ZERO-ORDER OPTIMIZATION VIA LEARNABLE DIRECTION SAMPLING FOR LLM FINE-TUNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Zeroth-order (ZO) optimization estimates gradients through function evaluations along sampled perturbation directions. Standard isotropic sampling yields weak expected gradient alignment, limiting estimator quality and introducing dimension-dependent convergence rates. We propose a plug-and-play framework that learns the perturbation distribution to improve directional alignment. Our design follows from a descent bound that identifies expected gradient alignment as the objective for adapting the mean of a Gaussian sampling distribution. For an idealized directional algorithm with exact policy updates, we characterize the evolution of alignment and derive convergence guarantees with no explicit dimension dependence. Building on this formulation, we develop a practical finite-difference algorithm that updates the sampling policy using REINFORCE estimates and integrates with existing ZO optimizers. The resulting framework is particularly relevant to large language model (LLM) fine-tuning, where avoiding backpropagation reduces training memory requirements. Experiments demonstrate consistent improvements across models, ZO optimizers, and fine-tuning schemes under matched forward-evaluation budgets, supporting learned direction sampling as an effective mechanism for improving ZO fine-tuning. Our code is available at `https://anonymous.4open.science/r/zo_ldsd`.

1 INTRODUCTION

Zeroth-order (ZO) methods iteratively solve optimization problems using only function evaluations, making them applicable when gradients are unavailable or costly to compute (Ghadimi & Lan, 2013; Duchi et al., 2015; Nesterov & Spokoiny, 2017; Liu et al., 2020). These methods have attracted renewed interest for memory-efficient fine-tuning of large language models (LLMs), where forward-only parameter updates avoid the memory overhead of backpropagation (Malladi et al., 2023; Zhang et al., 2024). Fine-tuning adapts pretrained models to diverse downstream tasks, often with limited labelled data (Howard & Ruder, 2018; Zhang et al., 2019; Lester et al., 2021). The fine-tuning objective is commonly formulated as an unconstrained optimization problem:

$$\min_{x \in \mathbb{R}^d} f(x), \tag{1}$$

where x are model parameters, d is their dimension and $f(x)$ is the loss function.

Gradient-based methods such as SGD Robbins & Monro (1951); Amari (1993) and Adam Kingma & Ba (2014) dominate due to efficiency and compatibility with backpropagation Semenov et al. (2025); Wen et al. (2025). However, backpropagation incurs a large memory footprint: storing activations and optimizer states can exceed inference memory by $\sim 12\times$ at scale (Malladi et al., 2023), limiting deployment on constrained hardware and complicating distributed training (Zhu et al., 2023; Gao et al., 2024; Liao et al., 2024).

Memory-efficient alternatives include coordinate-wise and block-coordinate Richtárik & Takáč (2014) methods, which reduce per-step storage but slow convergence by a factor of d Li et al. (2018). For example, BAdam Luo et al. (2024) trades memory for slower convergence proportional to the number of blocks.

Zero-order (ZO) methods offer a different trade-off: they approximate gradients using only forward evaluations, avoiding backpropagation and further reducing memory (Malladi et al., 2023; Zhang

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.

et al., 2024). The first recognition of this idea in the context of LLM fine-tuning came with the MeZO (Malladi et al., 2023) method, where the authors employed a ZO SGD (Ghadimi & Lan, 2013) update based on the following gradient estimator:

$$v \cdot \langle \nabla f(x), v \rangle \approx v \cdot \frac{f(x + \tau v) - f(x - \tau v)}{2\tau}, \tag{2}$$

where the random direction $v \in \mathbb{R}^d$ is typically sampled from $\mathcal{N}(0, 1)$ Nesterov & Spokoiny (2017), and $\tau > 0$ is the smoothing parameter. Estimators of the form (2) constitute a standard design choice for ZO schemes and are widely used in the modern literature Chen et al. (2017); Malladi et al. (2023); Petrov et al. (2025). Moreover, in the limit $\tau \to 0$ the method recovers directional SGD.

The performance of optimization methods is fundamentally governed by the quality of the gradient estimator g^t, which is directly tied to its variance $\mathbb{E}\|g^t - \nabla f(x^t)\|^2$. A well-known source of variance arises from stochastic minibatching, which has been extensively studied and effectively mitigated by variance reduction (VR) techniques (Defazio et al., 2014; Li et al., 2021), including in the ZO setting (Gautam et al., 2024). However, ZO methods are subject to an additional, intrinsic source of variance stemming from the randomness of the search direction v. Consequently, they inherit a principal drawback: the effective convergence rate degrades by a factor of d Zhan et al. (2024). Moreover, as the dimensionality d increases, the probability that a random direction forms a large angle with the true gradient grows, further deteriorating the accuracy of estimation. Consequently, until recently, the use of ZO methods was mostly restricted to low-dimensional settings Wang et al. (2018). However, in light of modern applications in LLM fine-tuning, where the dimensionality d reaches 10^{10} Seung et al. (2025), this limitation has become even more pronounced.

This work aims to mitigate the restrictive role of direction-induced variance in ZO schemes. We view the ZO optimization process through the lens of Reinforcement Learning (RL), interpreting the distribution mean of the direction v as a policy. By introducing a plug-and-play extension to the ZO framework, we enable a memory-efficient exploitation of information from previous iterations to improve the quality of gradient approximation. Our goal is both to provide theoretical insight and to empirically validate the proposed approach in real-world application settings.

Our contributions are as follows:

• A plug-and-play sampling framework for directional and ZO methods that learns to avoid uninformative directions.

• A theory of expected gradient-alignment dynamics under smooth non-convex objectives, including convergence guarantees with no explicit d dependence. Novel design leads to non-standard challenges, theoretical analysis and bounds.

• Evaluation on challenging LLM fine-tuning benchmarks, demonstrating effectiveness and practical relevance.

2 Brief Related Work

Standard ZO approaches differ primarily in the choice of sampling distribution, including spherical (Flaxman et al., 2004), Gaussian (Ghadimi & Lan, 2013; Nesterov & Spokoiny, 2017), and coordinate-wise (Duchi et al., 2015) sampling. Practical improvements include adaptive and momentum-based variants such as ZO-AdaMM (Chen et al., 2019) and JAGUAR (Veprikov et al., 2024; Petrov et al., 2025). Recent work has also explored curvature-aware methods: HiZOO (Zhao et al., 2024) incorporates Hessian-informed preconditioning, while LOREN (Seung et al., 2025) learns low-rank anisotropic perturbations within a natural evolution strategies framework. A complementary line of work connects ZO optimization to reinforcement learning: Qiu et al. (2025) establish an equivalence with REINFORCE (Williams, 1992). Related neural approaches include ZO-LSTM (Ruan et al., 2019), which learns ZO update rules and Gaussian sampling covariances using recurrent networks, and ZO-RL (Zhai et al., 2024), which generates perturbation directions using an actor-critic policy. However, ZO-LSTM and ZO-RL rely on flexible neural parameterizations of the search mechanism that lack theoretical grounding and convergence guarantees. Moreover, the additional controller state and training costs preclude their direct application to large-scale tasks such as LLM fine-tuning. In contrast, our design is motivated by a descent bound that identifies the policy objective. This yields convergence guarantees and scalability to large-scale tasks.

A more detailed discussion of related work is deferred to Appendix A.

3 Algorithm and Theoretical Analysis under Directional Setup

3.1 Notation and Assumptions

For any vector $x \in \mathbb{R}^d$ we denote by $\|x\| = \sqrt{\sum_{i=1}^d x_i^2}$ the Euclidean l_2 norm. For any non-zero $x \neq 0$ we define its normalized version as $\overline{x} = \frac{x}{\|x\|}$. To keep the main-text statements compact, for nonnegative quantities $g(x)$ and $f(x)$ we write $g(x) \leq \mathcal{O}(f(x))$ as shorthand for $g(x) \leq cf(x)$, where $c > 0$ is an absolute numerical constant independent of the problem parameters. The corresponding explicit constants are provided in Appendix C. We conduct the theoretical analysis in a non-convex setting under the standard assumptions (Bertsekas, 1997; Ghadimi & Lan, 2013; Richtárik et al., 2021; Li et al., 2021).

Assumption 3.1. Let f be L-smooth, i.e., for all $x_1, x_2 \in \mathbb{R}^d$ it holds that: $\|\nabla f(x_1) - \nabla f(x_2)\| \leq L\|x_1 - x_2\|$.

Assumption 3.2. Let there exist a minimizer of f i.e., an x^* that $f(x) \geq f(x^*)$ holds for all $x \in \mathbb{R}^d$.

3.2 Motivation

We consider solving the problem (1) using optimization along random directions, i.e., via a Directional Gradient Descent (DGD) (Silver et al., 2021):

$$x^{t+1} = x^t - \gamma_x^t \cdot \overline{v^t}\langle \overline{v^t}, \nabla f(x^t)\rangle, \tag{3}$$

where the sampling distribution of v^t is arbitrary. We note that ZO oracle (2) reduces to DGD estimator under the $\tau \to 0$ condition. It makes the role of the direction sampling distribution identical in both formulations and allows us to focus on the DGD setup in the theoretical analysis.

Intuitively, the quality of the estimator improves as the sampled direction becomes better aligned with the gradient, and it attains its maximum accuracy in the case of a deterministic choice of v^t collinear with $\nabla f(x^t)$. To formalize this intuition, we introduce the notion of *gradient alignment*:

$$C^t := \left\langle \overline{v^t}, \overline{\nabla f(x^t)} \right\rangle^2, \tag{4}$$

and use it in Lemma 3.3, inspired by (Cheng et al., 2021).

Lemma 3.3. *Under Assumptions 3.1 and 3.2 for iterations of DGD* (3) *the following inequality holds:*

$$\sum_{t=0}^{T-1} \frac{1}{T}\mathbb{E}\left[\left(\gamma_x^t - \frac{L{\gamma_x^t}^2}{2}\right)\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]\|\nabla f(x^t)\|^2\right] \leq \frac{f(x^0) - f(x^*)}{T},$$

where $\mathcal{F}^t$ is a σ-algebra generated by $\{x^0, x^1, \ldots, x^t\}$.

One of its key features is that Lemma 3.3 applies to an arbitrary sampling distribution. Crucially for our analysis, it does not rely on the unbiasedness of the oracle, in contrast to commonly used approaches (Nesterov & Spokoiny, 2017).

The simplest choice of the sampling distribution is $v^t \sim \mathcal{N}(\mu^t \equiv 0, \varepsilon^2 I)$: a Gaussian distribution with zero mean, which yields an unbiased estimator in the DGD setting. After normalizing the sampled vector, this yields a uniform distribution over the unit sphere $\overline{v^t} \sim RS(1)^d_{\|\cdot\|}$. Notably, by leveraging Lemma 3.3, we recover the convergence rate for DGD with $v^t \overset{i.i.d.}{\sim} \mathcal{N}(0, I)$ established in (Nesterov & Spokoiny, 2017), as stated in Corollary 3.4.

Corollary 3.4. *Let in the setting of Lemma 3.3 $v^t \overset{i.i.d.}{\sim} \mathcal{N}(\mu^t \equiv 0, \varepsilon^2 I)$ and $\gamma_x^t \leq \frac{1}{2L}$. Then the following bound holds:*

$$\mathbb{E}\|\nabla f(\overline{x}^T)\|^2 \leq \frac{2d}{\gamma_x^t T}(f(x^0) - f(x^*)),$$

where $\overline{x}^T$ is chosen uniformly from $\{x^0, x^1, \ldots, x^{T-1}\}$.

The proof of Corollary 3.4 is provided in Appendix C and directly reveals that the d factor arises from $\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right] \sim 1/d$ for $v^t \overset{i.i.d.}{\sim} \mathcal{N}(0, I)$.

The above intuition shows that, on the one hand, Lemma 3.3 indeed captures the variance effect induced by the randomness of the direction, and on the other hand, its application yields tight bounds. This makes expected gradient alignment $\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]$ a key concept in the analysis of the DGD setup. To obtain theoretical acceleration, it is necessary to maintain this quantity at an $\mathcal{O}(1)$ level, in contrast to the baseline $\mathcal{O}(1/d)$ regime. To this end, we propose to leverage information about the objective function obtained during previous iterations by learning the parameters of the sampling distribution and introducing control over them to increase the value of $\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]$.

Using a Gaussian distribution gives rise to two design choices: adapting the mean vector μ^t or the covariance matrix Σ^t. We focus on learning μ^t and fix $\Sigma^t \equiv \varepsilon^2 I$, since in the context of memory-efficient optimization storing a full covariance matrix is infeasible, whereas learning the mean vector incurs only an additional $\mathcal{O}(d)$ memory overhead. Moreover, the geometry of the problem provides a further justification for this parameterization: the central geometrical quantity is gradient alignment C^t, defined in (4), which depends only on the angle between the true gradient $\nabla f(x^t)$ and the sampled direction v^t. At the same time, parameterizing the Gaussian distribution through its mean μ^t, with fixed covariance $\varepsilon^2 I$, provides two essential properties:

(i) it allows the maximum of the induced probability density on the unit sphere $\mathbb{S}^{d-1}$ to be placed in any direction, while the concentration around that direction can be controlled through the policy norm $\|\mu^t\|$.

(ii) it assigns equal probability density to directions at the same angular distance from the current policy μ^t.

