SAMPLE-ADAPTIVE ZEROTH-ORDER POST-TRAINING

Anonymous authors
Paper under double-blind review

ABSTRACT

Zeroth-order (ZO) optimization offers a memory-efficient alternative to backpropagation for fine-tuning large language models. However, existing ZO methods typically sample training examples uniformly, overlooking their evolving importance and leaving avoidable estimation variance. To address this limitation, we propose Sample-Adaptive Zeroth-Order Optimization (SAZO), which learns sample importance online and adaptively allocates queries to accelerate convergence. SAZO extracts gradient-norm information from existing directional differences to estimate sample importance without access to true gradients. Furthermore, to track evolving importance from sparse minibatch feedback, SAZO maintains per-example exponential moving averages that aggregate observations across updates. The method requires no additional model evaluations and only $O(N)$ additional scalar state compared with MeZO. We prove that SAZO asymptotically matches the leading convergence term of the *minimum-variance* sampling oracle, and achieves a strictly faster convergence rate than standard methods. Extensive experiments across four model families demonstrate improved average downstream performance and up to $6\times$ faster convergence to the same training loss in terms of query count.

1 INTRODUCTION

Fine-tuning adapts pre-trained large language models (LLMs) to diverse downstream tasks, including sentiment analysis, natural language inference, and question answering (Devlin et al., 2019). However, standard fine-tuning incurs substantial memory overhead from stored activations, gradients, and optimizer states (Li et al., 2026), requiring up to $12\times$ the inference memory in the settings evaluated by Malladi et al. (2023). Parameter-efficient fine-tuning (PEFT), including prefix tuning (Li & Liang, 2021) and low-rank adaptation (LoRA) (Hu et al., 2021), reduces trainable parameters but still stores activations for backpropagation. To overcome this remaining activation-memory bottleneck, zeroth-order (ZO) methods such as MeZO bypass backpropagation and estimate updates from two forward evaluations, achieving a memory footprint close to inference (Malladi et al., 2023). This memory advantage comes with an optimization challenge: the resulting gradient estimates remain noisy, limiting the convergence speed of ZO fine-tuning (Nesterov & Spokoiny, 2017).

Inspired by first-order adaptive optimizers such as Adam (Kingma & Ba, 2015), mainstream approaches to accelerating ZO optimization focus on adaptation in parameter space. HiZOO (Zhao et al., 2025c) and PaZO (Zhao et al., 2025b) use Hessian-informed preconditioning, ZO-AdaMM (Chen et al., 2019) and R-AdaZO (Shu et al., 2025) adapt coordinate-wise update scales using historical moment estimates, and LOZO (Chen et al., 2025) exploits low-rank perturbation subspaces. However, ZO methods face a fundamental information bottleneck: each directional query reveals only a scalar projection rather than a full gradient. Identifying an unrestricted d-dimensional adaptation vector requires $\Omega(d)$ such observations, so the cost of learning reliable parameter-wise adaptation can offset its intended acceleration. We therefore ask:

> **A.** *Is there additional information in ZO queries that current optimizers leave untapped?*
>
> **B.** *If so, how can we adaptively exploit this information to accelerate ZO convergence?*

Since each update uses only a small subset of the training data, its effectiveness depends in part on which examples are selected. Our empirical analysis identifies an untapped source of information:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

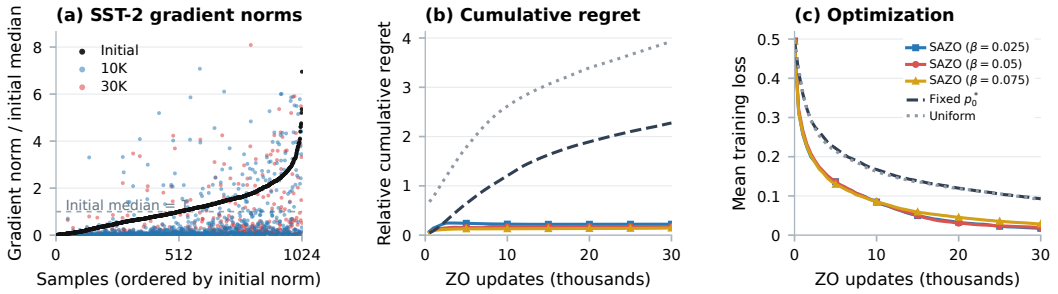

Figure 1: **From sample heterogeneity to allocation regret and optimization.** (a) Per-example gradient norms on SST-2 at initialization, 10K, and 30K updates, normalized by the initial median. Points retain the sample order induced by the initial norms, showing both cross-sample heterogeneity and temporal changes. (b) Sampling regret is the excess gradient-estimator variance over the variance-minimizing allocation (see Appendix E.5 for definitions and detailed settings); SAZO reduces this quantity versus uniform and fixed initial-oracle sampling. (c) The improved allocation accelerates loss reduction.

the evolving importance of individual training examples. Their gradient norms differ substantially and change during training (Figure 1(a); Appendix E.5), suggesting that queries need not be equally valuable across examples. Uniform sampling overlooks this heterogeneity, leaving avoidable estimation variance. To use this information for query allocation, a natural approach is importance sampling, inspired by first-order methods (Katharopoulos & Fleuret, 2018; Borsos et al., 2018; El Hanchi & Stephens, 2020): allocate more queries to examples with larger gradient norms. Although sampling heterogeneity also exists in first-order methods, we argue that zeroth-order algorithms are more sensitive to it, since the estimated gradient variance of zeroth-order methods using only function-value access is much larger than that of first-order methods[1].

However, applying importance sampling principle to ZO fine-tuning presents two challenges. **First, sample importance is not directly observable.** The variance-minimizing allocation depends on per-example gradient norms, $p^{\star}_{i,t} \propto \|\nabla f_i(x_t)\|$, whereas ZO queries reveal only random directional measurements. We therefore use the scaled magnitudes of existing directional differences to estimate these norms; their expectations approach the corresponding gradient norms as the perturbation radius vanishes, requiring no additional queries. **Second, evolving importance must be tracked from sparse minibatch feedback.** Each update reveals importance information only for examples in the sampled minibatch, while the importance of all examples may change as the model evolves. We therefore maintain per-example exponential moving averages to combine newly observed information with historical estimates, allowing importance tracking across intermittent observations. Uniform exploration provides opportunities to refresh infrequently sampled examples.

Together, these designs yield *Sample-Adaptive Zeroth-Order Optimization (SAZO)*. The learned scores guide sampling, while inverse-probability weighting preserves the expected full-data finite-difference update. SAZO requires no additional model evaluations and only $O(N)$ scalar state for N training examples. The benefit of tracking rather than fixing this allocation is visible in Figure 1(b): SAZO incurs lower relative cumulative sampling regret than uniform sampling and an oracle allocation fixed at initialization. This improved allocation also translates into faster training-loss reduction in Figure 1(c), connecting the recovered sample information to optimization progress.

Our main contributions are:

- We propose SAZO, which estimates sample importance from existing directional differences without extra model evaluations and tracks its evolution from sparse minibatch feedback using per-example exponential moving averages, requiring only $O(N)$ scalar state.
- We prove that SAZO asymptotically matches the leading convergence term of the minimum-variance sampling oracle, and achieves a strictly faster convergence rate than standard methods.

[1]The variance is actually $\Omega(d)$ times larger in the worst case.

- We conduct extensive experiments across four model families, demonstrating improved average downstream performance at MeZO-like memory and per-update cost. SAZO achieves up to $6\times$ query-count speedup over MeZO at the same training loss.

2 Related work

Sample-adaptive first-order optimization. Online variance minimization learns importance distributions from sampled feedback (Bouchard et al., 2016; Namkoong et al., 2017; Salehi et al., 2017). Static-regret analyses treat feedback as general loss sequences (Borsos et al., 2018; 2019); subsequent work connects sampling to Adam (Liu et al., 2020), federated convergence (Zeng et al., 2024; Zhao et al., 2025a), and joint convex optimization (Aolaritei et al., 2025). Avare exploits gradient continuity for dynamic-regret guarantees (El Hanchi & Stephens, 2020). SAZO addresses the ZO setting, where gradients are unavailable and directional feedback remains noisy, while accounting for the coupling between sampling and optimization.

Accelerating ZO post-training. MeZO establishes memory-efficient language-model fine-tuning using forward evaluations (Malladi et al., 2023). Subsequent approaches accelerate optimization through curvature-based preconditioning (Zhao et al., 2025c;b), low-rank perturbations (Chen et al., 2025), variance reduction (Gautam et al., 2024), and directional alignment (Ma & Huang, 2025). Other methods improve update normalization (Dang et al., 2026), memory-efficient optimizer adaptivity (Dbouk et al., 2026), or adaptive layer selection (Wang et al., 2026). These studies focus on constructing parameter updates and allocating perturbations within the model, where learning reliable parameter-wise adaptation is intrinsically difficult in high dimensions; SAZO instead adapts query allocation across examples.

3 Sample-adaptive zeroth-order optimization

Consider the finite-sum problem $\min_{x\in\mathbb{R}^d} F(x) = N^{-1}\sum_{i=1}^N f_i(x)$, where f_i denotes the loss of training example i. Given a uniformly sampled index I_t and an independent Gaussian direction $z_t \sim \mathcal{N}(0, I_d)$, ZO optimization uses the finite difference

$$s_{i,t} = \frac{f_i(x_t + \mu_t z_t) - f_i(x_t - \mu_t z_t)}{2\mu_t}. \tag{1}$$

to update $x_{t+1} = x_t - \eta s_{I_t,t} z_t$, with perturbation radius $\mu_t > 0$ and step size $\eta > 0$. Since $\mathbb{E}[z_t z_t^\top] = I_d$, the estimator satisfies $\mathbb{E}[s_{I_t,t} z_t \mid x_t] = \nabla F(x_t) + O(\mu_t^2)$ and is unbiased up to finite-difference error. Let $g_{i,t} = \nabla f_i(x_t)$ and $a_{i,t} = \|g_{i,t}\|$, and let $\mathbb{E}_t$ denote expectation conditional on the history before the current draws. The estimator's second moment is

$$\mathbb{E}_t \|s_{I_t,t} z_t\|^2 = \frac{d+2}{N}\sum_{i=1}^N a_{i,t}^2 + O(\mu_t^2). \tag{2}$$

The $O(\mu_t^2)$ terms account for finite-difference error under the Hessian regularity conditions in Assumption 1; their coefficients may depend on the sampling probabilities and gradient norms (Appendix C). The maximum perturbation radius is denoted by $\mu = \max_{t\le T} \mu_t$.

3.1 Sample-adaptive updates

The update above assigns equal query weight to every example, as does standard first-order SGD (Katharopoulos & Fleuret, 2018). Yet empirical evidence reveals pronounced sample heterogeneity in post-training: gradient norms vary substantially across examples and evolve during optimization (Figures 1(a) and 8). This heterogeneity exhibits interpretable sample-level structure that cannot be explained simply by label class, sequence length, or vocabulary (Appendices E.5 and E.6). These observations motivate SAZO (Algorithm 1), which uses importance sampling to assign different query weights across examples to reduce update variance and accelerate optimization.

