussian. Uniqueness of Fourier transforms of finite measures identifies $\rho_{r,\sigma}(Z)\, dZ$ with $\nu_{r,\sigma} - \nu_{\infty,\sigma} - \beta_\sigma/r$.

Since $|f(W + Z)| \le 2C_f(1 + \|W\|_F)^p(1 + \|Z\|_F)^p$, integration against this remainder gives

$$\|g_{r,\sigma}^f(W) - g_{\infty,\sigma}^f(W) - \frac{1}{r}\mathcal{B}_\sigma f(W)\|_F \le \frac{2C_f C_{\sigma,p}}{r^2}(1 + \|W\|_F)^p, \tag{92}$$

which proves Equation (21). Multiplying the expansion by r shows that $r(g_{r,\sigma}^f - g_{\infty,\sigma}^f)$ converges to $\mathcal{B}_\sigma f$ in $\|\cdot\|_{\infty,p}$. If this limit is nonzero, the difference cannot decay faster than $1/r$. □

The expansion identifies the dense-Gaussian limit and the leading finite-rank error $\mathcal{B}_\sigma f/r$. The two terms in Equation (83) respectively arise from the change in smoothing law and the score mismatch, while the remaining error is $O(r^{-2})$.

Since $g^f_{\infty,\sigma} = \nabla F_{\infty,\sigma}$, both fields in this comparison use the same fixed perturbation radius. Their difference therefore isolates the effect of finite rank rather than the smoothing shared with dense Gaussian ES. In particular,

$$\|g^f_{r,\sigma} - \nabla F_{\infty,\sigma}\|_{\infty,p} \le \frac{K_{\sigma,p}}{r} + \frac{R_{\sigma,p}}{r^2}, \qquad K_{\sigma,p} := \|\mathcal{B}_\sigma f\|_{\infty,p}. \tag{93}$$

For bounded f, this is an ordinary uniform bound. For polynomially growing f, the denominator in $\|\cdot\|_{\infty,p}$ allows the field error to grow at the same polynomial rate with W. When $\mathcal{B}_\sigma f$ is nonzero, the leading term does not vanish, so the $1/r$ rate cannot be improved in this norm. This strengthens the $O(1/r)$ comparison of Sarkar et al. (2026): their Edgeworth argument establishes the rate, whereas the kernel expansion identifies its coefficient and leaves an $O(r^{-2})$ residual. Appendix E.2 combines the field-error bounds with the finite-population variance from Section 4 to derive finite-iteration guarantees for approaching stationary points of the original objective, and, at fixed perturbation radius, of the dense-Gaussian-smoothed objective.

C Proofs for LOO-ROLL

C.1 Proof of Proposition 5.1

Proof of Proposition 5.1. Conditional on $\mathcal{F}_t$, the term $E_{t,s}F_{t,s}/\sigma$ has expectation $g^f_{r,\sigma}(W_t)$. The baseline $\overline{F}_{t,-s}$ depends only on the other directions and is therefore conditionally independent of $E_{t,s}$. Centering gives $\mathbb{E}[E_{t,s}\overline{F}_{t,-s} \mid \mathcal{F}_t] = 0$. Averaging over s proves the claim. □

C.2 Proof of Proposition 5.2

Proof of Proposition 5.2. In the affine model, the constant fitness cancels from every centered score. Write $z_i = \mathrm{vec}(E_i)$, $v = \mathrm{vec}(G)$, and $y_i = z_i^\top v$. The vectors z_i are independent, centered, and have identity covariance. With $\bar{z} = N^{-1}\sum_i z_i$ and $\bar{y} = N^{-1}\sum_i y_i$, the vectorized estimator is the sample covariance applied to v:

$$\mathrm{vec}(\widehat{g}^{\mathrm{LOO}}(W;N)) = \frac{1}{N-1}\sum_{i=1}^{N}(z_i - \bar{z})(y_i - \bar{y}). \tag{94}$$

Expanding the two means gives the following centered decomposition:

$$\mathrm{vec}(\widehat{g}^{\mathrm{LOO}}(W;N)) - v = \frac{1}{N}\sum_i (z_i y_i - v) - \frac{1}{N(N-1)}\sum_{i<j}(z_i y_j + z_j y_i). \tag{95}$$

Every cross-pair summand has conditional mean zero given either of its two directions. Consequently, these summands are uncorrelated with the first sum and with one another, including when two pairs share an index. By Theorem 4.1, $\mathbb{E}[\|z_i y_i - v\|^2] = \kappa_r\|v\|^2$. Independence and identity covariance also give

$$\mathbb{E}[\|z_i y_j + z_j y_i\|^2] = 2\mathbb{E}[\|z_i\|^2]\mathbb{E}[y_j^2] + 2\|\mathbb{E}[z_i y_i]\|^2 = 2(d+1)\|v\|^2, \qquad i \ne j. \tag{96}$$

There are $N(N-1)/2$ unordered pairs. Adding their variances therefore proves Equation (24). Substituting $2N$ for the LOO-ROLL population and dividing by the antithetic variance $\kappa_r\|G\|_F^2/N$ proves Equation (25). □

D Additional LLM comparisons

The fixed-corpus rank comparisons are reported in Tables 4 and 5, while Table 6 reports the next-token prediction results.

Table 4: Fixed-validation changes in held-out task reward from rank one to rank eight, in percentage points, over five matched seeds. Positive values favor rank eight and negative values favor rank one. The evaluation set contains 256 examples and is shared across ranks and budgets. "Matched evaluations" uses the same ES iterations and model evaluations. "Matched wall time" extends rank-one training to the measured rank-eight runtime. Brackets give paired 95% Student-t intervals.

Model	Task	Matched evaluations	Matched wall time
Qwen3-0.6B	Countdown	$-.60\%$ $[-3.28, 2.07]$	-1.57% $[-3.29, .15]$
Qwen3-0.6B	GSM8K	$+2.50\%$ $[-.82, 5.82]$	-5.70% $[-14.57, 3.16]$
Qwen3-1.7B	Countdown	$+1.50\%$ $[-.29, 3.30]$	$+1.88\%$ $[-.25, 4.01]$
Qwen3-1.7B	GSM8K	$+.94\%$ $[-.76, 2.64]$	-1.33% $[-2.88, .23]$

Table 5: Percentage change in next-token prediction loss from rank one to rank eight over five paired seeds and 256 fixed held-out sequences. Negative values indicate lower loss at rank eight. Brackets give paired 95% Student-t intervals. Bold entries exclude zero.

Model	Matched evaluations	Matched wall time
Qwen3-0.6B	$\mathbf{-.58\%}$ $\mathbf{[-.86, -.30]}$	$-.11\%$ $[-.29, .07]$
Qwen3-1.7B	$\mathbf{-2.23\%}$ $\mathbf{[-2.53, -1.92]}$	$\mathbf{-.93\%}$ $\mathbf{[-1.34, -.53]}$
Qwen3-4B	$\mathbf{-1.76\%}$ $\mathbf{[-1.95, -1.57]}$	$\mathbf{+1.92\%}$ $\mathbf{[1.48, 2.35]}$
Qwen3-8B	$\mathbf{-2.18\%}$ $\mathbf{[-2.60, -1.77]}$	$\mathbf{+2.14\%}$ $\mathbf{[1.95, 2.33]}$

Table 6: Percentage change in fixed-validation negative log likelihood from antithetic EGGROLL to LOO-ROLL. Values are mean $\pm$ sample standard deviation over five paired seeds, computed as $100(\text{LOO-ROLL} - \text{EGGROLL})/\text{EGGROLL}$. Negative values indicate lower loss and favor LOO-ROLL, while positive values favor EGGROLL. Bold marks mean improvements for LOO-ROLL.