Thus, mean parameterization is fully expressive in choosing the preferred direction and its concentration, while preserving the rotational invariances inherent to the problem geometry defined by the gradient alignment.

3.3 Theoretical Algorithm

The above considerations, along with Lemma 3.3, motivate the design of Algorithm 1: LDSD (**L**earnable **D**irection **S**ampling **D**escent).

Algorithm 1 LDSD

1: **Input:** $x^0, \mu^0, T, \varepsilon$.
2: **for** $t = 0, 1, 2, \ldots, T-1$ **do**
3: Sample $v^t \sim \mathcal{N}(\mu^t, \varepsilon^2 I)$
4: Set $\overline{v^t} = v^t / \|v^t\|$
5: $g_x^t = \overline{v^t} \left\langle \overline{v^t}, \nabla f(x^t) \right\rangle$
6: $g_\mu^t = \nabla_\mu \mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]\Big|_{\mu=\mu^t}$
7: $\mu^{t+1} = \mu^t + \gamma_\mu^t g_\mu^t$
8: $x^{t+1} = x^t - \gamma_x^t g_x^t$
9: **end for**

Lines 3, 5, 8 of Algorithm 1 reflect the classical use of DGD estimator. In contrast, Lines 6 and 7 incorporate the proposed mechanism for learning the sampling direction. Specifically, to attain the values $\mathbb{E}\left[C^t|\,\mathcal{F}^{t-1}\right]$ of expected gradient alignment as large as possible, we interpret the mean vector μ^t as a policy and $\mathbb{E}\left[C^t|\,\mathcal{F}^{t-1}\right]$ as the reward signal. Accordingly, at Line 3 of Algorithm 1, a direction v^t is sampled from the current policy, while at Lines 6 and 7 the policy is updated via a gradient ascent step.

Importantly, in practice the derivative of the expected gradient alignment $\nabla_\mu \mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]\Big|_{\mu=\mu^t}$ is not accessible. Consequently, while the analysis employs the update in Line 6, for practical implementation we propose using a well-established policy gradient approach (Williams, 1992). We refer to Sections 3.6 and 4 for further details on this implementation.

3.4 The dynamics of gradient alignment

Now we present Theorem 3.5, which characterizes the evolution of the expected gradient alignment $\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]$ over a single iteration of Algorithm 1 and constitutes the core of the theoretical analysis of the proposed approach.

Theorem 3.5. *Let Assumption 3.1 hold and $d \geq 16$. Then for iterations of Algorithm 1 with $\varepsilon \leq \mathcal{O}\left(d^{-\frac{3}{2}} \cos(\beta^t)(1 - \cos(\beta^t))\|\mu^t\|\right), \gamma_\mu^t \leq \mathcal{O}\left(\|\mu^t\|^2\right), \gamma_x^t \leq \mathcal{O}\left(\frac{\gamma_\mu^t \cos^2(\beta^t)(1-\cos(\beta^t))^2}{L\|\mu^t\|^2}\right)$ the*

following inequality holds:

$$\mathbb{E}\left[C^{t+1}\middle|\,\mathcal{F}^t\right] \geq \mathbb{E}\left[C^t\middle|\,\mathcal{F}^{t-1}\right] + \gamma_\mu^t \left(\frac{\cos(\beta^t)(1-\cos(\beta^t))}{24\|\mu^t\|}\right)^2,$$

where β^t is the angle between μ^t and $\nabla f(x^t)$.

Under the assumptions of Theorem 3.5, the expected gradient alignment $\mathbb{E}\left[C^t \mid \mathcal{F}^{t-1}\right]$ increases monotonically, improving the policy μ^t and strengthening the descent guarantee in Lemma 3.3. Two degenerate configurations require attention. First, $\mu^0 = 0$ is a saddle point of the alignment objective (Figure 1), where $\nabla_\mu \mathbb{E}\left[C^t \mid \mathcal{F}^{t-1}\right] = 0$; hence, Line 7 leaves the policy, and therefore the expected alignment, unchanged. Second, the objective gradient also vanishes when μ^t is collinear with or orthogonal to $\nabla f(x^t)$, so strict improvement is not guaranteed in these cases. These configurations form a measure-zero set and are therefore avoided almost surely.

Geometric interpretation. Figure 1 visualizes the saddle-like alignment landscape in dimension $d = 2$. As μ^t becomes orthogonal to $\nabla f(x^t)$, sampled directions tend to form large angles with the gradient, producing small values of C^t and its conditional expectation; alignment produces larger values. The landscape is symmetric under $\mu^t \mapsto -\mu^t$ as C^t depends only on the squared cosine of the angle β^t.

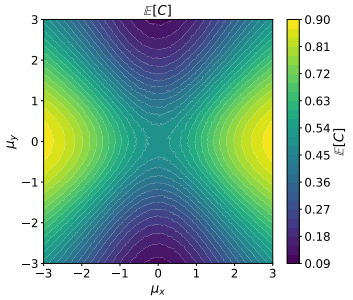


Figure 1: Landscape of the function $\mathbb{E}\left[C^t \middle| \mathcal{F}^{t-1}\right]$ with respect to μ^t, for $\nabla f(x^t) = (1, 0)^\top$.

Each iteration of Algorithm 1 first performs gradient ascent in μ^t, corresponding in the RL interpretation to a policy update with expected gradient alignment as the reward objective. The subsequent model update in Line 8 changes the direction of $\nabla f(x^t)$, geometrically rotating the alignment landscape while μ^t remains fixed. Assumption 3.1 bounds this rotation in terms of γ_x^t; equivalently, it bounds the change in the environment and the loss of relevance of the current policy. Theorem 3.5 balances this drift against the improvement produced by the policy update.

The saddle structure of $\mathbb{E}[C^t \mid \mathcal{F}^{t-1}]$ immediately implies non-concavity. It prevents us from directly deriving its increase using standard results from gradient ascent theory, thereby significantly complicating the theoretical analysis in Theorem 3.5 (a high-level sketch and the complete proof are given in Appendix C). The following result extends the one-step guarantee to the full trajectory of Algorithm 1.

Lemma 3.6. *Let in the setting of Theorem 3.5 $\delta \leq \cos(\beta^0) \leq 1 - \delta$ with $\delta \in \left(0; \frac{1}{2}\right)$, $\|\mu^t\| \leq M$ for all t and $\varepsilon = \mathcal{O}\left(d^{-\frac{3}{2}}\delta M\right), \gamma_x^t = \mathcal{O}\left(\frac{\delta^2}{L}\right)$. Then $\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]$ increase monotonically to the value*

$$\cos\left(\frac{\delta}{32d} + L\gamma_x^t + \arccos(1-\delta)\right)^2 (1 - e^{-1}) = \mathcal{O}(1)$$

and then never decrease below.

Lemma 3.6 establishes once the policy is initialized such that $\delta \leq \cos(\beta^0) \leq 1-\delta$, for an appropriate choice of $\gamma_x^t, \gamma_\mu^t, \varepsilon$ the expected gradient alignment $\mathbb{E}\left[C^t\middle|\,\mathcal{F}^{t-1}\right]$ increases monotonically up to a certain level and subsequently oscillates within a neighbourhood of 1. This stands in contrast to the $\mathcal{O}(1/d)$ behaviour established in Corollary 3.4.

3.5 Convergence Guarantees

Lemma 3.6 induces two complementary initialization regimes, distinguished by whether prior directional information is available. Without such information, μ^0 is initialized randomly and the expected gradient alignment grows progressively until reaching its saturation region. Appendix C.1 analyzes this transient regime. Because the alignment quality changes along the trajectory, the resulting convergence criterion takes the non-standard form of a weighted average of squared gradient norms. Moreover, the admissible step size γ_x depends, through δ, on the initial angle between μ^0 and $\nabla f(x^0)$, yielding the usual trade-off between the probability that a random initialization satisfies the condition and the strength of the resulting bound.

When an approximate direction of $\nabla f(x^0)$ is available, μ^0 can instead be initialized directly in the high-alignment regime of Lemma 3.6, where $\mathbb{E}\left[C^t \mid \mathcal{F}^{t-1}\right]$ remains in a neighbourhood of 1. This bypasses the initial alignment-growth phase and leads to a standard convergence guarantee. The following lemma formalizes this second regime.

Lemma 3.7. *Let Assumptions 3.1, 3.2 be satisfied,* $d \geq 16$ *and* $\cos(\beta^0) \geq \frac{1}{4}$. *Then for iterations of Algorithm 1 with* $\varepsilon = \mathcal{O}\left(d^{-\frac{3}{2}} M\right), M = \min_t\{\|\mu^t\|\}, \gamma_x^t = \mathcal{O}\left(\frac{1}{L}\right), \gamma_\mu^t = \mathcal{O}\left(\|\mu^t\|^2\right)$ *the following bound holds:*

$$\mathbb{E}[\|\nabla f(\overline{x}^T)\|^2] \leq \mathcal{O}\left(\frac{1}{T\gamma_x}\right)(f(x^0) - f(x^*)).$$

Discussion. The bound in Lemma 3.7 has no explicit dependence on the dimension d, unlike the polynomial dependence incurred by classical directional, coordinate, and ZO methods (Nesterov & Spokoiny, 2017; Veprikov et al., 2024; Hanzely et al., 2018). For uniformly random directions, such a slowdown is unavoidable: existing lower bounds establish an inherent $\mathcal{O}(d)$ factor (Lee & Sidford, 2013). Our result does not contradict this limitation since it learns a biased distribution over search directions. Notably, the strengthened initialization condition requires only an approximate direction of $\nabla f(x^0)$, which can be obtained by aggregating several ZO estimates from (2), or by computing a first-order gradient solely at the initial iteration. The latter remains substantially less restrictive than using first-order gradients throughout training.

3.6 Toy Experiment

To validate the theoretical results, we conduct a toy experiment consisting of training a linear regression model on the a9a (Chang & Lin, 2011) dataset with access to directional derivatives. As a baseline, we consider an algorithm that employs the estimator with K Monte Carlo samples:

$$g_x^t = \frac{1}{K}\sum_{k=1}^{K} \overline{v}_k^t \langle \overline{v}_k^t, \nabla f(x^t)\rangle, \tag{5}$$

with $v_k^t \sim \mathcal{N}(0, I)$. For `LDSD`, we use the same approximation (5) with the choice $v_k^t \sim \mathcal{N}(\mu^t, \varepsilon^2 I)$. For the proposed learning mechanism, we apply a log-derivative trick (Williams, 1992) to derive a Monte Carlo estimator of $\nabla_\mu \mathbb{E}\left[C^t | \mathcal{F}^{t-1}\right]\Big|_{\mu=\mu^t}$ from Line 6 of Algorithm 1. We denote the density of a Gaussian distribution by $\pi_\mu(\cdot)$ and introduce $C_k^t = \langle \overline{v_k^t}, \overline{\nabla f(x^t)}\rangle^2$ for each sampled direction v_k^t. Accordingly, $\nabla_\mu \mathbb{E}\left[C^t | \mathcal{F}^{t-1}\right]\Big|_{\mu=\mu^t} = \int C^t \left(\nabla_\mu \log \pi_{\mu^t}\right) \mathrm{d}\pi_{\mu^t} \approx \frac{1}{K}\sum_{k=1}^{K} C_k^t \left(\nabla_\mu \log \pi_{\mu^t}(v_k^t)\right)$. Since $\pi_{\mu^t}(\cdot)$ is a Gaussian distribution density, we have $\nabla_\mu \log \pi_{\mu^t}(v_k^t) = \frac{v_k^t - \mu^t}{\varepsilon^2}$. Hence, for policy updates we utilize $g_\mu^t = \frac{1}{K}\sum_{k=1}^{K}(C_k^t - b^t)\frac{v_k^t - \mu^t}{\varepsilon^2}$, where $b^t = \frac{1}{K}\sum_{k=1}^{K} C_k^t$ is a mean-baseline, introduced to reduce the estimator variance Williams (1992).

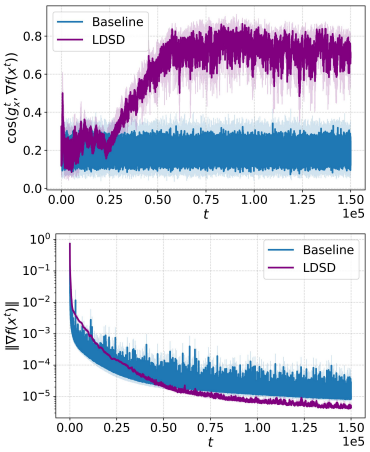


Figure 2: Comparison of LDSD and the baseline

For fixed K, the accuracy of this Monte Carlo policy-gradient estimate may still depend on the dimension d. In practice, however, stable estimates are often obtained with only $K = \mathcal{O}(1)$ samples (Schulman et al., 2017). We set $K = 5$ both for the baseline and `LDSD` and compare them in terms of the alignment between the true gradient and its estimator $\cos(g_x^t, \nabla f(x^t))$, as well as the norm of the gradient $\|\nabla f(x^t)\|$: the quantity involved in convergence guarantees. The results are reported in Figure 2.

The first plot in Figure 2 reveals that learning the sampling direction in `LDSD` indeed leads to strong alignment between the estimator g_μ^t and the true gradient $\nabla f(x^t)$. This behaviour fully agrees with Lemma 3.6 and the intuition discussed in Section 3: the alignment initially increases and subsequently oscillates in a neighbourhood of $\mathcal{O}(1)$.