Let $p_{i,t} > 0$ denote the sampling probability of example i, with $\sum_i p_{i,t} = 1$. For $I_t \sim p_t$ drawn independently of z_t, the importance-weighted update is

$$\widehat{g}_t = \frac{s_{I_t,t}}{N p_{I_t,t}} z_t, \qquad x_{t+1} = x_t - \eta \widehat{g}_t. \tag{3}$$

Algorithm 1 Sample-Adaptive Zeroth-Order Optimization

Require: $x_1, m_{i,1} > 0$, step budget T, radii $\mu_t > 0$, learning rate $\eta > 0$
Require: Exploration $\varepsilon \in (0, \frac{1}{2}]$ and EMA coefficient $\beta \in (0, 1)$
for $t = 1, \ldots, T$ **do**
  $p_{i,t} \leftarrow (1-\varepsilon)\dfrac{m_{i,t}}{\sum_j m_{j,t}} + \dfrac{\varepsilon}{N}$, for $i = 1, \ldots, N$. ▷ Importance sampling
  Draw $I_t \sim p_t$ and fresh $z_t \sim \mathcal{N}(0, I_d)$.
  $s_{I_t,t} \leftarrow \dfrac{f_{I_t}(x_t + \mu_t z_t) - f_{I_t}(x_t - \mu_t z_t)}{2\mu_t}$.
  $x_{t+1} \leftarrow x_t - \eta \dfrac{s_{I_t,t}}{N p_{I_t,t}} z_t$. ▷ Unbiased gradient estimator
  Copy $m_{t+1} \leftarrow m_t$.
  $m_{I_t,t+1} \leftarrow (1-\beta) m_{I_t,t} + \beta\sqrt{\dfrac{\pi}{2}}|s_{I_t,t}|$. ▷ Sparse update with EMA smoothing
end for

Importance weighting preserves the expected full-data ZO update, with $\mathbb{E}_t \widehat{g}_t = \nabla F(x_t) + O(\mu_t^2)$. Its second moment is

$$\mathbb{E}_t \|\widehat{g}_t\|^2 = \frac{d+2}{N^2} \sum_{i=1}^{N} \frac{a_{i,t}^2}{p_{i,t}} + O(\mu_t^2). \tag{4}$$

Under uniform sampling, $p_{i,t} = 1/N$, this second moment reduces to Eq. (2). For positive gradient norms, the distribution that minimizes the leading term and the resulting second moment are

$$p_{i,t}^\star = \frac{a_{i,t}}{\sum_j a_{j,t}}, \qquad \mathbb{E}_t \|\widehat{g}_t\|^2 \big|_{p_t = p_t^\star} = \frac{d+2}{N^2} \left(\sum_{i=1}^{N} a_{i,t} \right)^2 + O(\mu_t^2). \tag{5}$$

3.2 Estimating sample importance from directional differences

The ideal allocation in Eq. (5) requires the current per-example gradient norms. As identified in the Introduction, SAZO addresses two obstacles to learning them in ZO optimization.

Unobserved gradients. ZO optimization cannot access $a_{i,t} = \|\nabla f_i(x_t)\|$, but the directional difference $s_{i,t}$ already computed for the model update is approximately a one-dimensional Gaussian projection of the gradient. Its scaled absolute value therefore gives a gradient-norm estimator that is unbiased up to finite-difference error, without additional queries:

$$Y_{i,t} = \sqrt{\frac{\pi}{2}} |s_{i,t}|, \qquad \mathbb{E}_t Y_{i,t} = a_{i,t} + O(\mu_t^2). \tag{6}$$

Sparse feedback. Each update reveals $Y_{i,t}$ only for the sampled example, even though the importance of every example may evolve as optimization proceeds. SAZO therefore maintains a separate exponential moving average for every example, carrying historical estimates across the intervals between observations:

$$m_{i,t+1} = \begin{cases} (1-\beta) m_{i,t} + \beta Y_{i,t}, & i = I_t, \\ m_{i,t}, & i \neq I_t. \end{cases} \tag{7}$$

A larger β responds faster to changing importance, while a smaller β places more weight on accumulated observations and suppresses directional noise.

Finally, SAZO normalizes these smoothed estimates and adds uniform exploration, following the exploration–exploitation principle in EXP3 (Auer et al., 2002):

$$p_{i,t} = (1-\varepsilon) \frac{m_{i,t}}{\sum_j m_{j,t}} + \frac{\varepsilon}{N}. \tag{8}$$

Exploration refreshes stale scores, while the learned distribution guides the next queries. This creates a coupled estimation–optimization loop, analyzed under Assumption 1 in Section 3.3.

Query and storage cost. SAZO reuses the directional differences already computed for the model update and therefore requires no additional ZO queries. It stores only one scalar importance estimate per example, for $O(N)$ additional space; with 1,000 training examples and FP64 scores, this is 8,000 bytes, or less than 8 KiB, and is negligible relative to the inference-memory footprint measured in Section 4.3.

Initialization and minibatch implementation. We initialize each m_i by averaging its directional-magnitude observations over four complete, stable-order burn-in passes without replacement, using ordinary MeZO updates; these updates and queries count toward the training budget. We then switch to adaptive sampling with replacement and EMA tracking. Each minibatch shares one perturbation direction, as in MeZO (Malladi et al., 2023); repeated draws are evaluated once, with their multiplicities retained in the weighted update (Appendix D).

Recommended hyperparameters. We recommend $\beta = 0.05$ and $\varepsilon = 0.1$, which are used for all main SAZO results. This single setting performs robustly across tasks and datasets and across masked and autoregressive models of different scales; the controlled sweeps in Section 4.4 further show low sensitivity over broad ranges of both parameters.

3.3 Convergence analysis

This section analyzes convergence under the following continuity assumptions. Smoothness conditions of this type are standard in zeroth-order optimization and are routinely used to analyze ZO fine-tuning of language models (Nesterov & Spokoiny, 2017; Malladi et al., 2023). They are also well matched to our setting: Lipschitz gradients enable descent and stable tracking as the parameters evolve, while Lipschitz Hessians control the error of SAZO's symmetric finite differences.

Assumption 1. *Each ∇f_i is globally L_i-Lipschitz, and F is L_F-smooth and bounded below by F_*. Each Hessian is globally H_i-Lipschitz. The positive radii μ_t are deterministic, and the initialization and finite-time moments appearing below are finite. Let $L_2^2 = N^{-1}\sum_i L_i^2$.*

At the step-t parameters, ignoring finite-difference error, the optimal allocation has second moment $V_t^\star = \frac{d+2}{N^2}\left(\sum_{i=1}^N a_{i,t}\right)^2$. Under the sampling distribution p_t in Algorithm 1, the corresponding second moment is $V_t(p_t) = \frac{d+2}{N^2}\sum_{i=1}^N \frac{a_{i,t}^2}{p_{i,t}}$. Their cumulative difference,

$$\mathrm{Regret}_T = \sum_{t=1}^{T} \mathbb{E}\left[V_t(p_t) - V_t^\star\right] \geq 0, \tag{9}$$

satisfies the following bound.

Theorem 1 (Sampling regret)**.** *Under Assumption 1, run Algorithm 1 with fixed x_1 and constant η, ε, β within the horizon. Let $\mu = \max_{t\leq T}\mu_t \leq 1$, $\sigma_G^2 = \pi/2 - 1$, $E_0 = \mathbb{E}\|m_1 - a_1\|^2$, $r = \beta\varepsilon/N$, $\Lambda_d = (d+2)L_2^2$, and*

$$A_T = \frac{1}{T}\sum_{t=1}^{T}\sum_{i}\mathbb{E}a_{i,t}^2, \qquad \mathcal{D}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}V_t^\star. \tag{10}$$

If $\kappa = 192\Lambda_d\eta^2/(\varepsilon r^2) < 1$, then

$$\frac{\mathrm{Regret}_T}{T} \leq \frac{1}{1-\kappa}\left[\frac{16(d+2)E_0}{TN\varepsilon r} + \frac{24(d+2)\sigma_G^2\beta^2}{N\varepsilon r}A_T + (\kappa + 4\varepsilon)\,\mathcal{D}_T\right] + O(\mu^4). \tag{11}$$

The bound separates initialization, observation noise, target drift, and exploration, with finite differences contributing an additive remainder. The condition $\kappa < 1$ requires the feedback amplification of tracking error through sampling and parameter updates to be weaker than the error contraction induced by the EMA. Theorem 1 shows that our importance tracking method can asymptotically achieve the minimum sampling variance.

Substituting this result into the descent lemma (Nesterov, 2004, Lemma 1.2.3) yields the following convergence bound.

Table 1: Leading convergence terms on a common trajectory. The SAZO bound additionally includes the adaptation and finite-difference terms in Theorem 2; hindsight allocation is a comparator.

Sampling allocation	Average second moment	Leading term
Uniform (MeZO)	$\mathcal{U}_T$	$\sqrt{2L_F\Delta_F\mathcal{U}_T/T}$
Best fixed in hindsight	$\mathcal{H}_T$	$\sqrt{2L_F\Delta_F\mathcal{H}_T/T}$
Optimal at each step (SAZO leading term)	$\mathcal{D}_T$	$\sqrt{2L_F\Delta_F\mathcal{D}_T/T}$

Theorem 2 (Convergence). *Under Assumption 1, fix N, d, L_i and positive schedule coefficients; use*

$$\eta = c_\eta T^{-1/2}, \qquad \varepsilon = c_\varepsilon T^{-1/10}, \qquad \beta = c_\beta T^{-3/10}.$$

Let $\Delta_F = F(x_1) - F_*$ and define $K_T = \frac{(d+2)\sigma_G^2 c_\beta}{c_\varepsilon^2} A_T + \left(\frac{N^2\Lambda_d c_\eta^2}{c_\varepsilon^3 c_\beta^2} + c_\varepsilon\right)\mathcal{D}_T$. For $\Delta_F, \mathcal{D}_T > 0$, take $c_\eta = \sqrt{2\Delta_F/(L_F\mathcal{D}_T)}$, which minimizes the leading expression for fixed $\mathcal{D}_T$. If the parameters are feasible and $\kappa \le 1/2$, then

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\|\nabla F(x_t)\|^2 \le \sqrt{\frac{2L_F\Delta_F\mathcal{D}_T}{T}} + O\left(\frac{L_F c_\eta K_T}{T^{3/5}} + \frac{L_F c_\eta (d+2)E_0}{c_\varepsilon^2 c_\beta T}\right) + O(\mu^2). \tag{12}$$

The first $O(\cdot)$ hides only universal numerical constants; finite-difference coefficients are detailed in Appendix C.

The first term is the leading optimization term; under fixed problem and schedule coefficients and uniformly bounded trajectory moments, the adaptation correction is $o(T^{-1/2})$. Here $\mathcal{D}_T$ is the expected horizon average of $V_t^\star$ defined above. The choice of c_η is conditional on this trajectory-dependent quantity; proofs and parameter qualifications appear in Appendices A and B.

Table 1 compares the leading convergence terms for uniform sampling (MeZO), the best fixed allocation in hindsight, and the optimal allocation at each step, which governs the SAZO bound. With $p^{\mathrm{unif}} = (1/N, \ldots, 1/N)$ and Δ_N the probability simplex, define

$$\mathcal{U}_T = \frac{1}{T}\sum_t \mathbb{E}V_t(p^{\mathrm{unif}}) = \frac{d+2}{N}A_T, \qquad \mathcal{H}_T = \mathbb{E}\left[\inf_{p\in\Delta_N}\frac{1}{T}\sum_t V_t(p)\right].$$

Cauchy–Schwarz and minimization over the allocations give $\mathcal{U}_T \ge \mathcal{H}_T \ge \mathcal{D}_T$, and hence

$$\sqrt{\frac{2L_F\Delta_F\mathcal{U}_T}{T}} \ge \sqrt{\frac{2L_F\Delta_F\mathcal{H}_T}{T}} \ge \sqrt{\frac{2L_F\Delta_F\mathcal{D}_T}{T}}. \tag{13}$$

In particular, the gap between the uniform and optimal horizon-averaged second moments is

$$\mathcal{U}_T - \mathcal{D}_T = \frac{d+2}{T}\sum_{t=1}^{T}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left(a_{i,t} - \frac{1}{N}\sum_{j=1}^{N}a_{j,t}\right)^2\right] \ge 0. \tag{14}$$

Equality holds if and only if all examples have the same gradient norm at every step, almost surely. This comparison uses a common parameter trajectory and does not rank the separate trajectories generated by different samplers.

4 Experiments

We evaluate SAZO from five complementary perspectives: downstream performance on masked and generative language models, convergence across model scales, scaling with the training-population size, computational efficiency, and robustness to its sampling hyperparameters. Appendix E provides the detailed experimental settings and additional diagnostics.

Table 2: Experiments on RoBERTa-large (350M parameters) with $k = 512$ examples per class. LP: linear probing; FT: fine-tuning with Adam; PEFT: the better result between LoRA and prefix tuning for each task. MeZO and HiZOO results are reproduced. All tuning results report mean test accuracy, with standard deviations in subscripts. Bold: best ZO result.