Model	Equal iterations	Equal evaluations	Equal wall time
Qwen3-0.6B	$+0.20\% \pm 0.44$	$+0.11\% \pm 0.30$	$\mathbf{-0.55\% \pm 0.46}$
Qwen3-1.7B	$+0.21\% \pm 0.58$	$\mathbf{-0.10\% \pm 0.34}$	$\mathbf{-1.20\% \pm 0.62}$
Qwen3-8B	$+0.53\% \pm 0.57$	$+0.31\% \pm 0.49$	$\mathbf{-2.16\% \pm 0.81}$

E NETWORK EXTENSIONS AND FURTHER THEORETICAL RESULTS

The main results are stated for one matrix so that the finite-rank mechanism remains visible. A network objective couples many parameter matrices, but EGGROLL samples their perturbations independently. This section gives the corresponding joint resolvent, field-error, and variance statements.

Corollary E.1 (Network resolvent)**.** *Let $W = (W_1, \dots, W_Q)$, where $W_q \in \mathbb{R}^{m_q \times n_q}$, and let the blocks $E_{r_q}^{(q)}$ be independent Gaussian-product perturbations. Define*

$$\begin{aligned} F_{\boldsymbol{r},\sigma}(W) &= \mathbb{E}\big[f(W_1 + \sigma E_{r_1}^{(1)}, \dots, W_Q + \sigma E_{r_Q}^{(Q)})\big], \\ g_q(W) &= \frac{1}{\sigma}\mathbb{E}\left[E_{r_q}^{(q)} f(W_1 + \sigma E_{r_1}^{(1)}, \dots, W_Q + \sigma E_{r_Q}^{(Q)})\right]. \end{aligned} \tag{97}$$

Let $\mathcal{L}_q$ be the operator in Equation (12), acting on the qth vector-field component and differentiating with respect to W_q. If f admits the analogous Fourier representation and moment condition on the product parameter space, then

$$\left(I + \frac{\sigma^2}{r_q}\mathcal{L}_q\right) g_q = \nabla_{W_q} F_{\boldsymbol{r},\sigma} \quad \text{for every } q. \tag{98}$$

Equivalently, $g = J_{\boldsymbol{r},\sigma} \nabla F_{\boldsymbol{r},\sigma}$ for $J_{\boldsymbol{r},\sigma} = \mathrm{diag}_q(I + \sigma^2 \mathcal{L}_q / r_q)^{-1}$. The operator $J_{\boldsymbol{r},\sigma}$ is self-adjoint, positive, contractive, and firmly nonexpansive on the direct-sum L^2 space.

Proof. Independence gives the joint characteristic function

$$\Phi_{\boldsymbol{r}}(T_1, \dots, T_Q) = \prod_{q=1}^{Q} \Phi_{r_q}(T_q). \tag{99}$$

Substitution of the Fourier representation of f, followed by differentiation of the qth factor and application of Equation (8), gives the multiplier

$$i\Phi_{\boldsymbol{r}}(\sigma \boldsymbol{T}) \left(I_{m_q} + \frac{\sigma^2 T_q T_q^\top}{r_q}\right)^{-1} T_q. \tag{100}$$

The multiplier of $\nabla_{W_q} F_{\boldsymbol{r},\sigma}$ is $iT_q \Phi_{\boldsymbol{r}}(\sigma \boldsymbol{T})$. The blockwise identity follows, and the remaining properties follow from Theorem 2.2 on each diagonal block. □

For the smoothness bounds, write $d_q = m_q n_q$, $D = \sum_q d_q$, and $\mathbf{E} = (E_{r_1}^{(1)}, \dots, E_{r_Q}^{(Q)})$. Equip this product space with the direct-sum Frobenius norm $\|\mathbf{E}\|^2 = \sum_q \|E_{r_q}^{(q)}\|_F^2$. Independence across blocks gives

$$\mu_4(\boldsymbol{r}) := \mathbb{E}[\|\mathbf{E}\|^4] = D(D+2) + \sum_{q=1}^{Q} \frac{2d_q(m_q + n_q + 1)}{r_q}. \tag{101}$$

Corollary E.2 (Network field error under standard smoothness)**.** *Let h be a loss on the product of the block spaces, and define $g_{\boldsymbol{r},\sigma}^h$ using the joint perturbation $\mathbf{E}$. If h is L-smooth in the direct-sum norm, then*

$$\|g_{\boldsymbol{r},\sigma}^h(\mathbf{W}) - \nabla h(\mathbf{W})\| \le \frac{L\sigma}{2} \mu_4(\boldsymbol{r})^{3/4}. \tag{102}$$

If the Hessian is M-Lipschitz in the same norm, then

$$\|g_{\boldsymbol{r},\sigma}^h(\mathbf{W}) - \nabla h(\mathbf{W})\| \le \frac{M\sigma^2}{6} \mu_4(\boldsymbol{r}). \tag{103}$$

Proof. The smoothness inequalities in Lemma B.2 depend only on the norm of the full perturbation. Identity covariance supplies the first-order term, and Equation (101) supplies the required fourth moment. □

Corollary E.3 (Network-level variance). *Let block $q = 1, \dots, Q$ have shape $m_q \times n_q$, rank r_q, independent perturbation, and gradient G_q. Write $D = \sum_q m_q n_q$ and $\|G\|^2 = \sum_q \|G_q\|_F^2$. The concatenated single-direction estimator satisfies*

$$\mathbb{E}[\|X - G\|^2] = (D+1)\|G\|^2 + \sum_{q=1}^{Q} \frac{2(m_q + n_q + 1)}{r_q} \|G_q\|_F^2. \tag{104}$$

Proof. Concatenate the matrix blocks into one vector. The Gaussian part of the fourth moment produces $(D+1)\|G\|^2$. Independence makes the cross-block fourth cumulants vanish, and Theorem 4.1 gives the remaining within-block terms. □

The additional variance from block q is weighted by $\|G_q\|_F^2$. Its contribution therefore depends on both the dimensions of the block and the amount of gradient energy concentrated there. The same formula supplies a network-level relative-noise constant for the convergence results in Appendix E.2. One valid choice is

$$C^2 = D + 1 + \max_q \frac{2(m_q + n_q + 1)}{r_q}. \tag{105}$$

E.1 Secondary corrections suggested by the analysis

The fixed-radius expansion and variance calculation suggest two further modifications. Although both are theoretically valid estimators, neither improved on ordinary EGGROLL in our model-scale studies. We record the results in this appendix rather than presenting them as viable practical contributions.