4 ZERO-ORDER FRAMEWORK

Building on Algorithm 1, which presents the theoretical directional instantiation of our update rules, we now develop the corresponding practical zero-order formulation in Algorithm 2. First, we replace the directional derivative in Line 5 of Algorithm 1 with the standard ZO finite-difference oracle (2) in Line 7 of Algorithm 2. Estimator (2) reduces to the directional derivative in the limit $\tau \to 0$, and its use follows the standard practice of translating theoretically tractable directional updates into implementable ZO procedures (Malladi et al., 2023; Chen et al., 2024). Second, as in Section 3.6, we compute g_μ^t via the REINFORCE estimator. Finally, REINFORCE naturally requires multiple sampled directions, which can also be used to improve the ZO estimator g_x^t. In particular, in Line 4 of Algorithm 2, we greedily select the sample that is most collinear with the true gradient, further biasing the estimate toward this direction. The effects of the learnable direction and the sample selection strategy are isolated in Section 5.4.

Aside from those mechanics, Algorithm 2 preserves the structural update logic of Algorithm 1. The procedure preserves the forward-only nature of ZO methods, requires $K + 1$ forward evaluations per iteration, and concentrates sampling mass on empirically useful directions through online updates of μ^t. While classical ZO schemes typically use $K = 1$, we set $K = 5$ in our experiments as an empirically chosen compromise between selection quality and oracle cost. An ablation over K is provided in Section 5.4.

Algorithm 2 ZO-LDSD

1: **Input:** Initial points x^0, μ^0, number of iterations T, number of oracle calls K, variance ε, stepsizes γ_x, γ_μ
2: **for** $t = 0, 1, 2, \ldots, T-1$ **do**
3: Sample $\{v^i\}_{i=1}^K \sim \mathcal{N}(\mu^t, \varepsilon^2 I)$
4: Set $v_*^t = \arg\min_{v \in \{v^i\}_{i=1}^K} \{f(x^t + \tau v)\}$
5: $g_x^t = \frac{f(x^t+\tau v_*^t) - f(x^t - \tau v_*^t)}{2\tau} \cdot v_*^t$
6: $g_\mu^t = \frac{1}{K} \sum_{i=1}^K \frac{K \cdot f(x^t+\tau v^i) - \sum_{j=1}^K f(x^t+\tau v^j)}{K-1} \cdot \frac{v^i - \mu^t}{\varepsilon^2}$
7: $x^{t+1} = x^t - \gamma_x \cdot g_x^t$
8: $\mu^{t+1} = \mu^t + \gamma_\mu \cdot g_\mu^t$
9: **end for**

It is important to note that this framework is not a standalone optimizer but a plug-in module: any existing ZO estimator for the parameter update can be combined with the adaptive sampling routine to improve empirical convergence. In particular, the update at line 7 of Algorithm 2 is modular and may be replaced by the parameter-update rule of any other ZO optimizer. This enables straightforward integration of the adaptive sampling mechanism, while leaving the base optimizer's update logic intact.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETUP

Fine-Tuning Task and Schemes. We evaluate the proposed zero-order methods on two benchmarks: SST-2 sentiment classification (Socher et al., 2013) and COPA causal reasoning (Roemmele et al., 2011), following the experimental protocol of (Zhang et al., 2024). SST-2 is a standard supervised benchmark for sentence-level binary sentiment classification and COPA is a commonsense causal reasoning benchmark; both are commonly used to assess fine-tuning procedures for language models.

Models. Experiments are performed on two representative models: RoBERTa-Large (Liu et al., 2019) and OPT-1.3B (Zhang et al., 2022). These models span different architectures and parameter counts, allowing us to examine the behavior of ZO algorithms across realistic fine-tuning settings while keeping the experimental scope focused. To extend the evaluation to the larger-model regime, we additionally consider Llama 2 (Touvron et al., 2023) and OPT-13B (Zhang et al., 2022) under the LoRA fine-tuning scheme.

Baselines. We use our framework on established zero-order baselines reported in (Zhang et al., 2024): ZO-SGD (Ghadimi & Lan, 2013), ZO-AdaMM (Chen et al., 2019), JAGUAR SignSGD Petrov et al. (2025). All experiments follow the evaluation of (Zhang et al., 2024) to ensure a fair comparison. Training details are deferred to Appendix B.2.

Comparison procedure. For a fair and controlled evaluation we compare each baseline in two variants: (i) the original baseline implementation, and (ii) the same baseline augmented with our zero-order adaptive sampling framework (Algorithm 2). Crucially, in (ii) we do not change any of the original optimizer hyperparameters from (i) (e.g., learning rate, momentum), those remain exactly as in the base method. The only hyperparameters we tune are those introduced by our framework.

Since Algorithm 2 requires 6 ZO oracle calls per iteration for $K = 5$, all methods are compared under an equal oracle-call budget rather than an equal number of iterations. Accordingly, baselines are evaluated using either $K = 5$ (6 calls per iteration) or $K = 1$ (2 calls per iteration with proportionally more iterations), as reported in Table 1. This ensures that observed gains are to the adaptive sampling mechanism rather than computational budget differences.

Table 1: Test accuracy on SST2 for OPT-1.3B and RoBERTa-Large. Best performance among ZO methods is in **bold**.

Method	Sampling	OPT-1.3B		RoBERTa-Large	
		FT	LoRA	FT	LoRA
ZO-SGD	Gaussian, $K = 1$, more iterations	0.928	0.854	0.899	0.913
	Gaussian, $K = 5$, same iterations	0.919	0.850	0.869	0.905
	Algorithm 2	**0.935**	**0.916**	**0.909**	**0.928**
ZO-AdaMM	Gaussian, $K = 1$, more iterations	0.925	0.912	0.909	0.812
	Gaussian, $K = 5$, same iterations	0.917	0.905	0.899	0.801
	Algorithm 2	**0.930**	**0.929**	**0.917**	**0.900**
JAGUAR	Gaussian, $K = 1$, more iterations	0.925	0.921	0.906	0.905
	Gaussian, $K = 5$, same iterations	0.914	0.911	0.896	0.893
	Algorithm 2	**0.930**	**0.928**	**0.922**	**0.927**

Table 2: Test accuracy on larger models. Best performance per optimizer group in **bold**.

Method	Sampling	Llama-7B		OPT-13B
		Copa	SST-2	SST-2
ZO-SGD	Gaussian, $K = 1$, more iterations	0.86	0.919	0.611
	Algorithm 2	**0.90**	**0.937**	**0.620**
ZO-AdaMM	Gaussian, $K = 1$, more iterations	0.82	0.913	0.917
	Algorithm 2	**0.85**	**0.930**	**0.936**
JAGUAR	Gaussian, $K = 1$, more iterations	0.87	0.922	0.905
	Algorithm 2	**0.92**	**0.940**	**0.927**

5.2 Results

Table 1 reports SST2 test accuracy for OPT-1.3B and RoBERTa-Large under different fine-tuning schemes. Table 2 reports test accuracy for Llama-7B and OPT-13B under the LoRA fine-tuning setup.

Discussion. The results show that replacing Gaussian sampling with our adaptive direction sampling produces consistent improvements across models, optimizers, and fine-tuning modalities. Notably, merely increasing the number of probes with Gaussian sampling does not yield the same benefit: performance gains arise from the combination of multiple candidates and adaptive selection. These findings indicate that our sampling scheme yields more informative finite-difference probes and reduces variance in the resulting ZO gradient proxies.

5.3 Memory consumption comparison

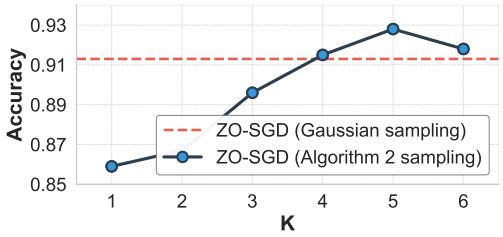
Table 3 reports peak GPU memory on SST-2 RoBERTa-Large (full FT). Across all optimizers Algorithm 2 adds moderate overhead (only the mean vector μ^t).

Table 3: Peak GPU memory (GB) on SST-2 RoBERTa-Large with full fine-tuning.

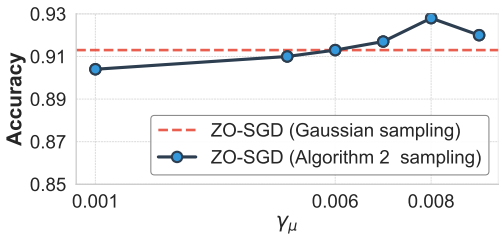
Method	Sampling	Peak mem, GB
ZO-SGD	Gaussian, $K = 2$	4.007
	Gaussian, $K = 6$	4.007
	Algorithm 2	**4.832**
ZO-AdaMM	Gaussian, $K = 2$	6.661
	Gaussian, $K = 6$	6.661
	Algorithm 2	**7.479**
JAGUAR	Gaussian, $K = 2$	3.508
	Gaussian, $K = 6$	4.011
	Algorithm 2	**4.943**

5.4 Ablation Studies

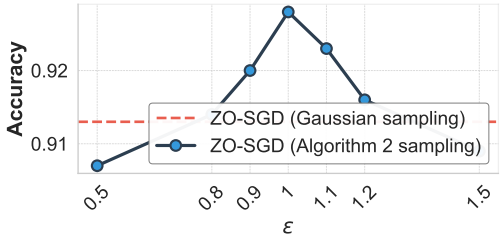
In this section, we provide several ablation studies on hyperparameters from Algorithm 2. In all experiments we use SST-2 dataset with the RoBERTa-large model under the LoRA fine-tuning scheme. Figure 3a shows accuracy rising with K up to $K = 5$ and then slightly falling; Figure 3b exhibits improvement as γ_μ increases until an intermediate optimum after which performance degrades; Figure 3c displays a U-shaped dependence on ε with a clear peak where our method outperforms the Gaussian baseline. These results motivate the default hyperparameter choices used in our experiments. Figure 3d presents an additional ablation of the ZO gradient estimator. We introduce a baseline that removes Line 4 from Algorithm 2 and averages the estimators from Line 5 over three sampled directions (keeping the total number of forward passes fixed at $K = 6$). This comparison justifies the use of the $argmin$ strategy in experiments and isolates its contribution from that of the learnable direction mechanism, which accounts for the majority of the observed gains.



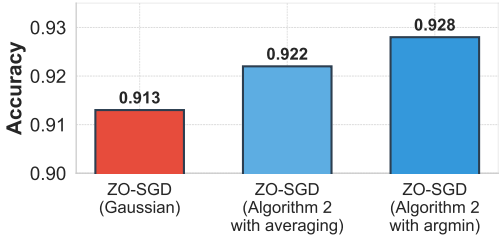
(a) Test accuracy of ZO-SGD (Algorithm 2 sampling) for different values of K.



(b) Test accuracy of ZO-SGD (Algorithm 2 sampling) for different values of γ_μ.



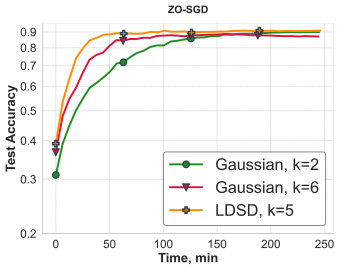
(c) Test accuracy of ZO-SGD (Algorithm 2 sampling) for different values of ε.



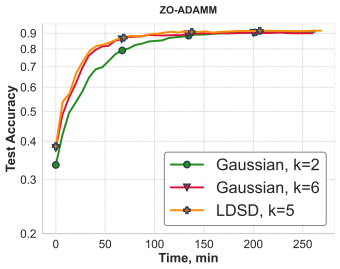
(d) Test accuracy of ZO-SGD for different types of sampling.

5.5 Wall-clock time comparison

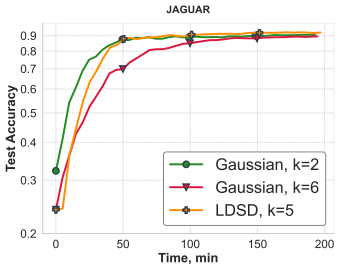
We compare test accuracy vs. wall-clock time on SST-2 RoBERTa-Large (full FT) for all three optimizers, matching the number of forward passes between our method and the Gaussian baseline. Algorithm 2 consistently reaches higher accuracy within the same time budget.



(a) ZO-SGD: test accuracy vs. wall-clock time on SST-2 RoBERTa-Large (full FT).



(b) ZO-AdaMM: test accuracy vs. wall-clock time on SST-2 RoBERTa-Large (full FT).



(c) JAGUAR: test accuracy vs. wall-clock time on SST-2 RoBERTa-Large (full FT).

6 Conclusion

In this work, we introduce a novel zero-order optimization framework that treats the sampling distribution over perturbation directions as a learnable policy. Theoretically, we identify gradient alignment as the key quantity governing convergence and leverage it to derive guarantees with no explicit dependence on the dimension d, in stark contrast to the $\mathcal{O}(d)$ degradation inherent in classical zero-order methods. Empirically, we validate our framework on LLM fine-tuning benchmarks, demonstrating consistent improvements over established ZO baselines including ZO-SGD, ZO-AdaMM, and JAGUAR SignSGD. Our framework operates as a plug-and-play module that enhances existing ZO optimizers without modifying their core hyperparameters. Extensive ablation studies confirm the robustness and practical applicability of our approach.