Method	Sentiment		Natural language inference			Question type	Avg.
	SST-2	SST-5	SNLI	MNLI	RTE	TREC	
Zero-shot	79.0	35.5	50.2	48.8	51.4	32.0	49.5
LP	$91.3_{\pm 0.5}$	$51.7_{\pm 0.5}$	$80.9_{\pm 1.0}$	$71.5_{\pm 1.1}$	$73.1_{\pm 1.5}$	$89.4_{\pm 0.5}$	76.3
FT	$93.9_{\pm 0.7}$	$55.9_{\pm 0.9}$	$88.7_{\pm 0.8}$	$84.4_{\pm 0.8}$	$82.7_{\pm 1.4}$	$97.3_{\pm 0.2}$	83.8
FT (PEFT)	$94.2_{\pm 0.2}$	$55.3_{\pm 0.7}$	$88.3_{\pm 0.5}$	$83.9_{\pm 0.6}$	$83.2_{\pm 1.3}$	$97.4_{\pm 0.2}$	83.7
MeZO	$93.3_{\pm 0.7}$	$53.2_{\pm 1.4}$	$83.0_{\pm 1.0}$	$78.3_{\pm 0.5}$	$78.6_{\pm 2.0}$	$94.3_{\pm 1.3}$	80.1
MeZO (PEFT)	$93.4_{\pm 0.4}$	$53.6_{\pm 0.5}$	$84.8_{\pm 1.1}$	$79.8_{\pm 1.2}$	$77.6_{\pm 1.3}$	$95.0_{\pm 0.7}$	80.7
HiZOO	$93.5_{\pm 0.4}$	$53.5_{\pm 1.2}$	$83.3_{\pm 1.4}$	$77.2_{\pm 1.5}$	$79.1_{\pm 1.2}$	$94.9_{\pm 1.7}$	80.3
HiZOO (PEFT)	$\mathbf{94.3}_{\pm 0.5}$	$54.2_{\pm 0.4}$	$85.7_{\pm 0.7}$	$79.7_{\pm 1.0}$	$80.9_{\pm 1.6}$	$93.1_{\pm 0.5}$	81.3
SAZO	$\mathbf{94.3}_{\pm 0.5}$	$53.6_{\pm 0.8}$	$85.3_{\pm 0.4}$	$79.6_{\pm 0.3}$	$\mathbf{82.3}_{\pm 2.4}$	$96.4_{\pm 0.6}$	81.9
SAZO (PEFT)	$94.0_{\pm 0.8}$	$\mathbf{54.5}_{\pm 0.5}$	$\mathbf{86.5}_{\pm 0.4}$	$\mathbf{81.6}_{\pm 1.4}$	$80.9_{\pm 1.7}$	$\mathbf{96.6}_{\pm 0.5}$	**82.4**

4.1 Performance and convergence across model families

4.1.1 Masked language models

We first examine whether SAZO's optimization strategy translates into downstream accuracy. Using RoBERTa-large (350M) (Liu et al., 2019) with $k = 512$ examples per class, we consider six tasks spanning sentiment classification, natural language inference, and question classification. Table 2 compares zero-shot learning, LP, and full-parameter and parameter-efficient FT, MeZO, HiZOO, and SAZO; across tasks, SAZO outperforms zero-shot and LP and approaches FT.

SAZO improves both full-parameter and parameter-efficient ZO tuning. Relative to the stronger of MeZO and HiZOO, SAZO raises the average accuracy by 1.6 percentage points under full-parameter tuning and by 1.1 points under PEFT. The same pattern appears on individual tasks: the gains on TREC are 1.5 and 1.6 points for full-parameter tuning and PEFT, respectively, while those on MNLI are 1.3 and 1.8 points.

4.1.2 Generative language models

We then test whether the benefits extend to generative language models (GLMs). Table 3 compares full-parameter SAZO, MeZO, and HiZOO on OPT-1.3B (Zhang et al., 2022), Phi-2 (2.7B), and Llama3-8B across eight classification, multiple-choice, and question-answering tasks.

SAZO achieves SOTA average performance compared with other ZO optimizers. On OPT-1.3B, Phi-2, and Llama3-8B, SAZO improves the average over the stronger of MeZO and HiZOO by 1.2, 1.7, and 3.9 percentage points, respectively. The gains are especially large on Llama3-8B CB and DROP, where SAZO exceeds the stronger baseline by 10.8 and 8.4 points.

4.1.3 Convergence across model scales

SAZO accelerates optimization across model scales. Figure 2 compares SST-2 training-loss trajectories on RoBERTa-large, Phi-2, and Llama 3 at a common forward-pass budget of 24K. Each method uses its optimal learning rate from the same search grid for each model; detailed settings are provided in Appendices E.1 and E.2. SAZO reaches MeZO's endpoint losses in 4K, 6K, and 6K passes, while HiZOO requires 10.5K, 9K, and 6K passes, respectively.

4.2 Scaling with training-population size

SAZO remains effective as the training population grows. Having established gains across model scales, we vary the amount of training data. Figure 3 compares SST-2 test accuracy as the RoBERTa-

Table 3: Full-parameter generative-LM results. MeZO and HiZOO results are reproduced.

Model	Method	Classification				Multiple choice		Question answering		Avg.
		SST-2	BoolQ	CB	WiC	COPA	MultiRC	SQuAD	DROP	
OPT-1.3B	MeZO	88.3	64.0	69.6	57.4	76.0	**61.9**	73.1	**23.7**	64.2
	HiZOO	90.7	64.0	**71.4**	58.2	75.0	60.1	**76.6**	21.3	64.7
	SAZO	**90.8**	**65.0**	**71.4**	**60.8**	**77.0**	**61.9**	**76.6**	23.3	**65.9**
Phi-2	MeZO	86.6	72.4	75.0	54.4	86.0	78.2	85.7	37.8	72.0
	HiZOO	**88.9**	72.0	75.2	**59.2**	86.0	**79.2**	85.7	36.2	72.8
	SAZO	88.5	**78.2**	**76.8**	57.2	**89.0**	**79.2**	**86.0**	**41.1**	**74.5**
Llama3-8B	MeZO	92.2	76.7	69.6	**57.8**	88.0	77.6	86.7	57.1	75.7
	HiZOO	94.3	77.1	69.6	57.7	89.0	77.9	87.7	49.4	75.3
	SAZO	**94.4**	**80.8**	**80.4**	56.6	**91.0**	**78.8**	**89.3**	**65.5**	**79.6**

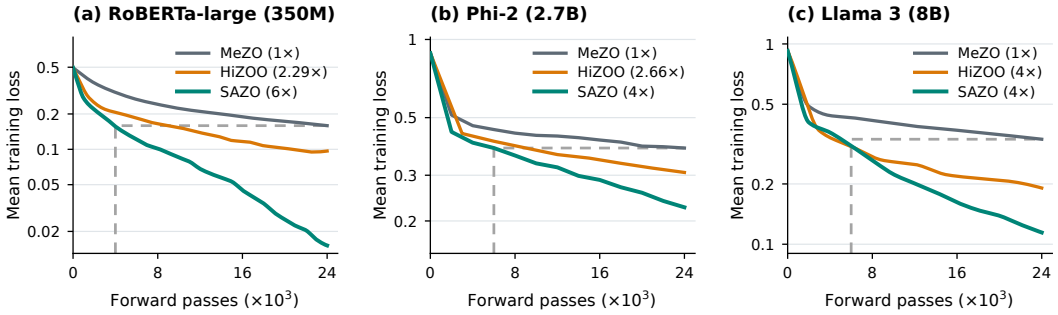

Figure 2: **SAZO substantially accelerates convergence across model scales.** At 24K forward passes, SAZO reaches MeZO's endpoint loss $6\times$, $4\times$, and $4\times$ faster in (a)–(c), respectively; HiZOO achieves $2.29\times$, $2.66\times$, and $4\times$.

large training population increases from $N = 1{,}024$ to $4N$. SAZO outperforms MeZO at every scale, with gains of 0.6, 1.1, 1.4, and 1.5 percentage points at N, $2N$, $3N$, and $4N$, respectively.

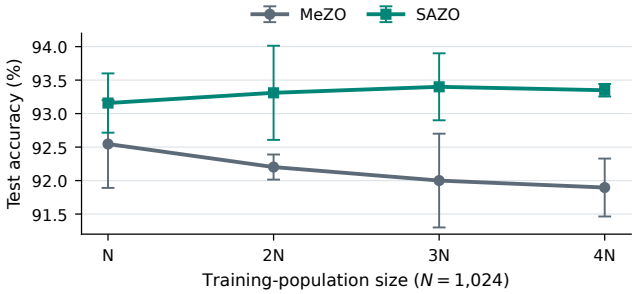

Figure 3: RoBERTa-large SST-2 test accuracy as the training population scales from $N = 1{,}024$ to $4N$. Curves show mean $\pm$ standard deviation over three seeds. See Appendix E.3 for details.

4.3 Memory and wall-clock efficiency

We finally examine whether the gains above preserve the computational advantages of ZO optimization. **SAZO retains MeZO's memory footprint.** Across the OPT configurations in Figure 4, the measured GB values for SAZO match those of MeZO. Compared with first-order methods, SAZO reduces memory requirements by factors of 6.8–$12.2\times$ for full FT and 3.3–$6.1\times$ for prefix FT (26 versus 316 GB at 13B for full FT). Thus, the downstream improvements do not sacrifice the low-memory advantage that makes ZO post-training attractive.

SAZO also maintains low per-update latency. Table 4 shows that its per-update time remains comparable to MeZO across the evaluated models. Together with faster loss reduction, this yields more effective updates at MeZO-like cost.

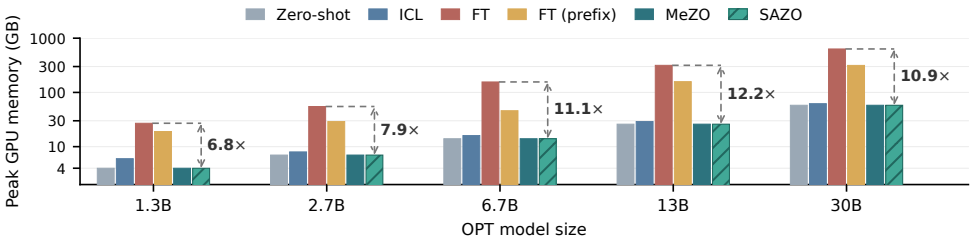

Figure 4: Peak GPU memory on MultiRC (mean length 400 tokens; batch size 1/device), summed across GPUs. Arrows show FT/SAZO ratios.

Table 4: Wall-clock time per update (seconds).

Method	RoBERTa-large	OPT-1.3B	Phi-2	Llama3-8B
MeZO	0.1153	0.1836	0.4235	0.9490
HiZOO	0.2738	0.4793	1.0519	2.4798
SAZO	0.1128	0.1839	0.4298	0.9805

4.4 Ablation studies

We conclude by examining sensitivity to the importance-score update. **SAZO is robust to the EMA coefficient.** Figure 5 varies this coefficient on SST-2. Across $\beta \in \{0.005, 0.05, 0.5\}$, the trajectories remain stable and reach closely clustered endpoint losses, with $\beta = 0.05$ performing best. In contrast, $\beta = 1$ discards the accumulated history, relies only on the latest directional observation, and diverges. This contrast highlights the importance of per-example moving averages for tracking evolving importance from intermittent feedback while smoothing directional noise. SAZO is similarly robust to the exploration coefficient ε over a broad range; Appendix E.4 provides the full sweep and experimental details.

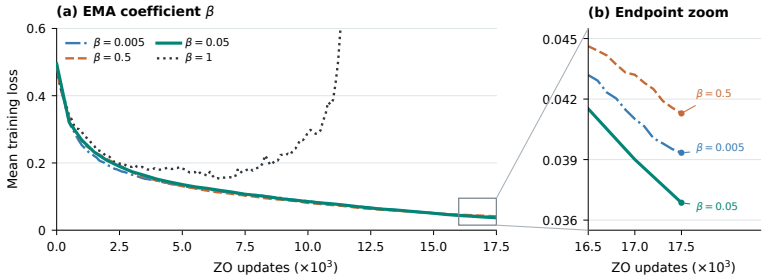

Figure 5: EMA-coefficient ablation on SST-2. Detailed settings are provided in Appendix E.4.