E.1.1 Rank extrapolation

The coefficient in Theorem 4.2 can be cancelled by combining two ranks. For $s > r$, define

$$g^{\mathrm{RE}}_{r,s,\sigma} := \frac{s g^f_{s,\sigma} - r g^f_{r,\sigma}}{s - r}. \tag{106}$$

The resulting bias can be bounded explicitly for any chosen rank ratio.

Corollary E.4 (Rank extrapolation at a prescribed tolerance). *Under the assumptions of Theorem 4.2, let $s = \alpha r$ for some $\alpha > 1$ such that s is an integer. For all sufficiently large r,*

$$\|g^{\mathrm{RE}}_{r,s,\sigma} - g^f_{\infty,\sigma}\|_{\infty,p} \le \frac{R_{\sigma,p}}{r^2} \frac{\alpha + 1}{\alpha(\alpha - 1)}. \tag{107}$$

If $\delta = \varepsilon r^2 / R_{\sigma,p}$, it is sufficient to choose

$$\alpha \ge \frac{\delta + 1 + \sqrt{(\delta+1)^2 + 4\delta}}{2\delta}. \tag{108}$$

Proof. Write $g^f_{k,\sigma} = g^f_{\infty,\sigma} + \mathcal{B}_\sigma f / k + R_k$ with $\|R_k\|_{\infty,p} \le R_{\sigma,p}/k^2$. Substitution into Equation (106) cancels $\mathcal{B}_\sigma f$ and gives Equation (107). Solving $(\alpha+1)/(\alpha(\alpha-1)) \le \delta$ gives Equation (108). □

Cancellation of the mean-field term comes at the cost of estimating two fields, and the extrapolation weights amplify their sampling errors. If the two estimates are independent and use N_r and N_s directions with single-direction variances V_r and V_s, then

$$\mathbb{E}\left[\|\hat{g}^{\mathrm{RE}}_{r,s,\sigma} - g^{\mathrm{RE}}_{r,s,\sigma}\|_F^2\right] = \frac{s^2 V_s / N_s + r^2 V_r / N_r}{(s-r)^2}. \tag{109}$$

At a fixed budget of 256 fitness evaluations, transformer-block experiments compare ordinary rank-one EGGROLL with extrapolation using ranks $(1, 2)$ and $(1, 4)$. Across two blocks and three radii, the extrapolated estimator has respectively 7.63–8.67 and 2.35–2.62 times the MSE of ordinary EGGROLL. The leading bias is below the sampling resolution in these settings, so cancelling it does not repay the variance introduced by dividing the population between two ranks.

Table 7: SVRG improvement over EGGROLL over five paired seeds. For next-token prediction, entries are the reduction in held-out loss, computed as EGGROLL loss minus SVRG loss. For GSM8K, entries are SVRG reward minus EGGROLL reward in percentage points. Positive values favor SVRG and negative values favor EGGROLL. Brackets give paired 95% Student-t intervals, and bold entries have intervals that exclude zero.

Model	Objective	Difference
Qwen3-0.6B	Next-token prediction loss	$-.0161$ $[-.0488, .0166]$
Qwen3-0.6B	GSM8K	-6.17% $[-17.43, 5.08]$
Qwen3-1.7B	Next-token prediction loss	$\mathbf{-.0197}$ $\mathbf{[-.0321, -.0072]}$
Qwen3-1.7B	GSM8K	-1.25% $[-2.98, .48]$

E.1.2 Variance reduction with a reference point

The dominant term in Equation (72) is shared with dense Gaussian ES. Following zeroth-order variance reduction for language-model fine-tuning (Gautam et al., 2024), let ξ denote the sampled data and define

$$Y_{r,\sigma}(W;\xi,E) := E\frac{\ell(W+\sigma E;\xi)-\ell(W-\sigma E;\xi)}{2\sigma}, \qquad g_{r,\sigma}(W) := \mathbb{E}\left[Y_{r,\sigma}(W;\xi,E)\right]. \tag{110}$$

At a reference point $\widetilde{W}$, let $\overline{g}_B(\widetilde{W})$ be an independent unbiased average of B updates and set

$$Y^{\mathrm{VR}}(W,\widetilde{W}) := Y_{r,\sigma}(W;\xi,E) - Y_{r,\sigma}(\widetilde{W};\xi,E) + \overline{g}_B(\widetilde{W}). \tag{111}$$

The corresponding control-variate guarantee for Gaussian-product directions is as follows.

Proposition E.5 (Low-rank variance reduction)**.** *The estimator in Equation (111) is unbiased for $g_{r,\sigma}(W)$. If*

$$\mathbb{E}\left[\|Y_{r,\sigma}(W;\xi,E) - Y_{r,\sigma}(V;\xi,E)\|_F^2\right] \le L_Y^2\|W-V\|_F^2 \tag{112}$$

and $\mathbb{E}[\|\overline{g}_B(\widetilde{W}) - g_{r,\sigma}(\widetilde{W})\|_F^2] \le V_{\mathrm{ref}}/B$, then

$$\mathbb{E}\left[\|Y^{\mathrm{VR}}(W,\widetilde{W}) - g_{r,\sigma}(W)\|_F^2\right] \le L_Y^2\|W-\widetilde{W}\|_F^2 + \frac{V_{\mathrm{ref}}}{B}. \tag{113}$$

Proof. Taking expectations in Equation (111) cancels the field at the reference point and leaves $g_{r,\sigma}(W)$. Independence of the reference estimate makes the centered errors orthogonal, and the two assumed bounds give Equation (113). □

We tested this estimator on Qwen3-0.6B and Qwen3-1.7B with five paired seeds using next-token prediction loss and GSM8K rewards. Table 7 reports held-out reward relative to matched EGGROLL checkpoints. All four means favor ordinary EGGROLL, with the 1.7B next-token prediction loss difference excluding zero. These results do not establish a practical benefit from maintaining the reference estimate and its additional evaluations.

E.2 Convergence consequences of the field bounds

This appendix translates the mean-field error and finite-population variance bounds into standard stationarity guarantees under smoothness.

Lemma E.6 (Descent with accumulated field error)**.** *Suppose h is L-smooth and bounded below by $h_\star$. Let $g = \nabla h + b$ and $x_{t+1} = x_t - \eta g(x_t)$. If $0 < \eta \le 1/L$, then*

$$\frac{1}{T}\sum_{t=0}^{T-1}\|\nabla h(x_t)\|^2 \le \frac{2(h(x_0)-h_\star)}{\eta T} + \frac{1}{T}\sum_{t=0}^{T-1}\|b(x_t)\|^2. \tag{114}$$

Proof. Write $a_t = \nabla h(x_t)$ and $b_t = b(x_t)$. Smoothness and the identity $-\langle a_t, a_t + b_t\rangle = -\|a_t\|^2/2 - \|a_t+b_t\|^2/2 + \|b_t\|^2/2$ give

$$h(x_{t+1}) \le h(x_t) - \frac{\eta}{2}\|a_t\|^2 - \frac{\eta}{2}(1-L\eta)\|a_t+b_t\|^2 + \frac{\eta}{2}\|b_t\|^2. \tag{115}$$

Summation proves the claim. □

Define

$$B^{(L)}_{r,\sigma} := \frac{L\sigma}{2}\mu_4(r)^{3/4}, \qquad B^{(M)}_{r,\sigma} := \frac{M\sigma^2}{6}\mu_4(r). \tag{116}$$

Combining Lemma E.6 with Lemma B.2 gives the following direct consequence.