LLM Usage

We used large language models (LLMs) to polish the language and improve the clarity of the manuscript.

REFERENCES

Shun-ichi Amari. Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4): 185–196, 1993.

Atılım Güneş Baydin, Barak A Pearlmutter, Don Syme, Frank Wood, and Philip Torr. Gradients without backpropagation. *arXiv preprint arXiv:2202.08587*, 2022.

Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3): 334–334, 1997.

Akhilan Boopathy and Ila Fiete. How to train your wide neural network without backprop: An input-weight alignment perspective. In *International Conference on Machine Learning*, pp. 2178–2205. PMLR, 2022.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013. ISBN 9780199535255. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky. The convex geometry of linear inverse problems. *Foundations of Computational mathematics*, 12(6):805–849, 2012.

Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):1–27, 2011.

Pin-Yu Chen, Huan Zhang, Yash Sharma, Jinfeng Yi, and Cho-Jui Hsieh. Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models. In *Proceedings of the 10th ACM workshop on artificial intelligence and security*, pp. 15–26, 2017.

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. *Advances in neural information processing systems*, 32, 2019.

Yiming Chen, Yuan Zhang, Liyuan Cao, Kun Yuan, and Zaiwen Wen. Enhancing zeroth-order fine-tuning for language models with low-rank structures, 2024. *URL https://arxiv.org/abs/2410.07698*, 2024.

Shuyu Cheng, Guoqiang Wu, and Jun Zhu. On the convergence of prior-guided zeroth-order optimization algorithms. *Advances in Neural Information Processing Systems*, 34:14620–14631, 2021.

Aaron Defazio, Francis Bach, and Simon Lacoste-Julien. Saga: A fast incremental gradient method with support for non-strongly convex composite objectives. *Advances in neural information processing systems*, 27, 2014.

Tim Dettmers, Mike Lewis, Sam Shleifer, and Luke Zettlemoyer. 8-bit optimizers via block-wise quantization. *arXiv preprint arXiv:2110.02861*, 2021.

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.

Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. *arXiv preprint cs/0408007*, 2004.

Lei Gao, Amir Ziashahabi, Yue Niu, Salman Avestimehr, and Murali Annavaram. Enabling efficient on-device fine-tuning of llms using only inference engines. *arXiv preprint arXiv:2409.15520*, 2024.

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.

Saeed Ghadimi and Guanghui Lan. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM journal on optimization*, 23(4):2341–2368, 2013.

Filip Hanzely, Konstantin Mishchenko, and Peter Richtárik. Sega: Variance reduction via gradient sketching. *Advances in Neural Information Processing Systems*, 31, 2018.

Geoffrey Hinton. The forward-forward algorithm: Some preliminary investigations. *arXiv preprint arXiv:2212.13345*, 2(3):5, 2022.

Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*, 2018.

Max Jaderberg, Wojciech Marian Czarnecki, Simon Osindero, Oriol Vinyals, Alex Graves, David Silver, and Koray Kavukcuoglu. Decoupled neural interfaces using synthetic gradients. In *International conference on machine learning*, pp. 1627–1635. PMLR, 2017.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of statistics*, pp. 1302–1338, 2000.

Yin Tat Lee and Aaron Sidford. Efficient accelerated coordinate descent methods and faster algorithms for solving linear systems. In *2013 ieee 54th annual symposium on foundations of computer science*, pp. 147–156. IEEE, 2013.

Brian Lester, Rami Al-Rfou, and Noah Constant. The power of scale for parameter-efficient prompt tuning. *arXiv preprint arXiv:2104.08691*, 2021.

Bingrui Li, Jianfei Chen, and Jun Zhu. Memory efficient optimizers with 4-bit states. *Advances in Neural Information Processing Systems*, 36:15136–15171, 2023.

Xingguo Li, Tuo Zhao, Raman Arora, Han Liu, and Mingyi Hong. On faster convergence of cyclic block coordinate descent-type methods for strongly convex minimization. *Journal of Machine Learning Research*, 18(184):1–24, 2018.

Zhize Li, Hongyan Bao, Xiangliang Zhang, and Peter Richtárik. Page: A simple and optimal probabilistic gradient estimator for nonconvex optimization. In *International conference on machine learning*, pp. 6286–6295. PMLR, 2021.

Changyue Liao, Mo Sun, Zihan Yang, Jun Xie, Kaiqi Chen, Binhang Yuan, Fei Wu, and Zeke Wang. Lohan: Low-cost high-performance framework to fine-tune 100b model on a consumer gpu. *arXiv preprint arXiv:2403.06504*, 2024.

Sijia Liu, Pin-Yu Chen, Bhavya Kailkhura, Gaoyuan Zhang, Alfred O Hero III, and Pramod K Varshney. A primer on zeroth-order optimization in signal processing and machine learning: Principals, recent advances, and applications. *IEEE Signal Processing Magazine*, 37(5):43–54, 2020.

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, 2019. URL `https://arxiv.org/abs/1907.11692`.

Qijun Luo, Hengxu Yu, and Xiao Li. Badam: A memory efficient full parameter optimization method for large language models. *Advances in Neural Information Processing Systems*, 37: 24926–24958, 2024.

Kai Lv, Yuqing Yang, Tengxiao Liu, Qipeng Guo, and Xipeng Qiu. Full parameter fine-tuning for large language models with limited resources. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 8187–8198, 2024.

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.

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

Arild Nøkland and Lars Hiller Eidnes. Training neural networks with local error signals. In *International conference on machine learning*, pp. 4839–4850. PMLR, 2019.

Georgii Sergeevich Novikov, Daniel Bershatsky, Julia Gusak, Alex Shonenkov, Denis Valerievich Dimitrov, and Ivan Oseledets. Few-bit backward: Quantized gradients of activation functions for memory footprint reduction. In *International Conference on Machine Learning*, pp. 26363–26381. PMLR, 2023.

Egor Petrov, Grigoriy Evseev, Aleksey Antonov, Andrey Veprikov, Nikolay Bushkov, Stanislav Moiseev, and Aleksandr Beznosikov. Leveraging coordinate momentum in signsgd and muon: Memory-optimized zero-order. *arXiv preprint arXiv:2506.04430*, 2025.

Junbin Qiu, Zhengpeng Xie, Xiangda Yan, Yongjie Yang, and Yao Shu. Zeroth-order optimization is secretly single-step policy optimization. *arXiv preprint arXiv:2506.14460*, 2025.

Peter Richtárik and Martin Takáč. Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Mathematical Programming*, 144(1):1–38, 2014.

Peter Richtárik, Igor Sokolov, and Ilyas Fatkhullin. Ef21: A new, simpler, theoretically better, and practically faster error feedback. *Advances in Neural Information Processing Systems*, 34: 4384–4396, 2021.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pp. 400–407, 1951.

Melissa Roemmele, Cosmin Adrian Bejan, and Andrew S Gordon. Choice of plausible alternatives: An evaluation of commonsense causal reasoning. In *AAAI spring symposium: logical formalizations of commonsense reasoning*, pp. 90–95, 2011.

Yangjun Ruan, Yuanhao Xiong, Sashank Reddi, Sanjiv Kumar, and Cho-Jui Hsieh. Learning to learn by zeroth-order oracle. *arXiv preprint arXiv:1910.09464*, 2019.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Andrei Semenov, Matteo Pagliardini, and Martin Jaggi. Benchmarking optimizers for large language model pretraining. *arXiv preprint arXiv:2509.01440*, 2025.

Hyunseok Seung, Jaewoo Lee, and Hyunsuk Ko. Low-rank curvature for zeroth-order optimization in llm fine-tuning. *arXiv preprint arXiv:2511.07971*, 2025.

David Silver, Anirudh Goyal, Ivo Danihelka, Matteo Hessel, and Hado van Hasselt. Learning by directional gradient descent. In *International Conference on Learning Representations*, 2021.

Utkarsh Singhal, Brian Cheung, Kartik Chandra, Jonathan Ragan-Kelley, Joshua B Tenenbaum, Tomaso A Poggio, and Stella X Yu. How to guess a gradient. *arXiv preprint arXiv:2312.04709*, 2023.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631–1642, 2013.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

Andrey Veprikov, Alexander Bogdanov, Vladislav Minashkin, and Aleksandr Beznosikov. New aspects of black box conditional gradient: Variance reduction and one point feedback. *Chaos, Solitons & Fractals*, 189:115654, 2024.

Yining Wang, Simon Du, Sivaraman Balakrishnan, and Aarti Singh. Stochastic zeroth-order optimization in high dimensions. In Amos Storkey and Fernando Perez-Cruz (eds.), *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pp. 1356–1365. PMLR, 09–11 Apr 2018. URL https://proceedings.mlr.press/v84/wang18e.html.

Kaiyue Wen, David Hall, Tengyu Ma, and Percy Liang. Fantastic pretraining optimizers and where to find them. *arXiv preprint arXiv:2509.02046*, 2025.

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

Zhou Zhai, Wanli Shi, Heng Huang, Yi Chang, and Bin Gu. Learning sampling policy to achieve fewer queries for zeroth-order optimization. In *International Conference on Artificial Intelligence and Statistics*, pp. 1162–1170. PMLR, 2024.

Heshen Zhan, Congliang Chen, Tian Ding, Ziniu Li, and Ruoyu Sun. Unlocking black-box prompt tuning efficiency via zeroth-order optimization. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 14825–14838, 2024.

Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Dewan, Mona Diab, Xian Li, Xi Victoria Lin, Todor Mihaylov, et al. Opt: Open pre-trained transformer language models, 2022. URL https://arxiv.org/abs/2205.01068.

Yihua Zhang, Pingzhi Li, Junyuan Hong, Jiaxiang Li, Yimeng Zhang, Wenqing Zheng, Pin-Yu Chen, Jason D. Lee, Wotao Yin, Mingyi Hong, Zhangyang Wang, Sijia Liu, and Tianlong Chen. Revisiting zeroth-order optimization for memory-efficient llm fine-tuning: A benchmark, 2024. URL https://arxiv.org/abs/2402.11592.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv preprint arXiv:1911.00536*, 2019.

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. *arXiv preprint arXiv:2402.15173*, 2024.

Ligeng Zhu, Lanxiang Hu, Ji Lin, Wei-Ming Chen, Wei-Chen Wang, Chuang Gan, and Song Han. Pockengine: Sparse and efficient fine-tuning in a pocket. In *Proceedings of the 56th Annual IEEE/ACM International Symposium on Microarchitecture*, pp. 1381–1394, 2023.

A Additional Related Work

A.1 Memory efficient approaches

We note that, beyond ZO approaches, the memory-efficient optimization literature includes several orthogonal techniques, such as gradient quantization to lower precision (Novikov et al., 2023), compression of optimizer statistics (Dettmers et al., 2021; Li et al., 2023), forward gradient learning (Baydin et al., 2022; Silver et al., 2021), greedy layer-wise strategy (Nøkland & Eidnes, 2019), synthetic gradients (Jaderberg et al., 2017) and others (Lv et al., 2024; Boopathy & Fiete, 2022; Hinton, 2022; Singhal et al., 2023).

Nevertheless, ZO methods have emerged as the most effective solutions for memory-efficient LLM fine-tuning Malladi et al. (2023). The most apparent distinction among existing ZO approaches lies in the choice of the sampling distribution for the random direction v. Commonly used options include the uniform distribution over the unit sphere $RS(1)^d_{\|\cdot\|}$ (Flaxman et al., 2004), the standard multivariate Gaussian distribution $\mathcal{N}(0, I)$ (Ghadimi & Lan, 2013; Nesterov & Spokoiny, 2017), and a uniform discrete distribution over the vectors of the canonical (one-hot) basis (Duchi et al., 2015). These choices yield unbiased gradient estimators of the smoothed function f, which form the foundation of the theoretical analyses of the ZO methods.

Recent works have explored complementary strategies for making ZO optimization more practical and memory-efficient. ZO-AdaMM Chen et al. (2019) generalizes adaptive-moment methods to the gradient-free regime, combining momentum with per-coordinate adaptive step-sizes using only function evaluations. JAGUAR Veprikov et al. (2024); Petrov et al. (2025) implements a momentum-focused ZO design and related matrix-aware extensions, emphasizing minimal per-iteration oracle cost and memory-efficient updates for large-model fine-tuning.

A.2 ZO optimization with adaptive direction sampling

Recent advances have aimed to inject richer curvature information into ZO procedures to improve stability and convergence. One of the first approaches in this direction was proposed in HiZOO Zhao et al. (2024): it augments zero-order gradient (2) with precondition matrix, using a Hessian-informed strategy to mitigate heterogeneous curvature across parameters.

Another recent perspective casts ZO finite-difference procedures as instances of policy optimization. In the work Qiu et al. (2025), authors show that standard ZO finite-difference estimators are mathematically equivalent to a single-step policy optimization objective and, equivalently, to a REINFORCE-style gradient estimator Williams (1992) with a particular baseline. Building on this insight, they propose ZoAR method, which incorporates policy-optimization–inspired variance reduction through an averaged baseline and query reuse. Another approach learns ZO search mechanisms with auxiliary neural networks. Ruan et al. (2019) propose ZO-LSTM, combining an UpdateRNN that predicts coordinatewise parameter updates with a QueryRNN that predicts the diagonal covariance of Gaussian perturbations; both are trained by truncated backpropagation through optimization trajectories. Zhai et al. (2024) propose ZO-RL, casting perturbation selection as a Markov decision process and training an actor-critic policy using objective decreases as rewards. Their implementations introduce additional state and training costs: ZO-LSTM maintains recurrent states for each optimized coordinate, while ZO-RL trains actor and critic networks and retains target networks and a replay buffer.