5 Conclusion

In this work, we propose SAZO, a sample-adaptive zeroth-order optimizer that reuses existing directional differences to estimate sample importance without additional model evaluations. Per-example exponential moving averages track how this importance evolves from sparse minibatch feedback and require only $O(N)$ scalar state. Our analysis asymptotically matches the variance-optimal dynamic oracle's leading convergence term, yielding a smaller convergence coefficient than uniform sampling when sample importance is heterogeneous. Extensive experiments demonstrate improved average downstream performance and up to $6\times$ faster convergence to the same training loss in terms of query counts.

AI USE STATEMENT

Generative AI assisted with research organization, literature retrieval, mathematical derivation, manuscript drafting, LaTeX preparation, numerical identity checks, and a separate adversarial review of this draft. It did not generate the reported training measurements.

REPRODUCIBILITY STATEMENT

The main algorithm and its analysis are specified in Section 3, with the analysis in Section 3.3. Appendix D specifies the with-replacement minibatch implementation and its shared-direction analysis, and Appendix E records the detailed experimental settings, curve configurations, and diagnostic definitions. Figure sources include the measured audit data and plotting code. Numerical checks of selected identities are supplementary checks, not substitutes for proofs or training results.

REFERENCES

Liviu Aolaritei, Bart P. G. Van Parys, Henry Lam, and Michael I. Jordan. Stochastic optimization with optimal importance sampling. *arXiv preprint arXiv:2504.03560*, 2025. URL https://arxiv.org/abs/2504.03560.

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002. doi: 10.1137/S0097539701398375. URL https://doi.org/10.1137/S0097539701398375.

Zalan Borsos, Andreas Krause, and Kfir Y. Levy. Online variance reduction for stochastic optimization. In *Conference On Learning Theory*, volume 75, pp. 324–357. PMLR, 2018. URL https://proceedings.mlr.press/v75/borsos18a.html.

Zalán Borsos, Sebastian Curi, Kfir Yehuda Levy, and Andreas Krause. Online variance reduction with mixtures. In *International Conference on Machine Learning*, volume 97, pp. 705–714. PMLR, 2019. URL https://proceedings.mlr.press/v97/borsos19a.html.

Guillaume Bouchard, Théo Trouillon, Julien Perez, and Adrien Gaidon. Online learning to sample. *arXiv preprint arXiv:1506.09016*, 2016. URL https://arxiv.org/abs/1506.09016v2.

Xiangyi Chen, Sijia Liu, Kaidi Xu, Xingguo Li, Xue Lin, Mingyi Hong, and David Cox. ZO-AdaMM: Zeroth-order adaptive momentum method for black-box optimization. In *Advances in Neural Information Processing Systems*, 2019.

Yiming Chen, Yuan Zhang, Liyuan Cao, Kun Yuan, and Zaiwen Wen. Enhancing zeroth-order finetuning for language models with low-rank structures. In *International Conference on Learning Representations*, volume 2025, pp. 62581–62607, 2025.

Sizhe Dang, Yangyang Guo, Yanjun Zhao, Xiaodong Zheng, Guang Dai, Ivor Tsang, and Haishan Ye. FZOO: Fast zeroth-order optimizer for fine-tuning large language models towards adam-scale speed. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2506.09034.

Hassan Dbouk, Nidham Gazagnadou, Matthias Reisser, and Christos Louizos. On adaptivity in zeroth-order optimization. *arXiv preprint arXiv:2605.03869*, 2026. URL https://arxiv.org/abs/2605.03869.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pp. 4171–4186, 2019.

Ayoub El Hanchi and David A. Stephens. Adaptive importance sampling for finite-sum optimization and sampling with decreasing step-sizes. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/b58f7d184743106a8a66028b7a28937c-Abstract.html.

Tanmay Gautam, Youngsuk Park, Hao Zhou, Parameswaran Raman, and Wooseok Ha. Variance-reduced zeroth-order methods for fine-tuning language models. *arXiv preprint arXiv:2404.08080*, 2024.

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. *arXiv preprint arXiv:2106.09685*, 2021.

Angelos Katharopoulos and François Fleuret. Not all samples are created equal: Deep learning with importance sampling. In *International Conference on Machine Learning*, volume 80, pp. 2525–2534. PMLR, 2018. URL https://proceedings.mlr.press/v80/katharopoulos18a.html.

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

Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)*, pp. 4582–4597, 2021.

Yanyi Li, Yimu Zhang, and Cong Fang. Prac: Principal-random subspace for llm activation compression and memory-efficient training. *arXiv preprint arXiv:2602.23111*, 2026.

Rui Liu, Tianyi Wu, and Barzan Mozafari. Adam with bandit sampling for deep learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://arxiv.org/abs/2010.12986.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019. URL https://arxiv.org/abs/1907.11692.

Shaocong Ma and Heng Huang. Revisiting zeroth-order optimization: Minimum-variance two-point estimators and directionally aligned perturbations. In *International Conference on Learning Representations*, volume 2025, pp. 13267–13293, 2025.

Sadhika Malladi, Tianyu Gao, Eshaan Nichani, Alex Damian, Jason D. Lee, Danqi Chen, and Sanjeev Arora. Fine-tuning language models with just forward passes. In *Advances in Neural Information Processing Systems*, 2023. URL https://arxiv.org/abs/2305.17333.

Hongseok Namkoong, Aman Sinha, Steve Yadlowsky, and John C. Duchi. Adaptive sampling probabilities for non-smooth optimization. In *International Conference on Machine Learning*, volume 70, pp. 2574–2583. PMLR, 2017. URL https://proceedings.mlr.press/v70/namkoong17a.html.

Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, 2004. doi: 10.1007/978-1-4419-8853-9. URL https://doi.org/10.1007/978-1-4419-8853-9.

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

Farnood Salehi, L. Elisa Celis, and Patrick Thiran. Stochastic optimization with bandit sampling. *arXiv preprint arXiv:1708.02544*, 2017. URL https://arxiv.org/abs/1708.02544.

Yao Shu, Qixin Zhang, Kun He, and Zhongxiang Dai. Refining adaptive zeroth-order optimization at ease. In *Proceedings of the International Conference on Machine Learning*, 2025.

Fei Wang, Li Shen, Liang Ding, Chao Xue, Ye Liu, and Changxing Ding. Universally empowering zeroth-order optimization via adaptive layer-wise sampling. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 40218–40239. Association for Computational Linguistics, 2026. URL https://aclanthology.org/2026.findings-acl.1999/.

Dun Zeng, Zenglin Xu, Yu Pan, Xu Luo, Qifan Wang, and Xiaoying Tang. Enhanced federated optimization: Adaptive unbiased client sampling with reduced variance. *arXiv preprint arXiv:2310.02698*, 2024. URL https://arxiv.org/abs/2310.02698v3.

Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, Todor Mihaylov, Myle Ott, Sam Shleifer, Kurt Shuster, Daniel Simig, Punit Singh Koura, Anjali Sridhar, Tianlu Wang, and Luke Zettlemoyer. OPT: Open pre-trained transformer language models. *arXiv preprint arXiv:2205.01068*, 2022. URL https://arxiv.org/abs/2205.01068.

Boxin Zhao, Lingxiao Wang, Ziqi Liu, Zhiqiang Zhang, Jun Zhou, Chaochao Chen, and Mladen Kolar. Adaptive client sampling in federated learning via online learning with bandit feedback. *Journal of Machine Learning Research*, 26(8):1–67, 2025a. URL https://jmlr.org/papers/v26/24-0385.html.

Hanzhen Zhao, Ding Shihong, Cong Fang, and Zhouchen Lin. Pazo: Preconditioned accelerated zeroth-order optimization for fine-tuning llms. *Advances in Neural Information Processing Systems*, 38:111163–111195, 2025b.

Yanjun Zhao, Sizhe Dang, Haishan Ye, Guang Dai, Yi Qian, and Ivor W. Tsang. Second-order fine-tuning without pain for LLMs: A hessian informed zeroth-order optimizer. In *International Conference on Learning Representations*, 2025c. URL https://arxiv.org/abs/2402.15173.

A PROOFS OF THE MAIN RESULTS

Let $\mathcal{F}_t$ denote the history before the current draws, so x_t, m_t, p_t are $\mathcal{F}_t$-measurable and $\mathbb{E}_t$ is conditional expectation. For the proofs, it is convenient to remove the common Gaussian factor from the second moments in Section 3.3. Throughout the appendices, define

$$\widetilde{V}_t(p) = \frac{V_t(p)}{d+2} = \frac{1}{N^2}\sum_i \frac{a_{i,t}^2}{p_i}, \qquad \bar{a}_t = \frac{1}{N}\sum_i a_{i,t}, \qquad \widetilde{V}_t^\star = \frac{V_t^\star}{d+2} = \bar{a}_t^2,$$

Writing $\rho_t = V_t(p_t) - V_t^\star$, use $\widetilde{\rho}_t = \rho_t/(d+2)$, $\widetilde{\mathrm{Regret}}_T = \mathrm{Regret}_T/(d+2)$, $\widetilde{\mathcal{D}}_T = \mathcal{D}_T/(d+2)$, $\widetilde{\mathcal{U}}_T = \mathcal{U}_T/(d+2)$, $\widetilde{\mathcal{H}}_T = \mathcal{H}_T/(d+2)$, and $\widetilde{K}_T = K_T/(d+2)$. Also set $C_d = L_F(d+2)$. These are exact rescalings of the main-text quantities.

For the explicit radius calculations below, set $D_6 = d(d+2)(d+4)$, $D_8 = d(d+2)(d+4)(d+6)$, $H_2^2 = N^{-1}\sum_i H_i^2$, $J_{i,t} = \pi H_i^2 D_6 \mu_t^4/72$, and $W_t = H_2^2 D_8 \mu_t^4/36$. Write $\overline{W}_T = T^{-1}\sum_t W_t$ and $\mathcal{B}_T = T^{-1}\sum_{t,i}\mathbb{E}[\beta^2 p_{i,t} J_{i,t}]$. With $r = \beta\varepsilon/N$, define

$$R_{\mu,T} = \frac{\mathcal{B}_T + NL_2^2\eta^2\overline{W}_T/\varepsilon}{N\varepsilon r^2}, \qquad \Omega_{\mu,T} = \sqrt{\widetilde{\mathcal{D}}_T\overline{W}_T} + \frac{L_F\eta}{\varepsilon}\sqrt{\frac{(d+2)A_T\overline{W}_T}{N}} + \frac{L_F\eta}{\varepsilon}\overline{W}_T.$$

These quantities expand the compact radius errors used in the main text. At a fixed horizon and fixed algorithm parameters, $R_{\mu,T} = O(\mu^4)$ and $\Omega_{\mu,T} = O(\mu^2)$ for $\mu = \max_t \mu_t \le 1$, with moment dependencies retained here. For local observation variance statements, the hidden coefficient also contains $a_{i,t}$ through $a_{i,t}\sqrt{J_{i,t}}$.

A.1 SUPPORTING MOMENT AND DESCENT BOUNDS

The following results use the normalized quantities defined above. Multiplying the normalized regret bound by $d+2$ gives Theorem 1; substituting $C_d = L_F(d+2)$ and $\widetilde{\mathcal{D}}_T = \mathcal{D}_T/(d+2)$ gives Theorem 2.

Proposition 1 (Feedback and update moments)**.** *For $\sigma_G^2 = \pi/2 - 1$, the finite-difference observations satisfy*

$$\begin{aligned} \mathbb{E}_t Y_{i,t} &= a_{i,t} + O(\mu^2), \\ \mathrm{Var}_t(Y_{i,t}) &= \sigma_G^2 a_{i,t}^2 + O(\mu^2). \end{aligned} \tag{15}$$

The corresponding update satisfies

$$\begin{aligned} \|\mathbb{E}_t\widehat{g}_t - \nabla F(x_t)\| &= O(\mu^2), \\ \mathbb{E}_t\|\widehat{g}_t\|^2 &\le 2(d+2)\widetilde{V}_t(p_t) + O(\mu^4). \end{aligned} \tag{16}$$

Lemma 1 (Descent decomposition)**.** *Under Assumption 1, let $\Psi_T = T^{-1}\sum_t \mathbb{E}\|\nabla F(x_t)\|^2$. Then*

$$\Psi_T \le \frac{\Delta_F}{\eta T} + \frac{C_d\eta}{2}\widetilde{\mathcal{D}}_T + \frac{C_d\eta}{2}\frac{\widetilde{\mathrm{Regret}}_T}{T} + O(\mu^2), \tag{17}$$

where $C_d = L_F(d+2)$ and $\Delta_F = F(x_1) - F_*$.