Theorem E.7 (Mean-field convergence under standard smoothness)**.** *Suppose h is L-smooth and bounded below by $h_\star$. Fix $\varepsilon > 0$, use $\eta = 1/L$, and set $H_0 = h(x_0) - h_\star$. Either of the sufficient choices*

$$0 < \sigma \le \sigma_L(r,\varepsilon) := \frac{\sqrt{2}\,\varepsilon}{L\mu_4(r)^{3/4}}, \tag{117}$$

or, when the Hessian is M-Lipschitz,

$$0 < \sigma \le \sigma_M(r,\varepsilon) := \left(\frac{3\sqrt{2}\,\varepsilon}{M\mu_4(r)}\right)^{1/2} \tag{118}$$

ensures $T^{-1}\sum_{t=0}^{T-1}\|\nabla h(x_t)\|^2 \le \varepsilon^2$ for $T \ge \lceil 4LH_0/\varepsilon^2\rceil$.

Proof. The applicable radius makes the corresponding field-error bound at most $\varepsilon/\sqrt{2}$. Substitution into Equation (114) completes the proof. □

E.2.1 Finite-population convergence

A finite-population update contains sampling noise around the mean field and systematic error relative to a chosen reference gradient. The next result separates these contributions, allowing either the original objective or dense Gaussian ES at the same radius to serve as the reference. As in Appendix B.5.3, we minimize a loss h through $x_{t+1} = x_t - \eta\widehat{g}_t$.

Theorem E.8 (Nonconvex convergence under relative population noise)**.** *Suppose h is L-smooth and bounded below by $h_\star$. Let $(\mathcal{F}_t)_{t\ge0}$ be the filtration generated by x_0 and all perturbation and evaluation randomness revealed before iteration t, so that x_t is $\mathcal{F}_t$-measurable. At every iterate, assume*

$$\mathbb{E}[\widehat{g}_t \mid \mathcal{F}_t] = g(x_t), \tag{119}$$

$$\mathbb{E}[\|\widehat{g}_t - g(x_t)\|^2 \mid \mathcal{F}_t] \le \frac{G^2 + C^2\|g(x_t)\|^2}{N} \tag{120}$$

for $C, G \ge 0$. If $0 < \eta \le [L(1 + C^2/N)]^{-1}$, then

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla h(x_t)\|^2] \le \frac{2(h(x_0) - h_\star)}{\eta T} + \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|g(x_t) - \nabla h(x_t)\|^2] + \frac{L\eta G^2}{N}. \tag{121}$$

Proof. Smoothness and conditional expectation give

$$\mathbb{E}_t[h(x_{t+1})] \le h(x_t) - \eta\langle\nabla h(x_t),\, g(x_t)\rangle + \frac{L\eta^2}{2}\left[\left(1 + \frac{C^2}{N}\right)\|g(x_t)\|^2 + \frac{G^2}{N}\right]. \tag{122}$$

Write $g = \nabla h + b$ and use

$$-\langle\nabla h,\, g\rangle = -\frac{1}{2}\|\nabla h\|^2 - \frac{1}{2}\|g\|^2 + \frac{1}{2}\|b\|^2. \tag{123}$$

The step-size condition makes the total coefficient of $\|g\|^2$ nonpositive. Taking full expectation and summing over t gives Equation (121). □

We first measure stationarity with respect to the original loss h. In this comparison, the field error includes both smoothing and finite-rank score mismatch, and the standard smoothness bounds from Section 3 control them together.

Corollary E.9 (Convergence relative to the original loss)**.** *Suppose that h is L-smooth and bounded below by $h_\star$, that the finite-population estimator satisfies Equations (119) and (120) with $g = g^h_{r,\sigma}$, and set $0 < \eta \le [L(1 + C^2/N)]^{-1}$. Define $B^{\mathrm{std}}_{r,\sigma} = B^{(L)}_{r,\sigma}$. If the Hessian is M-Lipschitz, one may instead take $B^{\mathrm{std}}_{r,\sigma} = B^{(M)}_{r,\sigma}$. Then*

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla h(x_t)\|^2] \le \frac{2(h(x_0) - h_\star)}{\eta T} + \left(B^{\mathrm{std}}_{r,\sigma}\right)^2 + \frac{L\eta G^2}{N}. \tag{124}$$

If $\sigma \le \sigma_L(r, \varepsilon)$, or $\sigma \le \sigma_M(r, \varepsilon)$ under the Lipschitz-Hessian assumption, then the middle term is at most $\varepsilon^2/2$.

Proof. Apply Lemma B.2 to the field-error term in Equation (121). □

Dividing the target tolerance among mean-field error, the finite optimization horizon, and irreducible population noise gives an explicit iteration complexity for a single matrix. Replacing $\mu_4(r)$ by $\mu_4(\boldsymbol{r})$ from Equation (101) gives the corresponding network statement.

Corollary E.10 (Explicit finite-population complexity)**.** *Under the assumptions of Corollary E.9, fix $\varepsilon > 0$ and let $\Delta = h(x_0) - h_\star$. Set $\eta = 1/(2L)$ and choose*

$$0 < \sigma \le \frac{2\varepsilon}{\sqrt{3}\, L\mu_4(r)^{3/4}}. \tag{125}$$

If the Hessian is M-Lipschitz, the alternative choice

$$0 < \sigma \le \left(\frac{2\sqrt{3}\,\varepsilon}{M\mu_4(r)}\right)^{1/2} \tag{126}$$

is sufficient. If

$$N \ge \max\left\{C^2, \frac{3G^2}{2\varepsilon^2}\right\} \tag{127}$$

and

$$T \ge \left\lceil \frac{12L\Delta}{\varepsilon^2} \right\rceil, \tag{128}$$

then

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla h(x_t)\|^2\right] \le \varepsilon^2. \tag{129}$$

Proof. The radius choices make the field-error bound at most $\varepsilon/\sqrt{3}$. The requirement $N \ge C^2$ gives $\eta \le [L(1 + C^2/N)]^{-1}$, while the second part of Equation (127) bounds the irreducible term by $G^2/(2N) \le \varepsilon^2/3$. Finally, Equation (128) bounds the optimization term by $4L\Delta/T \le \varepsilon^2/3$. Substitution into Equation (124) proves the claim. □

The guarantees above concern the original loss. To compare finite-rank EGGROLL directly with dense Gaussian ES at the same radius, let $H_{\infty,\sigma}(x) := \mathbb{E}[h(x + \sigma Z)]$ for $Z \sim \mathcal{N}(0, I)$. Using this smoothed loss as the reference removes the smoothing error shared by both methods and leaves the finite-rank discrepancy derived in Appendix B.7.