A further contribution is presented in Seung et al. (2025), which introduces LOREN: a curvature-aware ZO method that formulates gradient preconditioning as adaptive estimation of an anisotropic perturbation distribution, employs a low-rank block-diagonal preconditioner within a natural evolution strategies framework, and uses a REINFORCE leave-one-out estimator to improve convergence and search efficiency.

B Additional Experimental Details

B.1 Toy experiment

The tuned hyperparameters used in toy experiment are as follows:

- **Baseline DGD:**
  - $\gamma_x^t = 200$.
- **LSDS:**
  - $\gamma_x^t = 5$,
  - $\gamma_\mu^t = 1.4 \cdot 10^{-5}$,
  - $\varepsilon = 1.2 \cdot 10^{-2}$.

B.2 Experiments on LLM fine-tuning

All experiments in Section 5 were performed on NVIDIA H200 GPU (141 GB HBM3e memory) and used the same core sampling hyperparameters: $\varepsilon = 1$ and $\gamma_\mu = 10^{-3}$. For ZO-SGD setups we fixed momentum $= 0.9$. For ZO-Adamm setups we fixed $(\beta_1, \beta_2) = (0.9, 0.999)$. For ZO-Adamm setups we fixed JAGUAR momentum $\beta = 0.9$. All methods employed a cosine learning-rate schedule for γ_x. Table 4 reports the base learning rates γ_x used for each method and experimental condition.

Table 4: Base learning rates γ_x used for OPT-1.3B and RoBERTa-Large experiments with FT and LoRA.

Method	OPT-1.3B		RoBERTa-Large	
	FT	LoRA	FT	LoRA
ZO-SGD (Gaussian sampling, 2 forwards, more iterations)	4.0×10^{-8}	1.0×10^{-6}	4.0×10^{-8}	5.0×10^{-4}
ZO-SGD (Gaussian sampling, 6 forwards, same iterations)	4.0×10^{-8}	1.0×10^{-6}	4.0×10^{-8}	5.0×10^{-4}
ZO-SGD (Algorithm 2 sampling)	4.0×10^{-8}	1.0×10^{-6}	4.0×10^{-8}	5.0×10^{-4}
ZO-AdaMM (Gaussian sampling, 2 forwards, more iterations)	1.0×10^{-5}	6.0×10^{-5}	3.0×10^{-5}	3.0×10^{-3}
ZO-AdaMM (Gaussian sampling, 6 forwards, same iterations)	1.0×10^{-5}	6.0×10^{-5}	3.0×10^{-5}	3.0×10^{-3}
ZO-AdaMM (Algorithm 2 sampling)	1.0×10^{-5}	6.0×10^{-5}	3.0×10^{-5}	3.0×10^{-3}
JAGUAR SignSGD (Gaussian sampling, 2 forwards, more iterations)	2.0×10^{-6}	1.0×10^{-5}	4.0×10^{-6}	3.0×10^{-4}
JAGUAR SignSGD (Gaussian sampling, 6 forwards, same iterations)	2.0×10^{-6}	1.0×10^{-5}	4.0×10^{-6}	3.0×10^{-4}
JAGUAR SignSGD (Algorithm 2 sampling)	2.0×10^{-6}	1.0×10^{-5}	4.0×10^{-6}	3.0×10^{-4}

C Missing proofs

Lemma 3.3. *Under Assumptions 3.1 and 3.2 for iterations of the form*

$$x^{t+1} = x^t - \gamma_x^t \overline{v^t} \langle \overline{v^t}, \nabla f(x^t) \rangle,$$

the following inequality holds.

$$\sum_{t=0}^{T-1} \frac{1}{T} \mathbb{E}\left[\left(\gamma_x^t - \frac{L {\gamma_x^t}^2}{2} \right) \mathbb{E}\left[C^t | \mathcal{F}^{t-1} \right] \|\nabla f(x^t)\|^2 \right] \leq \frac{f(x^0) - f(x^*)}{T}.$$

Proof. We consider the algorithm step and utilize Assumption 3.1.

$$\begin{aligned} f(x^{t+1}) - f(x^t) &= f(x^t - \gamma_x^t \overline{v}^t \langle \nabla f(x^t), \overline{v}^t \rangle) - f(x^t) \\ &\overset{3.1}{\leq} -\gamma_x^t \langle \nabla f(x^t), \overline{v}^t \rangle^2 + \frac{L}{2} \|\gamma_x^t \overline{v}^t \langle \nabla f(x^t), \overline{v}^t \rangle\|^2 \\ &= \left(\frac{L {\gamma_x^t}^2}{2} - \gamma_x^t \right) \langle \nabla f(x^t), \overline{v}^t \rangle^2 \\ &= \left(\frac{L {\gamma_x^t}^2}{2} - \gamma_x^t \right) \langle \overline{\nabla f(x^t)}, \overline{v}^t \rangle^2 \|\nabla f(x^t)\|^2. \end{aligned}$$

Conditioning on $\mathcal{F}^t$ and using the $\mathcal{F}^{t-1}$-measurability of $\nabla f(x^t)$ and γ_x^t, we obtain

$$\mathbb{E}\left[f(x^{t+1})|\mathcal{F}^{t-1}\right] - f(x^t) \leq \left(\frac{L{\gamma_x^t}^2}{2} - \gamma_x^t\right)\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]\|\nabla f(x^t)\|^2.$$

Summing over $t = 0, \ldots, T-1$ and considering full expectation we derive

$$\sum_{t=0}^{T-1}\frac{1}{T}\mathbb{E}\left[\left(\gamma_x^t - \frac{L{\gamma_x^t}^2}{2}\right)\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]\|\nabla f(x^t)\|^2\right] \leq \frac{f(x^0) - \mathbb{E}[f(x^{t-1})]}{T}.$$

Applying Assumption 3.2 finishes the proof. □

Corollary 3.4. *Let in the setting of Lemma 3.3 $v^t \overset{i.i.d.}{\sim} RS(1)^d_{\|\cdot\|}$ be distributed uniformly on the unit sphere and $\gamma_x^t \leq \frac{1}{2L}$. Then the following bound holds:*

$$\mathbb{E}\|\nabla f(\overline{x}^T)\|^2 \leq \frac{2d}{\gamma_x^t T}(f(x^0) - f(x^*)),$$

where $\overline{x}^T$ is chosen uniformly from $\{x^0, x^1, \ldots, x^{T-1}\}$.

Proof. The choice $\gamma_x^t \leq \frac{1}{2L}$ guarantees:

$$\gamma_x^t - \frac{L{\gamma_x^t}^2}{2} \geq \frac{\gamma_x^t}{2} \tag{6}$$

Then we need to evaluate $\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]$. We consider coordinate system where $e_1 \parallel \nabla f(x^t)$:

$$\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right] = \mathbb{E}_{v^t\sim\mathcal{N}(0,I)}\left[\frac{v_1^2}{\|v\|^2}\right].$$

Due to the distribution symmetry we conclude:

$$\mathbb{E}_{v^t\sim\mathcal{N}(0,I)}\left[\frac{v_1^2}{\|v\|^2}\right] = \mathbb{E}_{v^t\sim\mathcal{N}(0,I)}\left[\frac{v_2^2}{\|v\|^2}\right] = \cdots = \mathbb{E}_{v^t\sim\mathcal{N}(0,I)}\left[\frac{v_d^2}{\|v\|^2}\right].$$

Then,

$$d\cdot\mathbb{E}_{v^t\sim\mathcal{N}(0,I)}\left[\frac{v_1^2}{\|v\|^2}\right] = \mathbb{E}_{v^t\sim\mathcal{N}(0,I)}\left[\frac{v_1^2 + v_2^2 + \cdots + v_d^2}{\|v\|^2}\right] = \mathbb{E}_{v^t\sim\mathcal{N}(0,I)}[1] = 1.$$

And we reach:

$$\mathbb{E}_{v^t\sim\mathcal{N}(0,I)}\left[\frac{v_1^2}{\|v\|^2}\right] = \frac{1}{d} \tag{7}$$

Substitution of (7), (6) into Lemma 3.3 and rearranging the terms finishes the proof. □

Lemma C.1. *Let the function f be L-smooth, i.e.,*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^d,$$

then it holds

$$f(y) \geq f(x) + \langle\nabla f(x), y - x\rangle - \frac{L}{2}\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^d.$$

Proof. The claim follows from integrating the directional derivative of f along the segment connecting x and y and then applying the Lipschitz continuity of the gradient.

$$\begin{aligned} f(y) &= f(x) + \int_0^1 \langle\nabla f(x + \tau(y - x)), y - x\rangle d\tau \\ &= f(x) + \langle\nabla f(x), y - x\rangle \end{aligned}$$

$$
\begin{aligned}
& + \int_0^1 \langle \nabla f(x+\tau(y-x)) - \nabla f(x), y-x\rangle d\tau \\
\geq\ & f(x) + \langle \nabla f(x), y-x\rangle \\
& - \int_0^1 \|\nabla f(x+\tau(y-x)) - \nabla f(x)\| \|y-x\| d\tau \\
\geq\ & f(x) + \langle \nabla f(x), y-x\rangle - \int_0^1 L\tau \|y-x\|^2 d\tau \\
=\ & f(x) + \langle \nabla f(x), y-x\rangle - \frac{L}{2}\|y-x\|^2.
\end{aligned}
$$

□

Theorem 3.5. *Let Assumption 3.1 hold and $d \geq 16$. Then for iterations of the Algorithm 1 with $\varepsilon \leq \frac{d^{-\frac{3}{2}} \cos(\beta^t)(1-\cos(\beta^t))\|\mu^t\|}{480}, \gamma_\mu^t \leq \frac{\|\mu^t\|^2}{640}, \gamma_x^t \leq \frac{\gamma_\mu^t \cos^2(\beta^t)(1-\cos(\beta^t))^2}{2^8 \cdot 3^2 L\|\mu^t\|^2}$ the following inequality holds:*

$$
\mathbb{E}\left[C^{t+1} \middle| \mathcal{F}^t\right] \quad \geq \quad \mathbb{E}\left[C^t \middle| \mathcal{F}^{t-1}\right] + \gamma_\mu^t \left(\frac{\cos(\beta^t)(1-\cos(\beta^t))}{24\|\mu^t\|}\right)^2, \tag{8}
$$

where β^t is the angle between μ^t and $\nabla f(x^t)$.

Due to the non-standard challenges and the resulting proof technique, we provide high-level sketch before the proof.

Proof Sketch of Theorem 3.5. The proof of Theorem 3.5 is deferred to Section C. It tracks the joint evolution of two quantities: the gradient $\nabla f(x^t)$ and the mean vector μ^t. The main steps are as follows:

(i) We first reduce the gradient dynamics using Assumption 3.1, which yields a relation of the form

$$
\mathbb{E}\left[C^{t+1} \middle| \mathcal{F}^t\right] \geq -2a^t + F_{\overline{\nabla f(x^t)}}(\mu^{t+1}), \tag{9}
$$

with $F_z(\mu) = \mathbb{E}_{u\sim\mathcal{N}(\mu,\varepsilon^2 I)}[\langle z, \overline{u}\rangle^2]$ and $a^t = 4\frac{L\gamma_x^t}{2}\sqrt{1-\left(\frac{L\gamma_x^t}{2}\right)^2}$. We note that $\mathbb{E}\left[C^t \middle| \mathcal{F}^{t-1}\right] = F_{\overline{\nabla f(x^t)}}(\mu^t)$. Then, we need to reduce μ^{t+1} to μ^t in the RHS of (9) to obtain a recursion on expected gradient alignment.

(ii) The core objective is to prove strict monotonic growth of expected gradient alignment $\mathbb{E}\left[C^{t+1} \middle| \mathcal{F}^t\right]$. Achieving this requires carefully isolating a positive contribution that dominates the drift a^t induced by changes in $\nabla f(x^t)$. Then, to enable a quantitative analysis, we establish smoothness of the function $F_z(\mu)$ with respect to μ^t via a uniform bound on the Hessian norm. This yields an inequality of the form

$$
F_{\overline{\nabla f(x^t)}}(\mu^{t+1}) \geq F_{\overline{\nabla f(x^t)}}(\mu^t) + \left(\gamma_\mu^t - \frac{\tilde{L}{\gamma_\mu^t}^2}{2}\right)\|\nabla F_{\overline{\nabla f(x^t)}}(\mu^t)\|^2. \tag{10}
$$

(iii) Combining the above relations (9) and (10) yields a recursive inequality for $\mathbb{E}\left[C^{t+1} \middle| \mathcal{F}^t\right]$. To ensure a positive increment, it suffices to lower bound the contribution of the last term in (10). To this end, we decompose the expectation over u in $F_z(\mu)$ definition into integrals over an inner region $u \leq \frac{\|\mu^t\|}{2\varepsilon}$ and its complement.