Schedules and finite-difference accuracy. The schedules in Theorem 2 are constant within each run and indexed by the horizon. The leading-term choice of c_η requires $\mathcal{D}_T$ and Δ_F; since $\mathcal{D}_T$ depends on the resulting trajectory, this is a conditional tuning result. The finite-time bounds retain the trajectory moments without imposing a global gradient bound, convexity, or a strong-growth condition. For fixed problem and schedule coefficients and uniformly bounded trajectory moments, the $O(\mu^2)$ constant can be taken uniformly over feasible horizons. A run-wise radius $\mu = \mu_0 T^{-\alpha}$ with $\alpha > 3/10$ makes this correction $o(T^{-3/5})$; a fixed radius retains an additive error floor. Appendix C gives the explicit bounds.

A.2 Gaussian identities and the sampling oracle

In this appendix only, write $s_{i,t}^{\mu_t} = s_{i,t}$ for the actual difference used in the main text and introduce $s_{i,t}^0 = z_t^\top g_{i,t}$ as a Taylor expansion term at the same current parameter vector. There is no auxiliary optimization trajectory. Condition on $\mathcal{F}_t$ and suppress t. For any fixed g, $z^\top g$ has distribution $\|g\|Z$ with $Z \sim \mathcal{N}(0,1)$. Since $\mathbb{E}|Z| = \sqrt{2/\pi}$ and $\mathbb{E}Z^2 = 1$, the Gaussian moments of $Y_i^0 = \sqrt{\pi/2}|s_i^0|$ follow. Adding the Taylor remainder bounded in Appendix C gives Eq. (15). To compute the update moment, rotate g onto the first coordinate:

$$\mathbb{E}_z\|zz^\top g\|^2 = \|g\|^2\mathbb{E}\left[z_1^2\sum_{j=1}^d z_j^2\right] = (3+d-1)\|g\|^2.$$

Also $\mathbb{E}zz^\top = I_d$. Averaging over the selected index gives the leading moment $(d+2)\widetilde{V}_t(p_t)$. Adding the Taylor remainder gives Eq. (16). Finally,

$$\left(\sum_i a_i\right)^2 = \left(\sum_i \frac{a_i}{\sqrt{p_i}}\sqrt{p_i}\right)^2 \le \sum_i \frac{a_i^2}{p_i}.$$

Equality holds for $p_i \propto a_i$ on the positive-norm support. Infima on the boundary are understood by limits, with zero-norm terms contributing zero.

A.3 Stationary variance of the EMA

Fix the parameter vector and index i, and enumerate its visits by k. At a fixed radius, with fresh independent Gaussian directions, let $a_i^\mu = \mathbb{E}Y_i^{(k)}$ and $v_i^\mu = \mathrm{Var}(Y_i^{(k)})$. The central-difference estimates obey $a_i^\mu = a_i + O(\mu^2)$ with the explicit coefficients in Proposition 1. For deterministic initial score $m_i^{(0)}$, the recursion $m_i^{(k)} = (1-\beta)m_i^{(k-1)} + \beta Y_i^{(k)}$ gives

$$\mathbb{E}(m_i^{(k)} - a_i^\mu)^2 = (1-\beta)^{2k}(m_i^{(0)} - a_i^\mu)^2 + \frac{\beta}{2-\beta}\big[1-(1-\beta)^{2k}\big]v_i^\mu.$$

This follows by centering the recursion and summing a geometric series. At visit times, the variance converges to $\beta v_i^\mu/(2-\beta)$. Averaging reduces variance but retains the finite-difference bias in a_i^μ. The result concerns a fixed target; it does not remove the drift term on an evolving optimization trajectory.

A.4 A global normalization bound

Lemma 2. *For $a \ge 0$, $m > 0$, $0 < \varepsilon \le 1/2$, and $p = (1-\varepsilon)m/(\sum_i m_i) + \varepsilon\mathbf{1}/N$, let $G = N^{-1}\sum_i a_i$ and $e = m - a$. Then*

$$\widetilde{V}(p) - G^2 \le \frac{8}{N\varepsilon}\|e\|^2 + 4\varepsilon G^2. \tag{18}$$

Proof. Let $S = \sum_i a_i$, $M = \sum_i m_i$, $q = m/M$, and $u = \mathbf{1}/N$. To retain the normalization by N^2, work directly with the weighted squared norm $\|v\|_{p^{-1}}^2 = \sum_i v_i^2/p_i$. Expanding the square gives

$$N^2\big[\widetilde{V}(p) - G^2\big] = \|a - Sp\|_{p^{-1}}^2.$$

Write $\delta = a - Sq = -e + (M-S)q$. Since $p_i \ge \varepsilon/N$, $p_i \ge (1-\varepsilon)q_i$, and $(M-S)^2 \le N\|e\|^2$,

$$\begin{aligned}\|\delta\|_{p^{-1}}^2 &\le 2\sum_i \frac{e_i^2}{p_i} + 2(M-S)^2\sum_i \frac{q_i^2}{p_i} \\ &\le \left(\frac{2N}{\varepsilon} + \frac{2N}{1-\varepsilon}\right)\|e\|^2 = \frac{2N}{\varepsilon(1-\varepsilon)}\|e\|^2.\end{aligned}$$

The exploration contribution also admits a weighted bound:

$$\varepsilon^2\|q-u\|_{p^{-1}}^2 = \|q-p\|_{p^{-1}}^2 = \sum_i \frac{q_i^2}{p_i} - 1 \le \frac{\varepsilon}{1-\varepsilon}.$$

Using $a - Sp = \delta + S\varepsilon(q - u)$ therefore yields

$$\begin{aligned}\widetilde{V}(p) - G^2 &\le \frac{4}{N\varepsilon(1-\varepsilon)}\|e\|^2 + \frac{2\varepsilon}{1-\varepsilon}G^2 \\ &\le \frac{8}{N\varepsilon}\|e\|^2 + 4\varepsilon G^2.\end{aligned}$$

The argument includes $S = 0$ and requires no closeness condition between M and S. □

A.5 Sparse EMA with correlated target drift

The following lemma allows dependent inclusions and shared directional observations. Given the current history, let O_i be an inclusion indicator independent of the fresh Gaussian direction, with $\pi_i = \mathbb{E}_t O_i$. The indicators for different coordinates may be dependent. For this local lemma, let $Y_i = Y_i^0$ denote the leading Gaussian observation and update $m_i^+ = m_i + \beta O_i(Y_i - m_i)$, and suppose $\beta\pi_i \ge r > 0$ for all steps and coordinates. Define $\delta a_{i,t} = a_{i,t+1} - a_{i,t}$.

Lemma 3 (Visited-coordinate tracking). *With $E_0 = \mathbb{E}\|m_1 - a_1\|^2$ and $Q_T = \sum_{t<T}\mathbb{E}\|a_{t+1} - a_t\|^2$,*

$$\sum_t \mathbb{E}\|m_t - a_t\|^2 \le \frac{2E_0}{r} + \frac{3\sigma_G^2\beta^2}{r}\sum_{t,i}\mathbb{E}[\pi_{i,t}a_{i,t}^2] + \frac{12Q_T}{r^2}. \tag{19}$$

Proof. For $e_i = m_i - a_i$ and $\nu_i = Y_i - a_i$, write

$$X_i = (1 - \beta O_i)e_i + \beta O_i\nu_i, \qquad e_{i,t+1} = X_i - \delta a_{i,t}.$$

Conditional centering and independence of O_i and the new direction imply

$$\mathbb{E}_t X_i^2 = [1 - (2\beta - \beta^2)\pi_i]e_i^2 + \beta^2\pi_i\sigma_G^2 a_i^2 \le (1 - r)e_i^2 + \beta^2\pi_i\sigma_G^2 a_i^2.$$

Young's inequality with parameter $r/2$ gives

$$\mathbb{E}_t e_{i,t+1}^2 \le (1 - r/2)e_i^2 + \tfrac{3}{2}\beta^2\pi_i\sigma_G^2 a_i^2 + \frac{3}{r}\mathbb{E}_t(\delta a_{i,t})^2,$$

because $(1 + r/2)(1 - r) \le 1 - r/2$ and $1 + 2/r \le 3/r$. Summing for $t < T$ and over i, including the terminal contribution using $r/2 \le 1$, gives Eq. (19), in fact with 6 instead of 12 for the drift constant. We retain 12 to use common constants with the finite-difference extension. The argument never assumes that X_i and $\delta a_{i,t}$ are independent. □

For Algorithm 1, $\pi_i = p_i$ and $r = \beta\varepsilon/N$. Use $T^{-1}\sum_{t,i}\mathbb{E}[p_{i,t}a_{i,t}^2] \le A_T$. Adding the finite-difference observation remainder as in Appendix C gives, with $e_t = m_t - a_t$,

$$\sum_t \mathbb{E}\|e_t\|^2 \le \frac{2E_0}{r} + \frac{3\sigma_G^2\beta^2}{r}TA_T + \frac{12Q_T}{r^2} + O(T\mu^4). \tag{20}$$

A.6 Closing the drift and proving stationarity

All quantities are evaluated on the actual finite-difference trajectory. The perturbed sparse-EMA inequality and normalization give

$$\widetilde{\mathrm{Regret}}_T \le \frac{16E_0}{N\varepsilon r} + \frac{24\sigma_G^2\beta^2}{N\varepsilon r}TA_T + \frac{96}{N\varepsilon r^2}(Q_T + T\mathcal{B}_T) + 4\varepsilon T\widetilde{\mathcal{D}}_T.$$

Gradient continuity and the second-moment bound imply

$$Q_T \le NL_2^2\sum_{t<T}\mathbb{E}\|x_{t+1} - x_t\|^2 \le 2N\Lambda_d\eta^2(T\widetilde{\mathcal{D}}_T + \widetilde{\mathrm{Regret}}_T) + 2NL_2^2\eta^2 T\overline{W}_T/\varepsilon.$$

Substitution and absorption prove Theorem 1 after multiplication by $d + 2$, with additive remainder at most $192(d + 2)R_{\mu,T}/(1 - \kappa)$.

For descent, write $\widehat{g}_t = \widehat{g}_t^0 + h_t$ at the same parameters, selected index, and direction, and let $M_t = (d+2)\widetilde{V}_t(p_t)$. The Taylor bounds give

$$\|\mathbb{E}_t h_t\|^2 \le W_t, \qquad \mathbb{E}_t\|\widehat{g}_t\|^2 \le M_t + 2\sqrt{M_t W_t/\varepsilon} + W_t/\varepsilon.$$

Smoothness implies

$$\mathbb{E}_t F(x_{t+1}) \le F(x_t) - \eta\|\nabla F(x_t)\|^2 + \eta\|\nabla F(x_t)\|\sqrt{W_t} + \frac{L_F\eta^2}{2}\left(M_t + 2\sqrt{M_t W_t/\varepsilon} + W_t/\varepsilon\right).$$

Using $\Psi_T \le \widetilde{\mathcal{D}}_T$ and $T^{-1}\sum_t \mathbb{E}M_t \le (d+2)A_T/(N\varepsilon)$, Cauchy–Schwarz over time and randomness bounds the additional terms by $\Omega_{\mu,T}$. Summation proves Lemma 1 without multiplying its leading term by the coarse moment bound's factor two.

For Theorem 2, substitution of the schedules and $r = \beta\varepsilon/N$ gives, when $\kappa \le 1/2$,

$$\widetilde{\mathrm{Regret}}_T/T \le \frac{32E_0}{c_\varepsilon^2 c_\beta T^{1/2}} + \frac{384\widetilde{K}_T}{T^{1/10}} + 384R_{\mu,T}.$$

Insert this into the descent lemma. Holding $\widetilde{\mathcal{D}}_T$ fixed, the leading expression is minimized at $c_\eta = \sqrt{2\Delta_F/(C_d\widetilde{\mathcal{D}}_T)}$. Before collecting radius terms, the result is

$$\Psi_T \le \sqrt{\frac{2C_d\Delta_F\widetilde{\mathcal{D}}_T}{T}} + 192\left(\frac{C_d c_\eta\widetilde{K}_T}{T^{3/5}} + \frac{C_d c_\eta E_0}{c_\varepsilon^2 c_\beta T} + \frac{C_d c_\eta R_{\mu,T}}{\sqrt{T}} + \Omega_{\mu,T}\right).$$

This gives Eq. (12) after rescaling, with the radius-error convention stated in the main text. The general descent lemma covers undefined or infeasible oracle coefficients; the choice is conditional on the resulting trajectory quantity.