Corollary E.11 (Convergence relative to dense Gaussian ES)**.** *Suppose h satisfies the assumptions of Theorem 4.2, $H_{\infty,\sigma}$ is L_σ-smooth and bounded below by $H^\star_{\infty,\sigma}$, and the finite-population estimator satisfies Equations (119) and (120) with $g = g^h_{r,\sigma}$. For the constants $K_{\sigma,p}, R_{\sigma,p}, r_0$ in Equation (93), every $r \ge r_0$ and $0 < \eta \le [L_\sigma(1 + C^2/N)]^{-1}$ satisfy*

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla H_{\infty,\sigma}(x_t)\|^2\right] \le{}& \frac{2(H_{\infty,\sigma}(x_0) - H^\star_{\infty,\sigma})}{\eta T} + \frac{L_\sigma \eta G^2}{N} \\ &+ \left(\frac{K_{\sigma,p}}{r} + \frac{R_{\sigma,p}}{r^2}\right)^2 \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[(1 + \|x_t\|)^{2p}\right].\end{aligned} \tag{130}$$

If the displayed parameter moments are uniformly bounded along the trajectory, the mean-field contribution of finite rank to the squared-stationarity guarantee is $O(r^{-2})$ whenever the leading coefficient is nonzero. For bounded fitness, $p = 0$ and no trajectory-moment condition is needed.

Proof. Equation (93) gives

$$\|g^h_{r,\sigma}(x_t) - \nabla H_{\infty,\sigma}(x_t)\| \le \left(\frac{K_{\sigma,p}}{r} + \frac{R_{\sigma,p}}{r^2}\right)(1 + \|x_t\|)^p. \tag{131}$$

Square this bound and apply Theorem E.8 with $H_{\infty,\sigma}$ as the reference loss. □

The two reference objectives should not be conflated. Relative to h, the error contains smoothing and finite-rank effects and must be controlled through both σ and r. Relative to dense Gaussian ES at the same σ, the shared smoothing disappears and the mean-field difference is $O(1/r)$. The finite-population noise is the same in either comparison.

The constants in the variance assumption can be connected to quantities that are calculable or measurable for EGGROLL. The constant G^2 allows noise to remain when the mean update vanishes. The term $C^2\|g(x)\|^2$ is relative or multiplicative noise, also called a strong-growth component (Vaswani et al., 2019; Wang et al., 2023). For one matrix in the exact affine model, Theorem 4.1 gives $C^2 = mn + 1 + 2(m + n + 1)/r$. Appendix E gives the corresponding network expression. Relative noise restricts the stable step size, but it vanishes with the mean field and creates no additional error floor. The additive component G^2 produces the final term in Equation (121). Setting $C = 0$ recovers the usual bounded-variance assumption, while the exact affine EGGROLL model has $G = 0$.

E.3 Fourth moments of the Gaussian-product law

This appendix supplies the fourth-moment calculation used in Proposition 3.4. The coefficient in Equation (63) follows from the fourth moment of four entries of E_r. At rank one, $(E_1)_{ij} = a_i b_j$, where a and b are independent standard Gaussian vectors. For arbitrary row indices $i_1, \dots, i_4$ and column indices $j_1, \dots, j_4$,

$$\mathbb{E}\left[\prod_{s=1}^{4}(E_1)_{i_s j_s}\right] = \left(\delta_{i_1 i_2}\delta_{i_3 i_4} + \delta_{i_1 i_3}\delta_{i_2 i_4} + \delta_{i_1 i_4}\delta_{i_2 i_3}\right)\left(\delta_{j_1 j_2}\delta_{j_3 j_4} + \delta_{j_1 j_3}\delta_{j_2 j_4} + \delta_{j_1 j_4}\delta_{j_2 j_3}\right). \tag{132}$$

Each parenthesis contains the three possible pairings of four Gaussian factors. Expanding their product gives nine terms. In three of them, the row and column indices use the same pairing. For example, pairing the first entry with the second and the third with the fourth in both factors gives

$$\delta_{i_1 i_2}\delta_{j_1 j_2}\delta_{i_3 i_4}\delta_{j_3 j_4}. \tag{133}$$

These three terms are precisely the fourth moment of a dense Gaussian matrix with independent entries. The remaining six terms use different pairings for the row and column indices, such as

$$\delta_{i_1 i_2}\delta_{i_3 i_4}\delta_{j_1 j_3}\delta_{j_2 j_4}, \tag{134}$$

and arise from the rank-one factorization.

For general rank, write $E_r = r^{-1/2}\sum_{q=1}^{r} a_q b_q^\top$. A nonzero fourth-moment term either draws all four factors from one summand or draws two factors from each of two summands. The three dense-Gaussian terms receive both kinds of contributions, with total coefficient

$$\frac{r}{r^2} + \frac{r(r-1)}{r^2} = 1. \tag{135}$$

Each of the six additional terms requires all four factors to come from the same rank-one summand and therefore has coefficient $r/r^2 = 1/r$. Consequently, the fourth moment of E_r consists of the three dense-Gaussian terms plus the six additional terms scaled by $1/r$. To apply this identity in Equation (62), write the (i_0, j_0) entry as

$$\sum_{i_1, j_1, i_2, j_2, i_3, j_3} \partial_{i_1 j_1}\partial_{i_2 j_2}\partial_{i_3 j_3} f(W)\,\mathbb{E}\left[(E_r)_{i_0 j_0}(E_r)_{i_1 j_1}(E_r)_{i_2 j_2}(E_r)_{i_3 j_3}\right]. \tag{136}$$

Table 8: Training budgets for the fourteen LOO-ROLL comparisons. N counts independent EGGROLL directions, B is the training batch size, T_E is the EGGROLL update count, and T_L is the mean number of LOO-ROLL updates completed at equal time. Hours are mean EGGROLL runtimes over five seeds. NTP denotes next-token prediction.

Model	Task	N	B	T_E	T_L	Hours	Time ratio
Qwen3-0.6B	Countdown	128	8	150	374.4	2.668	0.9970
Qwen3-0.6B	GSM8K	128	4	150	368.4	4.145	0.9978
Qwen3-0.6B	NTP	32	4	80	131.8	0.151	1.0088
Qwen3-1.7B	Countdown	128	8	200	496.0	4.187	0.9980
Qwen3-1.7B	GSM8K	128	4	200	485.6	6.190	0.9985
Qwen3-1.7B	NTP	32	4	160	268.0	0.293	1.0199
Qwen3-1.7B	MATH	128	4	200	482.8	8.838	0.9985
Qwen3-8B	GSM8K	128	4	120	303.6	8.929	1.0009
Qwen3-8B	NTP	32	4	120	203.4	0.363	1.0271
Qwen3-8B	MATH	128	4	200	484.2	18.686	0.9998
Qwen3-14B	GSM8K	128	4	120	323.4	14.544	0.9675
Qwen3-14B	MATH	128	4	200	522.8	28.948	0.9998
SmolLM2-1.7B	Countdown	128	8	200	485.0	4.521	0.9981
SmolLM2-1.7B	GSM8K	128	4	200	444.8	3.250	0.9981

The three dense-Gaussian terms give the same quantity, $\partial_{i_0 j_0}\Delta f(W)$, after relabeling indices. The six additional terms give the same quantity, $(\mathcal{D}f(W))_{i_0 j_0}$, by symmetry of the third derivative. Summing them proves

$$\mathbb{E}\left[E_r D^3 f(W)[E_r, E_r, E_r]\right] = 3\nabla\Delta f(W) + \frac{6}{r}\mathcal{D}f(W). \tag{137}$$

F Complete empirical results

F.1 Training budgets

Table 8 gives the direction counts, batch sizes, update budgets, runtimes, and achieved time matching for all fourteen primary comparisons.