The integral over the outer region is controlled by choosing $\varepsilon \lesssim \|\mu\| d^{-\frac{1}{2}}$, allowing us to apply Gaussian concentration inequalities for Lipschitz functions (Boucheron et al., 2013).

The inner-region contribution is analysed via a Taylor expansion with a Lagrange remainder. Each term is treated separately. First, symmetry considerations eliminate zero order term. Second, the second-order term is controlled using the previously established Hessian bound. Third, the Laurent-Massart inequalities (Laurent & Massart, 2000) serve as a core ingredient for extracting the required positive increment from the first order term.

Putting these estimates together with (9), (10) and balancing the competing effects yields the final bound.

We present the complete proof below.

Proof of Theorem 3.5. The evolution of $\mathbb{E}_t[C^t]$ depends on the dynamics of two quantities: the gradient $\nabla f(x^t)$ itself and the mean vector μ^t. We begin with **STEP I** by examining the contribution of the gradient dynamics.

Using Assumption 3.1 we bound the $\|\overline{\nabla f(x^{t+1})} - \overline{\nabla f(x^t)}\|$ in the following way.

By the x^{t+1} update rule $\|x^{t+1} - x^t\| = \|\gamma_x^t g_x^t\| = \gamma_x^t \left\|\overline{v^t}\left\langle \overline{v^t}, \nabla f(x^t)\right\rangle\right\| \le \gamma_x^t \|\nabla f(x^t)\|$. Therefore, $\|\nabla f(x^{t+1}) - \nabla f(x^t)\| \le L\gamma_x^t \|\nabla f(x^t)\|$, which implies that the angle between $\nabla f(x^t)$ and $\nabla f(x^{t+1})$ is at most

$$2\arcsin\left(\frac{\|\nabla f(x^{t+1}) - \nabla f(x^t)\|}{2\|\nabla f(x^t)\|}\right) \le 2\arcsin\left(\frac{L\gamma_x^t\|\nabla f(x^t)\|}{2\|\nabla f(x^t)\|}\right) = 2\arcsin\left(\frac{L\gamma_x^t}{2}\right).$$

The angle between $\overline{\nabla f(x^t)}$ and $\overline{\nabla f(x^{t+1})}$ is the same. It justifies

$$\begin{aligned}\|\overline{\nabla f(x^{t+1})} - \overline{\nabla f(x^t)}\| &\le 2\sin\left(2\arcsin\left(\frac{L\gamma_x^t}{2}\right)\right)\\ &= 4\frac{L\gamma_x^t}{2}\sqrt{1-\left(\frac{L\gamma_x^t}{2}\right)^2} =: a^t \end{aligned} \tag{11}$$

And now we proceed with the definition of C^t.

$$\begin{aligned}&\mathbb{E}_{v^{t+1}\sim\mathcal{N}(\mu^{t+1},\varepsilon^2 I)}\left[C^{t+1}\middle|\,\mathcal{F}^t\right]\\ &= \frac{1}{(2\pi\varepsilon^2)^{d/2}}\int_{\mathbb{R}^d}\left\langle\overline{\nabla f(x^{t+1})},\overline{u}\right\rangle^2 e^{-\frac{\|u-\mu^{t+1}\|^2}{2\varepsilon^2}}\mathrm{d}u\\ &= \frac{1}{(2\pi\varepsilon^2)^{d/2}}\int_{\mathbb{R}^d}\left\langle\overline{\nabla f(x^t)} + (\overline{\nabla f(x^{t+1})} - \overline{\nabla f(x^t)}),\overline{u}\right\rangle^2 e^{-\frac{\|u-\mu^{t+1}\|^2}{2\varepsilon^2}}\mathrm{d}u\\ &= \frac{1}{(2\pi\varepsilon^2)^{d/2}}\int_{\mathbb{R}^d}\Bigg\{\left\langle\overline{\nabla f(x^t)},\overline{u}\right\rangle^2 + \left\langle\overline{\nabla f(x^{t+1})} - \overline{\nabla f(x^t)},\overline{u}\right\rangle^2\\ &\quad + 2\left\langle\overline{\nabla f(x^{t+1})} - \overline{\nabla f(x^t)},\overline{u}\right\rangle\left\langle\overline{\nabla f(x^t)},\overline{u}\right\rangle\Bigg\}e^{-\frac{\|u-\mu^{t+1}\|^2}{2\varepsilon^2}}\mathrm{d}u\\ &\ge \frac{1}{(2\pi\varepsilon^2)^{d/2}}\int_{\mathbb{R}^d}\Bigg\{\left\langle\overline{\nabla f(x^t)},\overline{u}\right\rangle^2\\ &\quad + 2\left\langle\overline{\nabla f(x^{t+1})} - \overline{\nabla f(x^t)},\overline{u}\right\rangle\left\langle\overline{\nabla f(x^t)},\overline{u}\right\rangle\Bigg\}e^{-\frac{\|u-\mu^{t+1}\|^2}{2\varepsilon^2}}\mathrm{d}u\end{aligned} \tag{12}$$

Using Cauchy–Schwarz inequality and the bound (11) we derive

$$\left\langle\overline{\nabla f(x^{t+1})} - \overline{\nabla f(x^t)},\overline{u}\right\rangle\left\langle\overline{\nabla f(x^t)},\overline{u}\right\rangle \le a^t. \tag{13}$$

Utilizing (13), we continue with derivation (12) by introducing a constraint on the integral of the last term with respect to the probability measure.

$$\mathbb{E}_{v^{t+1}\sim\mathcal{N}(\mu^{t+1},\varepsilon^2 I)}\left[C^{t+1}\middle|\,\mathcal{F}^t\right] \ge -2a^t + \frac{1}{(2\pi\varepsilon^2)^{d/2}}\int_{\mathbb{R}^d}\left\langle\overline{\nabla f(x^t)},\overline{u}\right\rangle^2 e^{-\frac{\|u-\mu^{t+1}\|^2}{2\varepsilon^2}}\mathrm{d}u. \tag{14}$$

Thus, the gradient dynamics have been eliminated.

The next **STEP II** dedicated to the mean vector dynamics.

We start by introducing the function $F_z(\mu) = \mathbb{E}_{u\sim\mathcal{N}(\mu,\varepsilon^2 I)}[\langle z, \overline{u}\rangle^2]$ for every pair $\mu, z \in \mathbb{R}^d$. With this notation (14) and the definition of $\mathbb{E}_{u\sim\mathcal{N}(\mu^{t+1},\varepsilon^2 I)}\left[C^{t+1}\middle|\,\mathcal{F}^t\right]$ can be reformulated as

$$\mathbb{E}_{v^{t+1}\sim\mathcal{N}(\mu^{t+1},\varepsilon^2 I)}\left[C^{t+1}\middle|\,\mathcal{F}^t\right] \geq -2a^t + F_{\overline{\nabla f(x^t)}}(\mu^{t+1}), \tag{15}$$

$$\mathbb{E}_{v^{t}\sim\mathcal{N}(\mu^{t},\varepsilon^2 I)}\left[C^{t}\middle|\,\mathcal{F}^{t-1}\right] = F_{\overline{\nabla f(x^t)}}(\mu^{t}). \tag{16}$$

Next, we observe that Lines 6 and 7 of Algorithm 1 correspond to performing a step of gradient ascent with respect to μ on the function $F_{\overline{\nabla f(x^t)}}(\mu)$. Along with (15) and (16) it defines the $\mathbb{E}_t\left[C^t\right]$ dynamics.

Firstly, we introduce the kernel and change C^t notation to reflect implicit dependence on arguments:

$$\varphi(x) = \frac{1}{(2\pi\varepsilon^2)^{d/2}}\exp\left(\frac{-\|x\|^2}{2\varepsilon^2}\right), \tag{17}$$

$$\psi_a(x) = (\overline{a}, \overline{x})^2 = \frac{(a,x)^2}{\|x\|^2}. \tag{18}$$

Then we can reformulate $F_{\overline{\nabla f(x^t)}}(\mu)$ as follows

$$F_a(\mu) = \int_{\mathbb{R}^d}\psi_a(u)\varphi(\mu-u)\mathrm{d}u. \tag{19}$$

After that we start the Hessian computation using the general Stokes' theorem.

$$\begin{aligned}
\nabla_\mu F_a(\mu) &= \int_{\mathbb{R}^d}\psi_a(u)\nabla_\mu\varphi(\mu-u)\mathrm{d}u \\
&= \int_{\mathbb{R}^d}\psi_a(u)(-\nabla_u\varphi(\mu-u))\mathrm{d}u \\
&= -\int_{\mathbb{R}^d}\nabla_u(\psi_a(u)\varphi(\mu-u))\mathrm{d}u + \int_{\mathbb{R}^d}\nabla_u(\psi_a(u))\varphi(\mu-u)\mathrm{d}u \\
&= \int_{\mathbb{R}^d}\nabla_u(\psi_a(u))\varphi(\mu-u)\mathrm{d}u.
\end{aligned} \tag{20}$$

Following same reasoning we can derive

$$\nabla^2_\mu F_a(\mu) = \int_{\mathbb{R}^d}\nabla^2_u\psi_a(u)\varphi(\mu-u)\mathrm{d}u = \mathbb{E}_{u\sim\mathcal{N}(\mu,\varepsilon^2 I)}\left[\nabla^2_u\psi_a(u)\right]. \tag{21}$$

Straightforward computation provides

$$\begin{aligned}
\nabla^2_u\psi_a(u) &= \frac{1}{\|u\|^8}\big(\|u\|^4(2aa^\top\|u\|^2 + 4au^\top\langle a,u\rangle - 2\langle a,u\rangle^2 I \\
&\quad -4ua^\top\langle a,u\rangle) - 4\|u\|^2\left(2a\|u\|^2\langle a,u\rangle - 2\langle a,u\rangle^2 u\right)u^\top\big).
\end{aligned} \tag{22}$$

It implies for arbitrary a so that $\|a\| = 1$ the following bound holds.

$$\|\nabla^2_u\psi_a(u)\| \leq 20\frac{\|a\|^2}{\|u\|^2} = \frac{20}{\|u\|^2}. \tag{23}$$

Using (21), (23) and convexity of $\|\cdot\|$ we reach the bound

$$\|\nabla^2_\mu F_a(\mu)\| \leq 20\mathbb{E}_{u\sim\mathcal{N}(\mu,\varepsilon^2 I)}\left[\frac{1}{\|u\|^2}\right]. \tag{24}$$

Assumption $d > 3$ ensures the convergence of the integral in (24) and provides the bound of the Hessian norm, thereby asserting the $\tilde{L}$-smoothness of the objective function $F_a(\mu)$ with $\tilde{L} \leq A\mathbb{E}_{u\sim\mathcal{N}(\mu,\varepsilon^2 I)}\left[\frac{1}{\|u\|^2}\right]$.

$$
\begin{aligned}
&\mathbb{E}_{u\sim\mathcal{N}(\mu^t,\varepsilon^2 I)}\left[\frac{1}{\|u\|^2}\right] \\
&= \frac{1}{(2\pi\varepsilon^2)^{\frac{d}{2}}}\int_{\mathbb{R}^d}\frac{1}{\|u\|^2}e^{-\frac{\|u-\mu^t\|^2}{2\varepsilon^2}}\mathrm{d}u \\
&= \frac{1}{(2\pi\varepsilon^2)^{\frac{d}{2}}}\left[\int_{\|u\|\leq\frac{\|\mu^t\|}{4}}\frac{1}{\|u\|^2}e^{-\frac{\|u-\mu^t\|^2}{2\varepsilon^2}}\mathrm{d}u + \int_{\|u\|>\frac{\|\mu^t\|}{4}}\frac{1}{\|u\|^2}e^{-\frac{\|u-\mu^t\|^2}{2\varepsilon^2}}\mathrm{d}u\right] \\
&\leq \frac{1}{(2\pi\varepsilon^2)^{\frac{d}{2}}}\int_{\|u\|\leq\frac{\|\mu^t\|}{4}}\frac{1}{\|u\|^2}e^{-\frac{\|u\|^2+\|\mu^t\|^2-2\langle u,\mu^t\rangle}{2\varepsilon^2}}\mathrm{d}u + \frac{16}{\|\mu^t\|^2} \\
&\leq \frac{1}{(2\pi\varepsilon^2)^{\frac{d}{2}}}\int_{\|u\|\leq\frac{\|\mu^t\|}{4}}\frac{1}{\|u\|^2}e^{-\frac{\|u\|^2+\frac{\|\mu^t\|^2}{2}}{2\varepsilon^2}}\mathrm{d}u + \frac{16}{\|\mu^t\|^2} \\
&\leq e^{-\frac{\|\mu^t\|^2}{4\varepsilon^2}}\frac{1}{(2\pi\varepsilon^2)^{\frac{d}{2}}}\int_{\mathbb{R}^d}\frac{1}{\|u\|^2}e^{-\frac{\|u\|^2}{2\varepsilon^2}}\mathrm{d}u + \frac{16}{\|\mu^t\|^2} \\
&= \frac{e^{-\frac{\|\mu^t\|^2}{4\varepsilon^2}}}{\varepsilon^2(d-2)} + \frac{16}{\|\mu^t\|^2}.
\end{aligned}
$$

In the last transition we used classical result for χ^2 distribution r moment $2^r\frac{\Gamma\left(\frac{d}{2}+r\right)}{\Gamma\left(\frac{d}{2}\right)}$. Then, $\tilde{L} \leq \frac{20e^{-\frac{\|\mu^t\|^2}{4\varepsilon^2}}}{\varepsilon^2(d-2)} + \frac{16\cdot 20}{\|\mu^t\|^2}$.