B Parameter balance and comparator qualifications

The radius-independent noise, drift, and exploration contributions have scales

$$\frac{\beta A_T}{\varepsilon^2}, \qquad \frac{N^2\Lambda_d\eta^2\widetilde{\mathcal{D}}_T}{\varepsilon^3\beta^2}, \qquad \varepsilon\widetilde{\mathcal{D}}_T.$$

For positive moments and smoothness, their formal balance is

$$\varepsilon \asymp N^{1/5}\Lambda_d^{1/10}\eta^{1/5}(A_T/\widetilde{\mathcal{D}}_T)^{1/5}, \tag{21}$$

$$\beta \asymp N^{3/5}\Lambda_d^{3/10}\eta^{3/5}(\widetilde{\mathcal{D}}_T/A_T)^{2/5}. \tag{22}$$

If these choices and the resulting trajectory satisfy feasibility and $\kappa \le 1/2$, then

$$\widetilde{\mathrm{Regret}}_T/T \lesssim N^{1/5}\Lambda_d^{1/10}\eta^{1/5}A_T^{1/5}\widetilde{\mathcal{D}}_T^{4/5} + \frac{E_0}{TN\sqrt{\Lambda_d}\eta} + O(R_{\mu,T}). \tag{23}$$

Only universal constants are suppressed. These choices require $A_T, \widetilde{\mathcal{D}}_T$, which depend on the parameters and resulting trajectory; this is a conditional balance, not an online tuning rule or an existence theorem for self-consistent parameters. The initialization term follows from $\varepsilon^2\beta \asymp N\sqrt{\Lambda_d}\eta$. The finite-difference correction remains additive.

Balancing noise and drift alone gives $\beta\varepsilon^{1/3}/\eta^{2/3} \asymp (N^2\Lambda_d\widetilde{\mathcal{D}}_T/A_T)^{1/3}$. A common numerical value across tasks is an empirical tuning constraint; the right side is problem-dependent.

B.1 Three fixed and dynamic costs

In addition to $\widetilde{\mathcal{H}}_T$ and $\widetilde{\mathcal{U}}_T$, define

$$\widetilde{\mathcal{G}}_T = \inf_{p\in\Delta_N}\frac{1}{T}\sum_t \mathbb{E}\widetilde{V}_t(p) = \left(\frac{1}{N}\sum_i\sqrt{\frac{1}{T}\sum_t \mathbb{E}a_{i,t}^2}\right)^2.$$

For each realized trajectory,

$$\inf_p \frac{1}{T}\sum_t \widetilde{V}_t(p) = \left(\frac{1}{N}\sum_i \sqrt{\frac{1}{T}\sum_t a_{i,t}^2}\right)^2.$$

Pointwise minimization, $\mathbb{E}\inf \le \inf \mathbb{E}$, and evaluation at the uniform distribution establish

$$\widetilde{\mathcal{D}}_T \le \widetilde{\mathcal{H}}_T \le \widetilde{\mathcal{G}}_T \le \widetilde{\mathcal{U}}_T. \tag{24}$$

$\widetilde{\mathcal{H}}_T$ is the expected hindsight cost, while $\widetilde{\mathcal{G}}_T$ chooses a fixed distribution before taking the trajectory expectation. Interchanging them would change the comparator. For $a_t = c_t v$ with fixed $v \ge 0$, the dynamic and hindsight fixed costs coincide. For the deterministic two-step example $a_1 = (1, 0)$, $a_2 = (0, 1)$, $\widetilde{\mathcal{D}}_T = 1/4$ while $\widetilde{\mathcal{H}}_T = \widetilde{\mathcal{G}}_T = \widetilde{\mathcal{U}}_T = 1/2$. This example illustrates a cost hierarchy and is not asserted to be a realized training trajectory.

Equation (14) follows by expanding the squared deviations around $\bar{a}_t$ and multiplying by $d + 2$. Consequently, $\widetilde{\mathcal{D}}_T < \widetilde{\mathcal{U}}_T$ whenever the example norms differ at some step with positive probability. The balanced expressions in Table 1 follow by minimizing $\Delta_F/c + C_d c X/2$ over $c > 0$ for each fixed positive normalized comparator quantity X. This is an algebraic comparison on the same trajectory; it does not require treating a hindsight-selected distribution as an implementable sampler.

C Finite-difference error

This appendix retains the original, unsmoothed gradients as the reference quantities. It covers Algorithm 1 and the with-replacement minibatch update in Appendix D.

The Hessian regularity is included in Assumption 1; set $H_2^2 = N^{-1}\sum_i H_i^2$. Let $D_6 = d(d+2)(d+4)$ and $D_8 = d(d+2)(d+4)(d+6)$. Taylor's integral remainder gives

$$b_{i,t} = s_{i,t}^{\mu_t} - z_t^\top g_{i,t}, \qquad |b_{i,t}| \le \frac{H_i \mu_t^2}{6}\|z_t\|^3.$$

For $\delta_{i,t} = \sqrt{\pi/2}(|s_{i,t}^{\mu_t}| - |z_t^\top g_{i,t}|)$, define

$$J_{i,t} = \frac{\pi H_i^2 D_6}{72}\mu_t^4, \qquad W_t = \frac{H_2^2 D_8}{36}\mu_t^4, \qquad \overline{W}_T = T^{-1}\sum_t W_t.$$

The reverse triangle inequality and Gaussian radial moments imply

$$\mathbb{E}_t \delta_{i,t}^2 \le J_{i,t}, \qquad \mathbb{E}_t\|b_{i,t} z_t\|^2 \le H_i^2 D_8 \mu_t^4/36. \tag{25}$$

In this appendix take $B = 1$ for Algorithm 1, or $B > 1$ for the experimental implementation. Let $\pi_{i,t}$ denote the corresponding inclusion probability and let $r > 0$ lower-bound $\beta\pi_{i,t}$ uniformly. Define

$$\chi_B = \frac{1}{B\varepsilon} + 1 - \frac{1}{B}, \quad A_{\pi,T} = \frac{1}{T}\sum_{t,i}\mathbb{E}[\pi_{i,t} a_{i,t}^2], \quad \mathcal{B}_T = \frac{1}{T}\sum_{t,i}\mathbb{E}[\beta^2 \pi_{i,t} J_{i,t}].$$

For the main algorithm, take $r = \beta\varepsilon/N$ and $\pi_i = p_i$. For the batch extension, $\pi_{i,t} = 1 - (1 - p_{i,t})^B$ and $r = \beta[1 - (1 - \varepsilon/N)^B]$.

Lemma 4 (Outer perturbation)**.** *Let $h_t^\mu = \widehat{g}_t^\mu - \widehat{g}_t^0$, where the two estimators use the same current parameters, indices, and direction, with finite or ideal differences. Then*

$$\|\mathbb{E}_t h_t^\mu\|^2 \le W_t, \qquad \mathbb{E}_t\|h_t^\mu\|^2 \le \chi_B W_t.$$

Proof. Conditional on z, the with-replacement sample average has expectation $N^{-1}\sum_i b_i z$. Jensen's inequality and $(N^{-1}\sum_i H_i)^2 \le H_2^2$ prove the first bound. For B independent with-replacement indices, the conditional second moment consists of $B^{-1}N^{-2}\sum_i \|b_i z\|^2/p_i$ plus $(1 - 1/B)\|N^{-1}\sum_i b_i z\|^2$. Use $p_i \ge \varepsilon/N$ and Eq. (25). □

Theorem 3 (Explicit finite-radius bound)**.** *Under Assumption 1, use the with-replacement update in Eq.* (28) *and the once-per-included-example EMA in Eq.* (29)*, with* $0 < \varepsilon \le 1/2$ *and* $0 < \beta < 1$*. Let*

$$\kappa_\mu = \frac{192\Lambda_d \eta^2}{B\varepsilon r^2} < 1.$$

Along the actual finite-difference trajectory, define $\widetilde{\mathrm{Regret}}_T, \widetilde{\mathcal{D}}_T$ *using the original gradient norms using the normalization above and Eq.* (10)*. Then*

$$\begin{aligned}\frac{\widetilde{\mathrm{Regret}}_T}{T} \le \frac{1}{1-\kappa_\mu}\Bigg[&\frac{16E_0}{TN\varepsilon r} + \frac{24\sigma_G^2\beta^2}{N\varepsilon r}A_{\pi,T} + \frac{96\mathcal{B}_T}{N\varepsilon r^2} \\ &+ \left(\frac{192\Lambda_d\eta^2}{\varepsilon r^2} + 4\varepsilon\right)\widetilde{\mathcal{D}}_T + \frac{192L_2^2\eta^2\chi_B}{\varepsilon r^2}\overline{W}_T\Bigg].\end{aligned} \tag{26}$$

If $C_d\eta(1-1/B) \le 1/4$*, then*

$$\Psi_T \le \frac{4\Delta_F}{\eta T} + \frac{4C_d\eta}{B}(\widetilde{\mathcal{D}}_T + \widetilde{\mathrm{Regret}}_T/T) + (2 + 4L_F\eta\chi_B)\overline{W}_T. \tag{27}$$

Proof. In Lemma 3, the new error is $X_i + \beta O_i\delta_i - \delta a_i$. Apply Young's inequality with parameter $r/2$, then bound the square of the last two terms by twice the sum of their squares. This gives

$$\sum_t \mathbb{E}\|m_t - a_t\|^2 \le \frac{2E_0}{r} + \frac{3\sigma_G^2\beta^2}{r}TA_{\pi,T} + \frac{12}{r^2}(T\mathcal{B}_T + Q_T).$$

No centering of δ_i is assumed. By the ideal batch second moment, Lemma 4, and the two-term square inequality,

$$\mathbb{E}_t\|\widehat{g}_t^\mu\|^2 \le 2(d+2)\left[\widetilde{V}_t(p_t)/B + (1-1/B)\|\nabla F(x_t)\|^2\right] + 2\chi_B W_t.$$

Consequently,

$$Q_T \le 2N\Lambda_d\eta^2(T\widetilde{\mathcal{D}}_T + \widetilde{\mathrm{Regret}}_T/B) + 2NL_2^2\eta^2\chi_B T\overline{W}_T.$$

Substitute into Lemma 2 and absorb the regret coefficient to prove Eq. (26). For descent, use $\mathbb{E}_t\widehat{g}_t^\mu = \nabla F(x_t) + b_t$ with $\|b_t\|^2 \le W_t$, and $-\langle\nabla F, b_t\rangle \le \|\nabla F\|^2/2 + W_t/2$. The coefficient of Ψ_T is at least $1/2 - C_d\eta(1-1/B) \ge 1/4$, giving Eq. (27). □

Radius dependence in the main theorem. For $M_{4,T} = T^{-1}\sum_t \mu_t^4$, the explicit bounds are

$$\mathcal{B}_T \le \frac{\pi N\beta^2 H_2^2 D_6}{72}M_{4,T}, \qquad \overline{W}_T = \frac{H_2^2 D_8}{36}M_{4,T}.$$

For the main algorithm, with fixed problem and schedule coefficients, bounded trajectory moments, and the horizon schedules in Theorem 2, one has $r = \Theta(T^{-2/5})$, $\chi_B = O(T^{1/10})$, and

$$R_{\mu,T} = O(T^{3/10}M_{4,T} + M_{4,T}), \qquad \Omega_{\mu,T} = O(\sqrt{M_{4,T}} + T^{-2/5}M_{4,T}).$$

The constants in these two order summaries depend on the fixed coefficients explicitly displayed above and in the main definitions. The radius-independent remainder in the main theorem hides only universal constants. Its separate $O(\mu^2)$ hides the radius coefficients displayed here. Under the stated fixed-parameter and bounded-moment conditions, $\eta R_{\mu,T} = O(T^{-1/5}\mu^4 + T^{-1/2}\mu^4)$ and $\Omega_{\mu,T} = O(\mu^2)$ uniformly in feasible T; thus the collected $O(\mu^2)$ is uniform in those horizons. Choosing $\mu_t = \mu_0 T^{-\alpha}$ for every step of a run gives $M_{4,T} = \mu_0^4 T^{-4\alpha}$. For $\alpha > 3/10$, both $\Omega_{\mu,T}$ and $\eta R_{\mu,T}$ are $o(T^{-3/5})$. This condition applies to a horizon-dependent radius and does not assert the same conclusion for $\mu_t = \mu_0 t^{-\alpha}$ with a fixed initial radius. The earlier coarse outer bound remains useful if one prefers squared-radius errors at the price of larger leading constants. Evaluation randomness and numerical rounding are not represented by these Taylor bounds.