F.2 Paired statistical comparisons

The main text reports mean performance under matched evaluation and wall-time budgets. Here, we test the paired seed-level differences underlying those comparisons. Table 9 gives the individual two-sided tests at equal evaluation cost. Table 10 gives both the individual and Holm-adjusted p-values for the fourteen equal-wall-time comparisons, treating them as one family.

F.3 Rank comparisons and transformer audits

Tables 11 to 14 report the complete rank and radius comparisons summarized in Figure 4, including the stress-radius runs omitted from the main fixed-validation comparison. Rank columns give mean final training fitness $\pm$ standard error over three independent seeds, with higher values indicating better performance. The experiments use commit `bcc215e` of the eggroll-vLLM repository with the LOO-ROLL modifications described in Section 5.

The training results combine mean-field effects, population noise, and the resulting optimization trajectory. The transformer-block audit isolates the first two at a fixed model state. Table 15 compares each finite-rank estimator with dense Gaussian ES through its single-direction MSE and compares its estimated mean with the backpropagated gradient through cosine similarity. The rank-one MSE is 26–32% above the dense baseline and falls to 3–4% above it at rank eight, while the estimated mean retains a cosine near 0.98 at every rank. Finite-rank variance is therefore visible in these small blocks, whereas the mean direction changes little over the tested radii.

Table 9: Two-sided paired tests at equal evaluation cost, with five paired seeds and full-test GSM8K and MATH-500 evaluation. The 14B MATH-500 comparison favors LOO-ROLL with $p = 0.04591$. The other comparisons have $p \geq 0.05$.

Model	Task	Individual p
Qwen3-0.6B	Countdown	0.19198
Qwen3-0.6B	GSM8K	0.09112
Qwen3-0.6B	Next-token prediction	0.46333
Qwen3-1.7B	Countdown	0.68251
Qwen3-1.7B	GSM8K	0.51541
Qwen3-1.7B	MATH-500	0.89277
Qwen3-1.7B	Next-token prediction	0.53072
Qwen3-8B	GSM8K	0.11829
Qwen3-8B	MATH-500	0.46159
Qwen3-8B	Next-token prediction	0.23088
Qwen3-14B	GSM8K	0.84124
Qwen3-14B	MATH-500	0.04591
SmolLM2-1.7B	Countdown	0.64704
SmolLM2-1.7B	GSM8K	0.53016

Table 10: Two-sided paired tests for the fourteen equal-time LOO-ROLL comparisons. Holm adjustment treats all fourteen settings as one family. Bold adjusted values are below 0.05, and all nine corresponding effects favor LOO-ROLL.

Model	Task	Individual p	Holm-adjusted p
Qwen3-0.6B	GSM8K	0.000020	**0.00026**
Qwen3-1.7B	GSM8K	0.001532	**0.0123**
Qwen3-1.7B	MATH-500	0.000062	**0.00074**
Qwen3-8B	GSM8K	0.000004	**0.00006**
Qwen3-8B	MATH-500	0.000373	**0.00335**
Qwen3-14B	GSM8K	0.000199	**0.00199**
Qwen3-14B	MATH-500	0.000125	**0.00137**
SmolLM2-1.7B	Countdown	0.654584	1.000000
SmolLM2-1.7B	GSM8K	0.003708	**0.0260**
Qwen3-0.6B	Countdown	0.710583	1.000000
Qwen3-1.7B	Countdown	0.022238	0.088953
Qwen3-0.6B	Next-token prediction loss	0.055755	0.167266
Qwen3-1.7B	Next-token prediction loss	0.012190	0.060951
Qwen3-8B	Next-token prediction loss	0.003985	**0.0260**

Table 11: Expanded useful-radius comparison over five matched seeds. Reward changes are rank eight minus rank one in percentage points, so positive values favor rank eight. Wall overhead is the percentage increase in mean iteration time from rank one to rank eight. Brackets give paired 95% Student-t intervals. The bold entry excludes zero.

Model	Task	Reward change	Rank-8 wall overhead
Qwen3-0.6B	Countdown	$-.69\%$ $[-2.73, 1.34]$	$+12\%$
Qwen3-0.6B	GSM8K	$+2.09\%$ $[-.23, 4.41]$	$+12\%$
Qwen3-1.7B	Countdown	$+.28\%$ $[-1.61, 2.17]$	$+22\%$
Qwen3-1.7B	GSM8K	$\mathbf{+.76\%}$ $\mathbf{[.07, 1.46]}$	$+12\%$
Qwen3-4B	GSM8K	$+.21\%$ $[-1.30, 1.73]$	$+25\%$
Qwen3-8B	GSM8K	$-.70\%$ $[-1.58, .18]$	$+19\%$

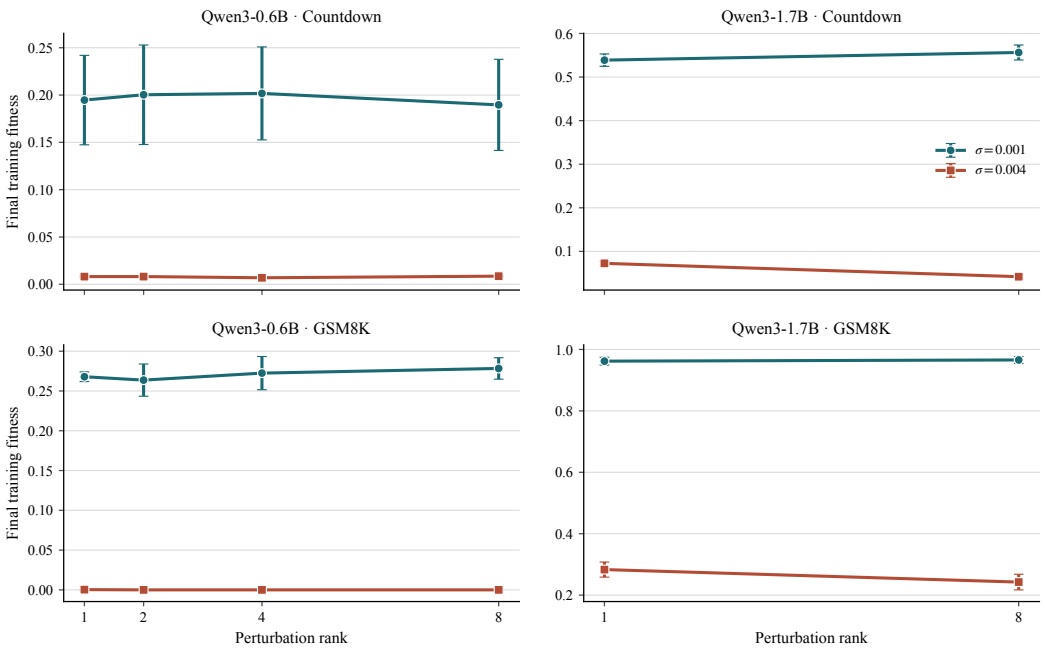


Figure 4: Mean final training fitness for the 0.6B and 1.7B rank studies. Bars are standard errors over three seeds. At $\sigma = 10^{-3}$, rank differences are small relative to seed variation. Increasing the radius to $\sigma = 4 \times 10^{-3}$ produces a much larger decline in fitness across ranks.