Applying Lemma C.1 to function F taking into account implicit step formulation $\mu^{t+1} = \mu^t + \gamma_\mu^t\nabla F_{\overline{\nabla f(x^t)}}(\mu^t)$ we get

$$
\begin{aligned}
F_{\overline{\nabla f(x^t)}}(\mu^{t+1}) &\geq F_{\overline{\nabla f(x^t)}}(\mu^t) + (\nabla F_{\overline{\nabla f(x^t)}}(\mu^t), \mu^{t+1}-\mu^t) - \frac{\tilde{L}}{2}\|\mu^{t+1}-\mu^t\|^2 \\
&= F_{\overline{\nabla f(x^t)}}(\mu^t) + \left(\gamma_\mu^t - \frac{\tilde{L}{\gamma_\mu^t}^2}{2}\right)\|\nabla F_{\overline{\nabla f(x^t)}}(\mu^t)\|^2 \qquad (25)
\end{aligned}
$$

By combining (15), (16) and (25) we derive

$$
\begin{aligned}
\mathbb{E}_{v^{t+1}\sim\mathcal{N}(\mu^{t+1},\varepsilon^2 I)}\left[C^{t+1}\middle|\mathcal{F}^t\right] &\geq -2a^t + F_{\overline{\nabla f(x^t)}}(\mu^t) + \left(\gamma_\mu^t - \frac{\tilde{L}{\gamma_\mu^t}^2}{2}\right)\|\nabla F_{\overline{\nabla f(x^t)}}(\mu^t)\|^2 \\
&\overset{(16)}{=} \mathbb{E}_{v^t\sim\mathcal{N}(\mu^t,\varepsilon^2 I)}\left[C^t\middle|\mathcal{F}^{t-1}\right] \\
&\quad - 2a^t + \left(\gamma_\mu^t - \frac{\tilde{L}{\gamma_\mu^t}^2}{2}\right)\|\nabla F_{\overline{\nabla f(x^t)}}(\mu^t)\|^2 \qquad (26)
\end{aligned}
$$

At the next step we work on the lower bound of $\|\nabla F_a(\mu)\|$ in particular μ-space regions.

$$
\begin{aligned}
\nabla F_a(\mu) &= \frac{1}{(2\pi\varepsilon^2)^{d/2}}\int_{\mathbb{R}^d}\frac{\langle a,u\rangle^2}{\|u\|^2}\frac{u-\mu}{\varepsilon^2}e^{-\frac{\|u-\mu\|^2}{2\varepsilon^2}}\mathrm{d}u \\
&= \frac{1}{\varepsilon(2\pi)^{d/2}}\int_{\mathbb{R}^d}\frac{\langle a,\varepsilon v+\mu\rangle^2}{\|\varepsilon v+\mu\|^2}ve^{-\frac{\|v\|^2}{2}}\mathrm{d}v \\
&= \frac{1}{\varepsilon(2\pi)^{d/2}}\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}\frac{\langle a,\varepsilon v+\mu\rangle^2}{\|\varepsilon v+\mu\|^2}ve^{-\frac{\|v\|^2}{2}}\mathrm{d}v
\end{aligned}
$$

$$+ \frac{1}{\varepsilon(2\pi)^{d/2}} \int_{\|v\|>\frac{\|\mu\|}{2\varepsilon}} \frac{\langle a, \varepsilon v + \mu\rangle^2}{\|\varepsilon v + \mu\|^2} v e^{-\frac{\|v\|^2}{2}} \mathrm{d}v. \tag{27}$$

We proceed with evaluation of the second term of 27.

$$\begin{aligned}
&\left\| \frac{1}{(2\pi)^{d/2}} \int_{\|v\|>\frac{\|\mu\|}{2\varepsilon}} \frac{\langle a, \varepsilon v + \mu\rangle^2}{\|\varepsilon v + \mu\|^2} v e^{-\frac{\|v\|^2}{2}} \mathrm{d}v \right\| \\
&\leq \frac{1}{(2\pi)^{d/2}} \int_{\|v\|>\frac{\|\mu\|}{2\varepsilon}} \frac{\langle a, \varepsilon v + \mu\rangle^2}{\|\varepsilon v + \mu\|^2} \|v\| e^{-\frac{\|v\|^2}{2}} \mathrm{d}v \\
&\leq \frac{1}{(2\pi)^{d/2}} \int_{\|v\|>\frac{\|\mu\|}{2\varepsilon}} \|v\| e^{-\frac{\|v\|^2}{2}} \mathrm{d}v
\end{aligned} \tag{28}$$

Then we use the fact that for d-dimensional $\xi \sim \mathcal{N}(0, I)$ holds the further bound (Chandrasekaran et al., 2012):

$$\mathbb{E}\left[\|\xi\|\right] = \sqrt{2}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \leq \sqrt{d}. \tag{29}$$

After that we use Cauchy–Schwarz inequality in $L_2(\mathbb{R}^d)$ space with Gaussian measure to derive

$$\begin{aligned}
\mathbb{E}\left[I_{\{\|\xi\|>t\}}\|\xi\|\right] &\leq \left(\mathbb{E}\left[\|\xi\|^2\right]\right)^{\frac{1}{2}} \left(\mathbb{E}\left[I^2_{\{\|\xi\|>t\}}\right]\right)^{\frac{1}{2}} \\
&= \sqrt{d}\left(\Pr\left(\|\xi\| > t\right)\right)^{\frac{1}{2}}.
\end{aligned} \tag{30}$$

Gaussian measure concentration inequality (Boucheron et al., 2013) for $L = 1$ Lipchitz function $\|\cdot\|$ yields

$$\Pr\left(\|\xi\| - \mathbb{E}[\|\xi\|] > t\right) \leq e^{-\frac{t^2}{2}}.$$

Combining it with (29) we get $\Pr\left(\|\xi\| - \sqrt{d} > t\right) \leq e^{-\frac{t^2}{2}}$ or

$$\Pr\left(\|\xi\| > \tau\right) \leq e^{-\frac{(\tau-\sqrt{d})^2}{2}},\ \tau > \sqrt{d}. \tag{31}$$

Substitution of (31) into (30) provides

$$\mathbb{E}\left[I_{\{\|\xi\|>\tau\}}\|\xi\|\right] = \sqrt{d}e^{-\frac{(\tau-\sqrt{d})^2}{4}},\ \tau > \sqrt{d}. \tag{32}$$

Finally, using (32) we transform (28) into

$$\left\| \frac{1}{\varepsilon(2\pi)^{d/2}} \int_{\|v\|>\frac{\|\mu\|}{2\varepsilon}} \frac{\langle a, \varepsilon v + \mu\rangle^2}{\|\varepsilon v + \mu\|^2} v e^{-\frac{v^2}{2}} \mathrm{d}v \right\| \leq \frac{1}{\varepsilon}\sqrt{d}e^{-\frac{\left(\frac{\|\mu\|}{2\varepsilon}-\sqrt{d}\right)^2}{4}},\ \frac{\|\mu\|}{2\varepsilon} > \sqrt{d}. \tag{33}$$

Then we turn to the evaluation of the first term in (27).

$$\begin{aligned}
&\frac{1}{\varepsilon(2\pi)^{d/2}} \int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}} \frac{\langle a, \varepsilon v + \mu\rangle^2}{\|\varepsilon v + \mu\|^2} v e^{-\frac{\|v\|^2}{2}} \mathrm{d}v \\
&= \frac{1}{\varepsilon(2\pi)^{d/2}} \int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}} \frac{\langle a, \varepsilon v + \mu\rangle^2}{\|\varepsilon v + \mu\|^2} v e^{-\frac{\|v\|^2}{2}} \mathrm{d}v \\
&= \frac{1}{\varepsilon(2\pi)^{d/2}} \int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}} \psi(\varepsilon v + \mu) v e^{-\frac{\|v\|^2}{2}} \mathrm{d}v
\end{aligned} \tag{34}$$

We proceed by using the Taylor expansion with the remainder term in Lagrange form.

$$\psi_a(\varepsilon v+\mu)v=\psi_a(\mu)v+\langle\nabla\psi_a(\mu),\varepsilon v\rangle v+\frac{1}{2}\left[\left(\varepsilon v^\top\right)H_\psi(\mu+\varepsilon v')\left(\varepsilon v\right)\right]v, \tag{35}$$

for $v'=\alpha v, 0\leq\alpha\leq 1$.

Symmetry considerations provides first term vanishing after integration over $\left\{\|v\|<\frac{\|\mu\|}{2\varepsilon}\right\}$. Symmetry also simplifies the second term of (35) the in following way

$$\begin{aligned}
&\frac{1}{\varepsilon(2\pi)^{d/2}}\left(\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}\langle\nabla\psi_a(\mu),\varepsilon v\rangle v e^{-\frac{\|v\|^2}{2}}\mathrm{d}v\right)_i\\
&=\frac{1}{(2\pi)^{d/2}}\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}\sum_{j=1}^{d}\nabla\psi_a(\mu)_j v_j v_i e^{-\frac{\|v\|^2}{2}}\mathrm{d}v\\
&=\frac{1}{(2\pi)^{d/2}}\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}\nabla\psi_a(\mu)_i v_i v_i e^{-\frac{\|v\|^2}{2}}\mathrm{d}v\\
&=\frac{\nabla\psi_a(\mu)_i}{(2\pi)^{d/2}}\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}v_1^2 e^{-\frac{\|v\|^2}{2}}\mathrm{d}v\\
&=\frac{\nabla\psi_a(\mu)_i}{(2\pi)^{d/2}}\frac{1}{d}\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}\|v\|^2 e^{-\frac{\|v\|^2}{2}}\mathrm{d}v.
\end{aligned}\tag{36}$$

We apply Laurent-Massart inequalities (Laurent & Massart, 2000) to construct the lower bound of the integral in (36).

$$\begin{aligned}
&\Pr\left(\|v\|^2-d\leq 2\sqrt{dx}+2x\right)\leq e^{-x},\\
&\Pr\left(d-\|v\|^2\geq 2\sqrt{dx}\right)\leq e^{-x}.
\end{aligned}\tag{37}$$

Assuming $x=1$ and $\frac{\|\mu\|}{2\varepsilon}\geq\sqrt{d+2\sqrt{d}+2}$ we get

$$\begin{aligned}
&\frac{1}{(2\pi)^{d/2}}\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}\|v\|^2 e^{-\frac{\|v\|^2}{2}}\mathrm{d}v\\
&\geq\frac{1}{(2\pi)^{d/2}}\int_{d-2\sqrt{d}<\|v\|^2<d+2\sqrt{d}+2}\|v\|^2 e^{-\frac{\|v\|^2}{2}}\mathrm{d}v\\
&\geq(d-2\sqrt{d})\Pr\left(d-2\sqrt{d}<\|v\|^2<d+2\sqrt{d}+2\right)\\
&\geq(d-2\sqrt{d})(1-2e^{-1})\\
&\geq\frac{(d-2\sqrt{d})}{4}
\end{aligned}\tag{38}$$

Combining (38) and (36) we derive

$$\begin{aligned}
\left\|\frac{1}{(2\pi)^{d/2}}\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}\langle\nabla\psi_a(\mu),v\rangle v e^{-\frac{v^2}{2}}\mathrm{d}v\right\| &\geq \frac{\|\nabla\psi_a(\mu)\|}{d}\frac{(d-2\sqrt{d})}{4}\\
&= \|\nabla\psi_a(\mu)\|\left(\frac{1}{4}-\frac{1}{2\sqrt{d}}\right)\\
&\geq \frac{\|\nabla\psi_a(\mu)\|}{8}
\end{aligned}\tag{39}$$

To justify the last transition we narrow our analysis to the case $d \geq 16$.