D With-replacement minibatches and shared directions

All experiments use with-replacement minibatch sampling. This appendix specifies the implementation and derives the effect of sharing one Gaussian direction within a batch. Algorithm 1 and the main convergence results are its $B = 1$ case.

D.1 MINIBATCH UPDATE AND REUSED EVALUATIONS

At step t, form p_t from the stored scores and draw B indices independently with replacement, $I_{t,1}, \dots, I_{t,B} \sim p_t$, independently of $z_t \sim \mathcal{N}(0, I_d)$. Let

$$K_{i,t} = \sum_{b=1}^{B} \mathbf{1}\{I_{t,b} = i\}, \qquad S_t = \{i : K_{i,t} > 0\}.$$

For each $i \in S_t$, evaluate the two-sided difference only once, using the common perturbations $x_t \pm \mu_t z_t$. If an example is drawn repeatedly, its stored difference is reused: it does not trigger another forward evaluation or a replacement draw. The model update is

$$\widehat{g}_t = \frac{z_t}{B} \sum_{b=1}^{B} \frac{s_{I_{t,b},t}}{N p_{I_{t,b},t}} = \frac{z_t}{BN} \sum_{i \in S_t} \frac{K_{i,t} s_{i,t}}{p_{i,t}}, \qquad x_{t+1} = x_t - \eta \widehat{g}_t. \tag{28}$$

Thus evaluation reuse preserves the multiplicities of the original B draws. Each included example updates its score once:

$$m_{i,t+1} = \begin{cases} (1-\beta) m_{i,t} + \beta \sqrt{\dfrac{\pi}{2}} |s_{i,t}|, & i \in S_t, \\ m_{i,t}, & i \notin S_t. \end{cases} \tag{29}$$

The distribution remains fixed within a batch; the new scores determine p_{t+1}. No inverse-probability factor is applied to the EMA. The two perturbed batches contain $|S_t|$ examples each, requiring $2|S_t| \le 2B$ example-level function evaluations. The expected number of distinct examples is $\sum_i [1 - (1 - p_{i,t})^B]$.

D.2 EFFECT OF THE SHARED DIRECTION ON VARIANCE

For fixed x_t, p_t, z_t, let $\bar{s}_t = N^{-1} \sum_i s_{i,t}$. Independence of the index draws gives the exact finite-difference identities

$$\mathbb{E}_{I|z_t} \widehat{g}_t = \bar{s}_t z_t, \tag{30}$$

$$\mathbb{E}_{I|z_t} \|\widehat{g}_t\|^2 = \|z_t\|^2 \left[\frac{1}{BN^2} \sum_i \frac{s_{i,t}^2}{p_{i,t}} + \left(1 - \frac{1}{B}\right) \bar{s}_t^2 \right]. \tag{31}$$

Hence sharing a direction and reusing duplicate evaluations preserve the expected full-data finite-difference update. The differences for different examples need not be independent.

Write $\bar{g}_t = \nabla F(x_t)$. Under Assumption 1, integrating over the shared Gaussian direction yields

Proposition 2 (Shared-direction minibatch moments)**.** *For the update in Eq.* (28)*,*

$$\begin{aligned} \mathbb{E}_t \widehat{g}_t &= \bar{g}_t + O(\mu^2), \\ \mathbb{E}_t \|\widehat{g}_t\|^2 &= (d+2) \left[\frac{\widetilde{V}_t(p_t)}{B} + \left(1 - \frac{1}{B}\right) \|\bar{g}_t\|^2 \right] + O(\mu^2). \end{aligned} \tag{32}$$

Its total variance is

$$\mathbb{E}_t \|\widehat{g}_t - \mathbb{E}_t \widehat{g}_t\|^2 = (d+1) \|\bar{g}_t\|^2 + \frac{d+2}{B} \left[\widetilde{V}_t(p_t) - \|\bar{g}_t\|^2 \right] + O(\mu^2). \tag{33}$$

The radius terms follow the convention in Appendix C; the displayed leading terms are exact for directional derivatives.

Proof. The central-difference remainder satisfies $s_{i,t} = z_t^\top g_{i,t} + b_{i,t}$ with $|b_{i,t}| \le H_i \mu_t^2 \|z_t\|^3 / 6$. Applying this expansion directly to Eqs. (30)–(31), the Gaussian identities

$$\mathbb{E}[z z^\top] = I_d, \qquad \mathbb{E}[\|z\|^2 (z^\top v)^2] = (d+2) \|v\|^2$$

give the displayed leading terms. Cross terms involving $b_{i,t}$ are $O(\mu^2)$ by Cauchy–Schwarz and the finite-radius moment bounds; squared remainders are $O(\mu^4)$. Subtracting $\|\mathbb{E}_t \widehat{g}_t\|^2 = \|\bar{g}_t\|^2 + O(\mu^2)$ gives Eq. (33). □

Advantage over MeZO under the same shared direction. MeZO also shares one Gaussian direction across a minibatch. Its leading directional variance is therefore the same $(d+1)\|\bar{g}_t\|^2$ as SAZO's; the allocation controls the remaining term. At the same parameter vector and batch size, write $p^{\mathrm{unif}} = \mathbf{1}/N$. The variance reduction relative to MeZO is

$$\mathrm{Var}_t(\widehat{g}_t^{\mathrm{MeZO}}) - \mathrm{Var}_t(\widehat{g}_t^{\mathrm{SAZO}}) = \frac{d+2}{B}\left[\widetilde{V}_t(p^{\mathrm{unif}}) - \widetilde{V}_t(p_t)\right] + O(\mu^2), \tag{34}$$

where $\mathrm{Var}_t(X) = \mathbb{E}_t\|X - \mathbb{E}_t X\|^2$. For the norm-proportional oracle, the leading gain is

$$\frac{d+2}{B}\left[\widetilde{V}_t(p^{\mathrm{unif}}) - \bar{a}_t^2\right] = \frac{d+2}{BN}\sum_i (a_{i,t} - \bar{a}_t)^2. \tag{35}$$

Thus unequal gradient norms give a strictly positive leading variance reduction. The common direction term cancels from the comparison: sharing a direction does not introduce an additional variance penalty for SAZO relative to MeZO. For the learned allocation, the leading gain equals the oracle gain minus $(d+2)\widetilde{\rho}_t/B$, directly connecting the benefit to the tracking analysis.

Larger batches reduce both methods' sampling variance by $1/B$, while norm-proportional allocation minimizes this term for every batch size. SAZO exploits sample heterogeneity within the same two perturbed model states used by MeZO, and duplicate evaluation reuse reduces the work needed for its B draws.

D.3 Consequences for tracking and convergence

Because an included example updates its EMA once, its refresh event is $O_{i,t} = \mathbf{1}\{K_{i,t} > 0\}$. Conditional on the history, the probability that all B independent draws miss example i is $(1-p_{i,t})^B$. Thus its inclusion probability and the resulting uniform refresh bound are

$$\pi_{i,t} = 1 - (1-p_{i,t})^B, \qquad r_B = \beta\left[1 - \left(1 - \frac{\varepsilon}{N}\right)^B\right]. \tag{36}$$

The floor $p_{i,t} \geq \varepsilon/N$ gives $\beta\pi_{i,t} \geq r_B$. Relative to the $B = 1$ rate $r = \beta\varepsilon/N$, one has $r \leq r_B \leq \min\{Br, \beta\}$ and $r_B = Br[1 + O(B\varepsilon/N)]$ when $B\varepsilon/N$ is small.

To connect inclusion to tracking contraction, set $e_{i,t} = m_{i,t} - a_{i,t}$ and use the leading Gaussian observation $Y_{i,t}^0$ from Appendix A. Independence of the index draws and the fresh direction gives

$$\mathbb{E}_t\left[\left(e_{i,t} + \beta O_{i,t}(Y_{i,t}^0 - m_{i,t})\right)^2\right] \leq (1 - r_B)e_{i,t}^2 + \beta^2\pi_{i,t}\sigma_G^2 a_{i,t}^2,$$

since $(2\beta - \beta^2)\pi_{i,t} \geq \beta\pi_{i,t} \geq r_B$. Sharing z_t correlates the observations across examples, but Lemma 3 controls a sum of coordinatewise squared errors and does not require independence across examples. Its argument therefore applies with $r = r_B$; the moving target and finite-difference remainder are included in Theorem 3. Each included example receives only one directional observation, and updating its EMA once avoids treating duplicate draws as fresh evidence.

To see how the shared-direction term enters optimization, average Eq. (32) in the smoothness descent inequality. With $C_d = L_F(d+2)$ and $\Psi_T = T^{-1}\sum_t \mathbb{E}\|\nabla F(x_t)\|^2$, this gives

$$\left[1 - \frac{C_d\eta}{2}\left(1 - \frac{1}{B}\right)\right]\Psi_T \leq \frac{\Delta_F}{\eta T} + \frac{C_d\eta}{2B}\left(\widetilde{\mathcal{D}}_T + \frac{\widetilde{\mathrm{Regret}}_T}{T}\right) + O(\mu^2). \tag{37}$$

For $C_d\eta(1 - 1/B) < 2$, the coefficient on the left is positive. This coefficient is common to SAZO and a uniform shared-direction update. Along a given parameter trajectory, uniform sampling contributes $\widetilde{\mathcal{U}}_T/B$ to the right-hand sampling term, whereas SAZO contributes $(\widetilde{\mathcal{D}}_T + \widetilde{\mathrm{Regret}}_T/T)/B$. The same dynamic-oracle improvement established in the main analysis therefore persists in the minibatch bound. Setting $B = 1$ recovers the main descent lemma.

The same distinction matters for target drift. The coarse moment bound with an $O(\mu^4)$ remainder gives

$$Q_T \leq 2N\Lambda_d\eta^2\left(T\widetilde{\mathcal{D}}_T + \frac{\widetilde{\mathrm{Regret}}_T}{B}\right) + O(T\mu^4),$$

using $\|\bar{g}_t\|^2 \le \bar{a}_t^2$. Combining this inequality with the $12Q_T/r_B^2$ term in Lemma 3 and the factor $8/(N\varepsilon)$ in Lemma 2, the coefficient of regret on the right-hand side is

$$\kappa_B = \frac{192\Lambda_d\eta^2}{B\varepsilon r_B^2}, \qquad \frac{\kappa_B}{\kappa} = \frac{1}{B}\left[\frac{\varepsilon/N}{1-(1-\varepsilon/N)^B}\right]^2. \tag{38}$$

Here κ is the $B = 1$ coefficient at the same η, ε, β. Thus $\kappa_B/\kappa = B^{-3}[1 + O(B\varepsilon/N)]$ when $B\varepsilon/N$ is small: sampling reduces the regret-dependent update moment by $1/B$, and the increased refresh rate reduces its tracking contribution by a further factor of approximately B^2. Theorem 3, with $r = r_B$ and $\kappa_\mu = \kappa_B$, combines these effects with the remaining noise, exploration, and finite-difference terms. Its $B = 1$ specialization recovers the main convergence analysis.

E Experimental settings

E.1 Detailed experiments on RoBERTa-large

We use the hyperparameters in Table 5 for SAZO experiments on RoBERTa-large. Regarding learning-rate scheduling and early stopping, we use a constant learning rate for all SAZO experiments. All SAZO experiments use $\beta = 0.05$, $\varepsilon = 0.1$, and four burn-in passes. For Figure 2(a), we select the optimal constant learning rate for each method from $\{1, 5\} \times \{10^{-7}, 10^{-6}, 10^{-5}\}$; all other settings match the corresponding full-parameter experiments in Table 2.