Table 12: Complete Qwen3-0.6B rank study. Entries are mean final training fitness $\pm$ standard error over three seeds. Higher values are better.

Task	σ	Rank 1	Rank 2	Rank 4	Rank 8
Countdown	.001	$.195 \pm .047$	$.200 \pm .053$	$.202 \pm .049$	$.190 \pm .048$
Countdown	.004	$.0081 \pm .0007$	$.0082 \pm .0007$	$.0069 \pm .0007$	$.0086 \pm .0004$
GSM8K	.001	$.268 \pm .006$	$.264 \pm .020$	$.272 \pm .021$	$.278 \pm .013$
GSM8K	.004	$.0003 \pm .0003$	$.0000 \pm .0000$	$.0000 \pm .0000$	$.0000 \pm .0000$

Table 13: Paired rank-eight minus rank-one reward changes over three matched seeds, in percentage points. Positive values favor rank eight. Brackets give two-sided 95% Student-t intervals, and bold entries exclude zero.

Model	Task	$\sigma = .001$	$\sigma = .004$
Qwen3-0.6B	Countdown	$-.51\%$ $[-5.83, 4.82]$	$+.04\%$ $[-.42, .51]$
Qwen3-0.6B	GSM8K	$+1.04\%$ $[-5.42, 7.50]$	$-.03\%$ $[-.17, .11]$
Qwen3-1.7B	Countdown	$+1.75\%$ $[-.15, 3.66]$	$\mathbf{-3.06\%}$ $\mathbf{[-5.18, -.94]}$
Qwen3-1.7B	GSM8K	$+.39\%$ $[-1.50, 2.29]$	$\mathbf{-4.07\%}$ $\mathbf{[-4.44, -3.70]}$

Table 14: Complete Qwen3-1.7B rank study. Rank columns give mean final training fitness $\pm$ standard error over three seeds. Higher fitness is better. The final column gives the paired rank-eight minus rank-one reward change in percentage points.

Task	σ	Rank 1	Rank 8	Reward change
Countdown	.001	$.539 \pm .014$	$.556 \pm .017$	$+1.75\%$
Countdown	.004	$.073 \pm .005$	$.042 \pm .004$	-3.06%
GSM8K	.001	$.962 \pm .013$	$.966 \pm .011$	$+.39\%$
GSM8K	.004	$.283 \pm .025$	$.243 \pm .025$	-4.07%

Table 15: Comparison of finite-rank and dense-Gaussian ES updates in two Qwen3-0.6B transformer blocks. Relative MSE is the single-direction squared error about the backpropagated block gradient, divided by the corresponding dense-Gaussian MSE, so 1 denotes equal sampling error and lower values are better. Mean-field cosine is the cosine similarity between the average of 8,192 antithetic updates and the backpropagated block gradient, so 1 denotes perfect directional agreement. Values average four perturbation radii within each of three independent perturbation seeds and are reported as means ± standard errors over seeds.

Quantity	Block	Rank 1	Rank 2	Rank 4	Rank 8
Relative MSE	MLP	$1.319 \pm .036$	$1.168 \pm .009$	$1.090 \pm .009$	$1.040 \pm .002$
Relative MSE	Attention	$1.264 \pm .018$	$1.127 \pm .018$	$1.089 \pm .025$	$1.028 \pm .015$
Mean-field cosine	MLP	$.9817 \pm .0013$	$.9830 \pm .0011$	$.9837 \pm .0003$	$.9827 \pm .0007$
Mean-field cosine	Attention	$.9800 \pm .0011$	$.9832 \pm .0009$	$.9836 \pm .0006$	$.9842 \pm .0012$

Table 16: Monte Carlo verification of the exact single-direction variance in Equation (72). Results use 100,000 independent sampled directions for a normalized 3×5 affine gradient.

Rank r	Exact variance	Monte Carlo variance	Relative error
1	34.00	34.24	0.70%
2	25.00	25.21	0.86%
4	20.50	20.62	0.57%
8	18.25	18.19	0.35%

F.4 Controlled numerical experiments

These experiments examine the parts of the theory that have exact numerical answers. Computing the covariance directly from sampled perturbation entries reproduces the tensor predicted by Equation (70), without evaluating the formula itself. Independent calculations also reproduce the resolvent multiplier at individual frequencies, the two Hessians in Proposition 3.2, the cubic identity in Proposition 3.4, and the $1/r$ rank expansion in Theorem 4.2. Together, these calculations verify the algebra before any model or optimization effects enter.

The finite-sampling experiment asks whether Monte Carlo estimates reproduce those exact quantities. Table 16 uses 100,000 directions for a 3×5 affine problem. The empirical single-direction variance is within 0.9% of Equation (72) at every rank. The same calculation recovers the stability transition in Figure 2: the largest Jacobian real part is 0.02925 at rank one and -0.00444 at rank two.

The quadratic case provides an exact test of the finite-population dynamics. Let $h(W) = \|W\|_F^2/2$, so that the exact population field is W by Proposition 3.3. If $\widehat{G}_N$ averages N directions, unbiasedness and Equation (20) give

$$\mathbb{E}\left[\|W - \eta \widehat{G}_N\|_F^2 \mid W\right] = \left[1 - 2\eta + \eta^2\left(1 + \frac{\kappa_r}{N}\right)\right]\|W\|_F^2. \tag{138}$$

Drawing a fresh independent population at every step and iterating this identity yields

$$\frac{\mathbb{E}[\|W_T\|_F^2]}{\|W_0\|_F^2} = \left[1 - 2\eta + \eta^2\left(1 + \frac{\kappa_r}{N}\right)\right]^T. \tag{139}$$

We test an 8×8 problem at $\eta = 0.3$ using 4,000 independent one-step trials per rank–population pair and 5,000 independent 25-step trajectories.

The synthetic result makes the finite-population prediction visible in a controlled setting. With few directions, the term κ_r/N can move the iteration across the stability boundary. Increasing N removes this difference, while increasing the matrix width makes the rank-one contribution small relative to the common $mn + 1$ term. The transformer experiments in Section 6 examine these predictions in a nonlinear model.