Table 5: Hyperparameter grids used for SAZO experiments on RoBERTa-large.

Experiment	Hyperparameter	Values
SAZO	Batch size	64
	Learning rate	$\{10^{-7}, 10^{-6}, 10^{-5}\}$
	Perturbation radius μ	10^{-3}
	Weight decay	0
SAZO (prefix)	Batch size	64
	Learning rate	$\{10^{-2}, 5\times10^{-3}, 10^{-3}\}$
	Perturbation radius μ	10^{-1}
	Weight decay	0
	Prefix tokens	5
SAZO (LoRA)	Batch size	64
	Learning rate	$\{10^{-5}, 5\times10^{-5}, 10^{-4}\}$
	Perturbation radius μ	10^{-3}
	Weight decay	0.1
	(r, α)	$(8, 16)$

E.2 Detailed results on various LLMs

We use the hyperparameters in Table 6 for SAZO experiments on GLMs. For Figure 2(b,c), we select the optimal constant learning rate for each method and model from $\{1, 5\} \times \{10^{-7}, 10^{-6}, 10^{-5}\}$; all other settings match the corresponding experiments in Table 3.

Table 6: Hyperparameter grids used for GLM experiments. All weight decay is set to 0. SAZO uses 20K steps and constant learning rates.

Experiment	Hyperparameter	Values
SAZO	Batch size	16
	Learning rate	$\{10^{-6}, 5\times10^{-7}, 10^{-7}\}$
	Perturbation radius μ	10^{-3}

E.3 Training-population scaling settings

Figure 3 uses full-parameter RoBERTa-large on SST-2 with training populations $N = 1{,}024$, $2N = 2{,}048$, $3N = 3{,}072$, and $4N = 4{,}096$. MeZO and SAZO use batch size 16, constant learning rate 10^{-6}, and 100K optimization steps. All learning-rate and checkpoint choices are fixed or validation-only; the test set is never used for selection. We report test accuracy averaged over seeds 13, 21, and 42, with population standard deviations as error bars.

E.4 Ablation experiment details

The ablations use one seed of full-parameter RoBERTa-large on SST-2 with $k = 512$, a fixed 1,024-example training population, batch size 64, constant learning rate 3×10^{-6}, perturbation radius 10^{-3}, and zero weight decay. We initialize the importance scores by averaging directional magnitudes over four passes, then use sampling with replacement and inverse-probability-weighted parameter updates. The β sweep in Figure 5 uses $\beta \in \{0.005, 0.05, 0.5, 1\}$ at fixed $\varepsilon = 0.1$. The exploration sweep uses $\varepsilon \in \{0.05, 0.1, 0.5, 1\}$ at fixed $\beta = 0.05$. Each plotted point is a forward-only mean-loss audit over the complete training population; audit queries are excluded from the update count. We retain every measured audit and apply neither curve smoothing nor extrapolation.

Figure 6 shows stable optimization across $\varepsilon \in \{0.05, 0.1, 0.5\}$, indicating that SAZO is robust to substantial changes in the exploration rate. At the extreme $\varepsilon = 1$, the adaptive component vanishes and sampling becomes uniform; this trajectory eventually diverges under the same learning rate.

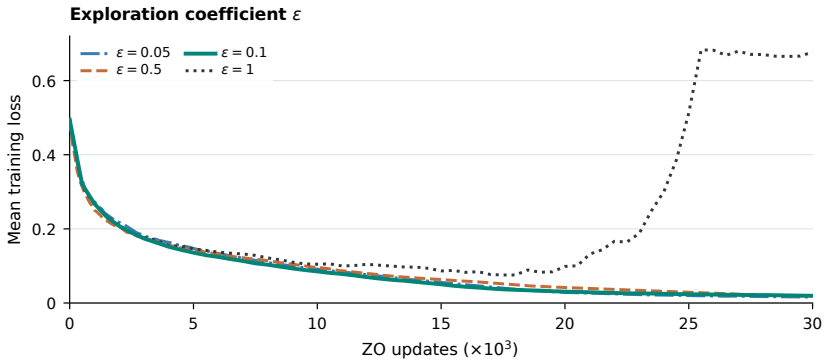

Figure 6: Exploration-coefficient ablation on SST-2. SAZO remains stable over $\varepsilon \in \{0.05, 0.1, 0.5\}$, whereas $\varepsilon = 1$, which reduces to uniform sampling, becomes unstable.

E.5 Sample-importance diagnostics

The heterogeneity experiments use full-parameter RoBERTa-large under uniform MeZO training, with $k = 512$, batch size 64, learning rate 10^{-6}, perturbation radius 10^{-3}, and zero weight decay. At frozen checkpoints, we compute exact per-example gradient norms over the trainable parameters. These gradients are used only for diagnostics. The oracle allocation is $p^{\star}_{i,t} = a_{i,t}/\sum_j a_{j,t}$, and relative importance is $Np^{\star}_{i,t}$. Top-10% sets are recomputed at each checkpoint; temporal changes are measured using Spearman rank correlation. For the SST-2 examples, the prompt appends `It was <mask>.`, with `great` and `terrible` as the positive and negative verbalizers.

Figure 1(a) reports the exact audits at initialization, 10K, and 30K updates. We divide all three checkpoints by the median initial norm and fix the horizontal order using the initial norms, so a given position refers to the same training example at every checkpoint. Every example is shown as a point, without smoothing or aggregation.

Length and label controls. We group examples by label or by the joint pair of prompt token count and label, and optimize the allocation subject to equal probabilities within each group. For a

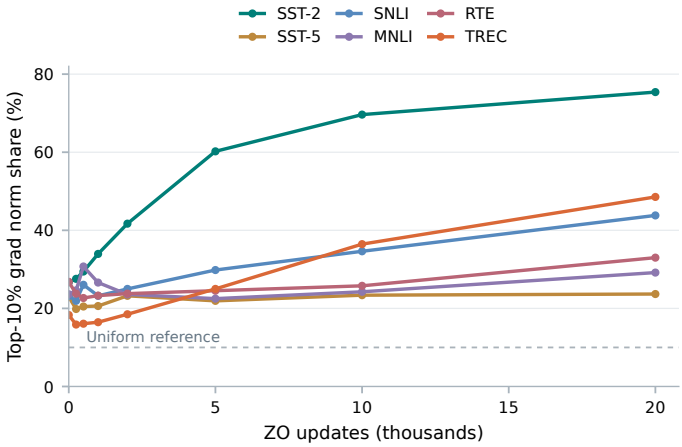

Figure 7: **Gradient-norm concentration across tasks.** The top 10% of examples account for an increasing share of the total gradient norm across six tasks. The top set is recomputed at each checkpoint; the dashed line marks the approximately 10% uniform reference. Exact gradients are measured at frozen RoBERTa-large MeZO checkpoints.

partition $\mathcal{G}$, the optimal group-restricted distribution is

$$b_{G,t} = \left(\frac{1}{|G|} \sum_{i \in G} a_{i,t}^2 \right)^{1/2}, \qquad p_{i,t}^{\mathcal{G}} = \frac{b_{G(i),t}}{\sum_{G \in \mathcal{G}} |G| b_{G,t}}.$$

The fraction of the uniform-to-oracle variance gap remaining is $[\widetilde{V}_t(p_t^{\mathcal{G}}) - \widetilde{V}_t(p_t^\star)]/[\widetilde{V}_t(\mathbf{1}/N) - \widetilde{V}_t(p_t^\star)]$, evaluated at 20K updates.

Online versus fixed sampling. The SST-2 comparison uses a common initial model, batch size 64, perturbation radius 10^{-3}, and 30K updates. We scan constant learning rates in $\{10^{-6}, 3 \times 10^{-6}\}$. SAZO uses $\beta \in \{0.025, 0.05, 0.075\}$ and $\varepsilon = 0.1$. The fixed sampler retains the initial exact gradient-norm-proportional distribution $p_0^\star$ without an exploration mixture; uniform sampling uses $1/N$. At each audit, the normalized instantaneous regret is

$$r_t = \frac{\widetilde{V}_t(p_t) - \widetilde{V}_t(p_t^\star)}{\widetilde{V}_t(p_t^\star)}. \tag{39}$$

Figure 1(b) reports relative cumulative regret,

$$\frac{\int_0^T [\widetilde{V}_t(p_t) - \widetilde{V}_t(p_t^\star)]\, dt}{\int_0^T \widetilde{V}_t(p_t^\star)\, dt}. \tag{40}$$

We integrate the numerator and denominator separately using the trapezoidal rule on all 61 recorded audits, spaced 500 updates apart from 0 to 30K, and then take their ratio. This differs from integrating r_t. The undefined ratio at $T = 0$ is omitted from the plot. Exact gradients at each method's current parameters define its reference oracle. Loss is evaluated over the complete training population.

E.6 Interpretable sample-level structure

Figures 8 and 9 connect population-level changes in sample importance to the model's predictions on individual inputs. Importance is measured by the normalized exact gradient norm $N\|\nabla f_i(x_t)\| / \sum_j \|\nabla f_j(x_t)\|$ at frozen RoBERTa-large MeZO checkpoints; a value of 1 corresponds to uniform allocation. These are diagnostic oracle quantities, rather than SAZO's estimated scores. Detailed audit settings are given in Appendix E.5.

Sample rankings change during training. Figure 8 compares each example's importance at initialization and after 20K updates on SST-2 and TREC. Points above the diagonal gain relative importance, while those below lose it. The initial-to-20K Spearman correlations are 0.52 and 0.12, respectively, showing that the initial ranking does not reliably preserve subsequent allocation priorities.

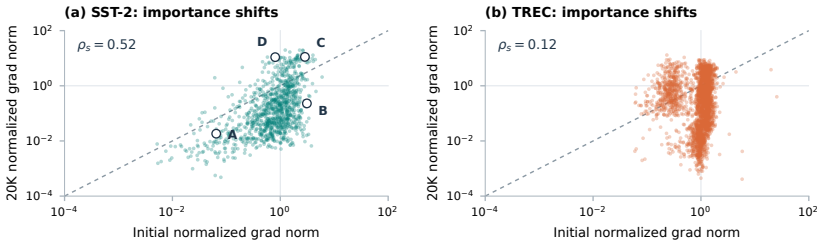

Figure 8: **Sample-importance rankings evolve during training.** Initial versus 20K normalized gradient norms $N\|\nabla f_i(x_t)\| / \sum_j \|\nabla f_j(x_t)\|$ on SST-2 and TREC, with one point per example and shared logarithmic axes. The diagonal marks unchanged importance, and ρ_s is the Spearman correlation. A–D identify the SST-2 examples examined in Figure 9. All gradient norms are exact audits at frozen MeZO checkpoints.

Importance changes have interpretable sample-level structure. The four SST-2 cases A–D in Figure 9 link the scatter plot to predictions at initialization, 5K, and 20K updates. A ("a great cast") stays correct and has low importance throughout. B ("over-dramatic at times") becomes correct as its normalized importance falls from 3.113 to 0.230. C ("smug grin") remains incorrect, while its importance rises from 2.863 to 11.170. D, containing both "amateurish" and "a minor treat," becomes incorrect and rises from 0.805 to 10.930. For D, the absolute gradient norm also increases, so the rise is not solely due to other examples' shrinking gradients.



Figure 9: **Representative examples reveal interpretable importance changes.** A–D correspond to the SST-2 points marked in Figure 8. Robot lights and status labels indicate correctness at each checkpoint. Numbers above the blue bars give the correct-class probability $\exp(-f_i(x_t))$; numbers below give normalized gradient norms. Bars share a linear scale from 0 to 12, with 1 corresponding to uniform allocation. All gradient norms are exact audits at frozen MeZO checkpoints. The input template appends `It was <mask>.` Appendix E.5 provides detailed settings and sample identities.

Together, the population shifts and these illustrative cases show that sample importance reflects the model's evolving response to an input. Words, length, and labels remain fixed, yet predictions and gradient contributions change. An allocation based only on initial importance can therefore become stale, motivating continued tracking during optimization.