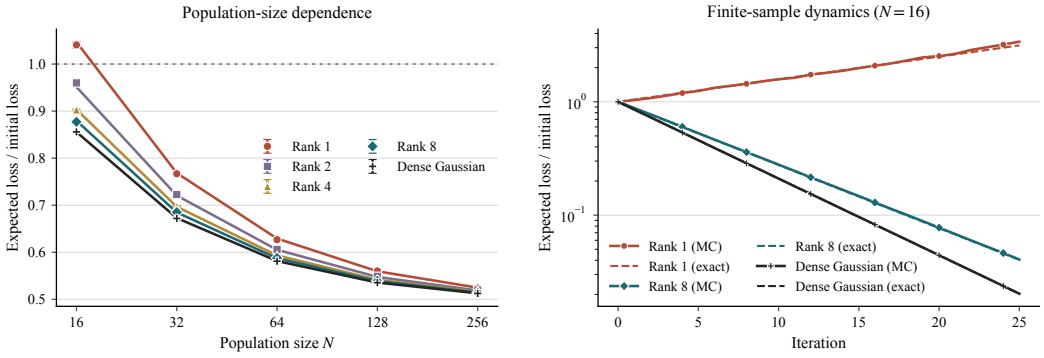


Figure 5: Finite-sample validation of Equation (139). **Left:** Monte Carlo one-step ratios agree with the exact curves across populations and ranks. The horizontal line marks stability. **Right:** at $N = 16$, the formula predicts the full mean-square trajectory. Rank one is unstable, whereas rank eight and dense Gaussian ES contract. Marked curves are Monte Carlo means and dashed curves are exact.

G Population centering and standardization

The operator analysis concerns the uncentered and unstandardized antithetic estimator, while the end-to-end experiments retain the population centering and standardization specified by EGGROLL. We therefore distinguish the direction of one realized update from the expectation of the normalized estimator. The proposition shows that prompt-wise centering and one common scale do not rotate a realized update. The corollary nevertheless shows that a scale computed from the same random population can change the expected field.

Proposition G.1 (Population centering and common standardization preserve each realized direction)**.** *Let pair j use perturbations $+E_j$ and $-E_j$, with prompt-level fitnesses $Y_{j,+,p}$ and $Y_{j,-,p}$. For arbitrary prompt offsets c_p, define*

$$Z_{j,\pm} = \frac{1}{P}\sum_{p=1}^{P}(Y_{j,\pm,p} - c_p). \tag{140}$$

Then $Z_{j,+} - Z_{j,-} = P^{-1}\sum_p(Y_{j,+,p} - Y_{j,-,p})$ exactly. If all $Z_{j,\pm}$ are subsequently divided by the same random scalar $S > 0$, the ES update is exactly S^{-1} times the unstandardized update. Consequently, per-prompt population centering has no effect on the update, and standardization changes only its realized magnitude.

Proof. The offsets cancel algebraically in every pair difference. The ES update is linear in those differences, so dividing every score by the same positive S multiplies their weighted sum by S^{-1} and cannot change its realized direction. □

Corollary G.2 (Bias induced by a same-population normalizer)**.** *At a fixed iterate, let $\widehat{g} = g + \xi$ with deterministic g and $\mathbb{E}[\xi] = 0$, and let $\alpha = S^{-1}$ have finite second moment. Then*

$$\|\mathbb{E}[\alpha\widehat{g}] - \mathbb{E}[\alpha]g\| \le \sqrt{\mathrm{Var}(\alpha)\mathbb{E}[\|\xi\|^2]}. \tag{141}$$

If α is measurable with respect to a sigma-field $\mathcal{G}$ and $\mathbb{E}[\xi \mid \mathcal{G}] = 0$, the left side is zero. Independence of α and ξ is sufficient. A same-population normalizer can otherwise change the expected field.

Proof. Since $\mathbb{E}[\xi] = 0$,

$$\mathbb{E}[\alpha\widehat{g}] - \mathbb{E}[\alpha]g = \mathbb{E}[(\alpha - \mathbb{E}[\alpha])\xi]. \tag{142}$$

Cauchy–Schwarz gives Equation (141). If α is $\mathcal{G}$-measurable and $\mathbb{E}[\xi \mid \mathcal{G}] = 0$, then $\mathbb{E}[\alpha\xi] = \mathbb{E}[\alpha\mathbb{E}[\xi \mid \mathcal{G}]] = 0$. □

The bound becomes explicit in the population size once the update mean and empirical scale satisfy concentration bounds.

Table 17: Comparison of mean updates before and after population score standardization, averaged over two transformer blocks, four perturbation radii, and three seeds. Cosine measures directional agreement between the raw and standardized means, with 1 indicating identical directions. Residual is the component of the standardized mean that cannot be explained by rescaling the raw mean, expressed as a percentage of the standardized mean's norm. Higher cosine and lower residual indicate a smaller directional effect.

N	Dense cosine	Dense residual	Rank-one cosine	Rank-one residual
32	.99954	3.03%	.99858	5.28%
64	.99979	2.04%	.99924	3.87%
128	.99990	1.44%	.99970	2.42%

Proposition G.3 (N-dependent standardization bias)**.** *For each population size N, write $\widehat{g}_N = g + \xi_N$ and $\alpha_N = S_N^{-1}$. Suppose that, for constants $V, A < \infty$ independent of N,*

$$\mathbb{E}[\|\xi_N\|^2] \le \frac{V}{N}, \qquad \mathrm{Var}(\alpha_N) \le \frac{A}{N}. \tag{143}$$

Then

$$\|\mathbb{E}[\alpha_N \widehat{g}_N] - \mathbb{E}[\alpha_N] g\| \le \frac{\sqrt{AV}}{N}. \tag{144}$$

Consequently, same-population standardization changes the raw population field by an $O(N^{-1})$ term, apart from the scalar factor $\mathbb{E}[\alpha_N]$.

Proof. Apply Equation (141) and substitute the two bounds in Equation (143). □

The first condition in Equation (143) is the ordinary N^{-1} variance reduction from averaging independent member updates. The second follows, for example, when the empirical scale concentrates at rate $N^{-1/2}$ around a positive population scale and its inverse has uniformly controlled second moments. A deterministic floor on the empirical scale supplies a direct sufficient condition. More precisely, if $S_N \ge s_0 > 0$ almost surely, the population scale is $s > 0$, and $\mathbb{E}[(S_N - s)^2] \le C_S/N$, then

$$\mathrm{Var}(S_N^{-1}) \le \mathbb{E}[(S_N^{-1} - s^{-1})^2] \le \frac{C_S}{s_0^2 s^2 N}, \tag{145}$$

so Equation (144) holds with $A = C_S/(s_0^2 s^2)$. Thus, the raw resolvent field remains the leading expected direction as the population grows.

G.1 Transformer-block audit

The released EGGROLL implementation divides all member scores by one standard deviation computed from the same population. We test the resulting finite-population effect with 4,096 coupled directions for each of two Qwen3-0.6B blocks, four radii from 10^{-3} to 8×10^{-3}, and three seeds. For each block, radius, and seed, the directions are partitioned into independent populations of size N. After fitting the best scalar multiple of the raw field, the directional residual is the norm of the remaining orthogonal component divided by the norm of the standardized field.

Table 17 shows that the normalized and raw fields remain closely aligned, and that their difference decreases with population size. At the training value $N = 128$, the rank-one cosine is 0.99970 with a 2.42% directional residual. The dense control gives 0.99990 and 1.44%. Fits across ranks $1, 2, 4, 8$ do not produce a coefficient whose direction and magnitude remain stable under the predicted σ^2 scaling, leaving that correction below the resolution of this audit.