Then we examine the last term of Taylor Expansion (36) using (23).

$$\begin{aligned}\left\|\frac{1}{2}\left[\left(\varepsilon v^{\top}\right) H_{\psi}(\mu+\varepsilon v')\left(\varepsilon v\right)\right] v\right\| &\leq \frac{\varepsilon^2}{2}\|v\|^3 \|H_{\psi}(\mu+\varepsilon v')\| \\ &\overset{(23)}{\leq} \frac{\varepsilon^2}{2}\|v\|^3 \frac{20}{\|\mu+\varepsilon v'\|^2}.\end{aligned}$$

Using norm convexity we conclude

$$\begin{aligned}&\left\|\frac{1}{\varepsilon(2\pi)^{d/2}}\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}\frac{1}{2}\left[\left(\varepsilon v^{\top}\right) H_{\psi}(\mu+\varepsilon v')\left(\varepsilon v\right)\right] v e^{-\frac{\|v\|^2}{2}}\mathrm{d}v\right\| \\ &\leq \frac{1}{\varepsilon(2\pi)^{d/2}}\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}\left\|\frac{1}{2}\left[\left(\varepsilon v^{\top}\right) H_{\psi}(\mu+\varepsilon v')\left(\varepsilon v\right)\right] v\right\| e^{-\frac{\|v\|^2}{2}}\mathrm{d}v \\ &\leq \frac{1}{\varepsilon(2\pi)^{d/2}}\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}\frac{\varepsilon^2}{2}\|v\|^3\|\frac{20}{\|\mu+\varepsilon v'\|^2}e^{-\frac{\|v\|^2}{2}}\mathrm{d}v \\ &\leq \frac{1}{\varepsilon(2\pi)^{d/2}}\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}\frac{\varepsilon^2}{2}\|v\|^3\|\frac{80}{\|\mu\|^2}e^{-\frac{\|v\|^2}{2}}\mathrm{d}v \\ &\leq \frac{40\varepsilon}{\|\mu\|^2(2\pi)^{d/2}}\int_{\|v\|<\frac{\|\mu\|}{2\varepsilon}}\|v\|^3 e^{-\frac{\|v\|^2}{2}}\mathrm{d}v \\ &\leq \frac{40\varepsilon}{\|\mu\|^2(2\pi)^{d/2}}\int_{\mathbb{R}^d}\|v\|^3 e^{-\frac{\|v\|^2}{2}}\mathrm{d}v \\ &\leq \frac{40\varepsilon}{\|\mu\|^2}d^{\frac{3}{2}}. \end{aligned} \tag{40}$$

The last transition utilizes Cauchy-Schwarz inequality and already mentioned values of $\mathbb{E}[\|v\|] \leq \sqrt{d}$ and $\mathbb{E}[\|v\|^2] = d$. Finally, using triangle inequality as $\|a+b\| \geq \|a\| - \|b\|$ we combine (27), (33), (39) and (40) to reach

$$\|\nabla F_a(\mu)\| \geq \frac{\|\nabla\psi_a(\mu)\|}{8} - \frac{40\varepsilon}{\|\mu\|^2}d^{\frac{3}{2}} - \frac{\sqrt{d}}{\varepsilon}e^{-\frac{\left(\frac{\|\mu\|}{2\varepsilon}-\sqrt{d}\right)^2}{4}}. \tag{41}$$

After that we can continue the recursion (26) applying (41).

$$\begin{aligned}&\mathbb{E}_{v^{t+1}\sim\mathcal{N}(\mu^{t+1},\varepsilon I)}\left[C^{t+1}\middle|\mathcal{F}^t\right] \geq \mathbb{E}_{v^t\sim\mathcal{N}(\mu^t,\varepsilon I)}\left[C^t\middle|\mathcal{F}^{t-1}\right] - 2a^t \\ &+\left(\gamma_\mu^t - \frac{\tilde{L}{\gamma_\mu^t}^2}{2}\right)\left(\frac{\|\nabla\psi_{\frac{\nabla f(x^t)}{\|\nabla f(x^t)\|}}(\mu^t)\|}{8} - \frac{40\varepsilon}{\|\mu^t\|^2}d^{\frac{3}{2}} - \frac{\sqrt{d}}{\varepsilon}e^{-\frac{\left(\frac{\|\mu^t\|}{2\varepsilon}-\sqrt{d}\right)^2}{4}}\right)^2. \end{aligned} \tag{42}$$

We begin the evaluation of (42) by requiring

$$\begin{cases}\frac{\|\nabla\psi_{\frac{\nabla f(x^t)}{\|\nabla f(x^t)\|}}(\mu^t)\|}{24} \geq \frac{40\varepsilon}{\|\mu^t\|^2}d^{\frac{3}{2}}, \\ \frac{\|\nabla\psi_{\frac{\nabla f(x^t)}{\|\nabla f(x^t)\|}}(\mu^t)\|}{24} \geq \frac{\sqrt{d}}{\varepsilon}e^{-\frac{\left(\frac{\|\mu^t\|}{2\varepsilon}-\sqrt{d}\right)^2}{4}}, \\ \gamma_\mu^t \geq \tilde{L}{\gamma_\mu^t}^2. \end{cases} \tag{43}$$

To start the evaluation of conditions (43) we introduce the angle β^t between the μ^t and $\nabla f(x^t)$.

$$\begin{aligned}\|\nabla\psi_{\overline{\nabla f(x^t)}}(\mu)\| &= \left\|\frac{2\langle\nabla f(x^t),\mu\rangle}{\mu^4}\{\nabla f(x^t)\|\mu\|^2-\mu\langle\nabla f(x^t),\mu\rangle\}\right\| \\ &\geq \frac{2}{\|\mu\|^3}\cos(\beta^t)\left(\|\mu\|^2-\|\mu\|^2\cos(\beta^t)\right) \\ &\geq \frac{2}{\|\mu\|}\cos(\beta^t)(1-\cos(\beta^t))\end{aligned} \tag{44}$$

Hence, choosing $\varepsilon \leq \frac{d^{-\frac{3}{2}}\cos(\beta^t)(1-\cos(\beta^t))\|\mu^t\|}{480}$ we satisfy the first condition. This choice of ε also justifies second restriction in (43). To satisfy the last inequality of (43) we choose $\gamma_\mu^t \leq \frac{\|\mu^t\|^2}{640} \leq \frac{1}{L}$. Then, (42) transforms into

$$\begin{aligned}\mathbb{E}_{v^{t+1}\sim\mathcal{N}(\mu^{t+1},\varepsilon I)}\left[C^{t+1}\middle|\mathcal{F}^t\right] &\geq \mathbb{E}_{v^t\sim\mathcal{N}(\mu^t,\varepsilon I)}\left[C^t\middle|\mathcal{F}^{t-1}\right]-2a^t+\frac{\gamma_\mu^t}{2}\frac{\|\nabla\psi(\mu^t)\|^2}{24^2} \\ &= \mathbb{E}_{v^t\sim\mathcal{N}(\mu^t,\varepsilon I)}\left[C^t\middle|\mathcal{F}^{t-1}\right]-2a^t \\ &\quad +\gamma_\mu^t\frac{2\cos^2(\beta^t)(1-\cos(\beta^t))^2}{24^2\|\mu^t\|^2}.\end{aligned} \tag{45}$$

After that we require

$$2a^t \leq \gamma_\mu^t\frac{\cos^2(\beta^t)(1-\cos(\beta^t))^2}{24^2\|\mu^t\|^2},$$

so that $2a^t$ is no larger than one half of the positive term in (45). Recalling the definition $a^t = 4\frac{L\gamma_x^t}{2}\sqrt{1-\left(\frac{L\gamma_x^t}{2}\right)^2}$ and using $\sqrt{1-\left(\frac{L\gamma_x^t}{2}\right)^2}\leq 1$, it is sufficient to choose

$$\gamma_x^t \leq \frac{\gamma_\mu^t\cos^2(\beta^t)(1-\cos(\beta^t))^2}{2^8\cdot 3^2L\|\mu^t\|^2}.$$

Finally we reach

$$\mathbb{E}_{v^{t+1}\sim\mathcal{N}(\mu^{t+1},\varepsilon I)}\left[C^{t+1}\middle|\mathcal{F}^t\right] \geq \mathbb{E}_{v^t\sim\mathcal{N}(\mu^t,\varepsilon I)}\left[C^t\middle|\mathcal{F}^{t-1}\right]+\gamma_\mu^t\left(\frac{\cos(\beta^t)(1-\cos(\beta^t))}{24\|\mu^t\|}\right)^2 \tag{46}$$

□

Corollary C.2. *In the setting of Theorem 3.5 for iterations of the Algorithm 1 with* $\varepsilon = \frac{d^{-\frac{3}{2}}\cos(\beta^t)(1-\cos(\beta^t))M}{480}, \gamma_\mu^t = \frac{\|\mu^t\|^2}{640}, \gamma_x^t = \frac{\cos^2(\beta^t)(1-\cos(\beta^t))^2}{3\cdot 2^{12}\cdot 5^2L}$ *the following inequality holds:*

$$\mathbb{E}\left[C^{t+1}\middle|\mathcal{F}^t\right] \geq \mathbb{E}\left[C^t\middle|\mathcal{F}^{t-1}\right]+\frac{\cos^2(\beta^t)(1-\cos(\beta^t))^2}{3\cdot 2^{10}\cdot 5^2}. \tag{47}$$

If additionally $\delta \leq \cos(\beta^t) \leq 1-\delta$ *with* $\delta \in \left(0;\frac{1}{2}\right)$ *and* $\varepsilon = \frac{d^{-\frac{3}{2}}\delta M}{960}, \gamma_x^t = \frac{\delta^2}{3\cdot 2^{14}\cdot 5^2L}$, *then*

$$\mathbb{E}\left[C^{t+1}\middle|\mathcal{F}^t\right] \geq \mathbb{E}\left[C^t\middle|\mathcal{F}^{t-1}\right]+\frac{\delta^2}{3\cdot 2^{12}\cdot 5^2}. \tag{48}$$

Proof. The proof follows by direct substitution into the result of Theorem 3.5. □

Lemma 3.6. *Let in the setting of Theorem 3.5* $\delta \leq \cos(\beta^0) \leq 1-\delta$ *with* $\delta \in \left(0;\frac{1}{2}\right)$ *and* $\varepsilon = \frac{d^{-\frac{3}{2}}\delta M}{960}, \gamma_x^t = \frac{\delta^2}{3\cdot 2^{14}\cdot 5^2L}$. *Then* $\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]$ *increase monotonically to the value* $\cos\left(\frac{\delta}{32d}+L\gamma_x^t+\arccos(1-\delta)\right)^2(1-e^{-1})$ *and then never decrease below.*

Proof. To justify the first part of statement related to $\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]$ monotonically growth it sufficient to show, that the angle β^t will decrease until $\cos(\beta^t)$ reaches $1-\delta$. We note that C^t dynamic analysis closely related to the evaluation of the angle β^t. We proceed by assuming $\delta \leq \cos(\beta^t) \leq q - \delta$ considering

$$\begin{aligned}\|\overline{\mu^{t+1}} - \overline{\nabla f(x^{t+1})}\| &\overset{(11)}{\leq} \|\overline{\mu^{t+1}} - \overline{\nabla f(x^t)}\| + a^t \\ &= \|\overline{\mu^t + \gamma_\mu^t \nabla_\mu \mathbb{E}_t\left[C^t\right]\Big|_{\mu=\mu^t}} - \overline{\nabla f(x^t)}\| + a^t \\ &= \|\overline{\mu^t + \gamma_\mu^t F_{\overline{\nabla f(x^t)}}(\mu^t)} - \overline{\nabla f(x^t)}\| + a^t. \end{aligned} \tag{49}$$

Then we recall decomposition (27) and that one integral there contains (36). We remind that $\nabla\psi(\mu) = \frac{2\langle\nabla f(x^t),\mu\rangle}{\mu^4}\{\nabla f(x^t)\|\mu\|^2 - \mu\langle\nabla f(x^t),\mu\rangle\}$ which implies its collinearity to $\overline{\mu^t} - \overline{\nabla f(x^t)}$. Using the same bounds as in (41) we continue derivation.

$$\begin{aligned}\|\overline{\mu^{t+1}} - \overline{\nabla f(x^{t+1})}\| \leq\ & \|\overline{\mu^t} - \overline{\nabla f(x^t)}\| + a^t \\ & -\frac{1}{2\|\mu^t\|}\gamma_\mu^t\left(\frac{\|\nabla\psi(\mu^t)\|}{8} - \frac{40\varepsilon}{\|\mu^t\|^2}d^{\frac{3}{2}} - \frac{\sqrt{d}}{\varepsilon}e^{-\frac{\left(\frac{\|\mu^t\|}{2\varepsilon}-\sqrt{d}\right)^2}{4}}\right). \end{aligned} \tag{50}$$

After that we note $\delta \leq \cos(\beta^t) \leq 1 - \delta$ with $\delta \in \left(0; \frac{1}{2}\right)$ implies both $\cos(\beta^t)$ and $1 - \cos(\beta^t)$ greater than or equal to δ and one of them greater then $\frac{1}{2}$, which yields $\cos(\beta^t)(1 - \cos(\beta^0)) \geq \frac{\delta}{2}$. Using it along with a choice of parameters in Theorem 3.5 (see transitions (42) to (46)) we conclude

$$\|\overline{\mu^{t+1}} - \overline{\nabla f(x^{t+1})}\| \leq \|\overline{\mu^t} - \overline{\nabla f(x^t)}\|, \tag{51}$$

Which means that once the condition $\delta \leq \cos(\beta^t) \leq 1 - \delta$ holds $\beta^t \geq \beta^{t+1}$. Hence, $\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right]$ growth monotonically until $\cos(\beta^t)$ exceeds $(1 - \delta)$. After that β^t can not fall below $\delta_0 := L\gamma_x^t + \arccos(1 - \delta)$, since the direction of the gradient changes less then $2\arcsin\left(\frac{L\gamma_x^t}{2}\right)$ per step, as was shown in the beginning of the Theorem 3.5 proof, and β^t can not increase further if $\delta \leq \cos(\beta^t) \leq 1 - \delta$ as shown above.

Finally, we evaluate

$$\begin{aligned}\mathbb{E}\left[C^t|\mathcal{F}^{t-1}\right] &= \frac{1}{(2\pi\varepsilon^2)^{d/2}}\int_{\mathbb{R}^d}\left\langle\overline{\nabla f(x^t)}, \overline{u}\right\rangle^2 e^{-\frac{\|u-\mu^t\|^2}{2\varepsilon^2}}\mathrm{d}u \\ &= \frac{1}{(2\pi\varepsilon^2)^{d/2}}\int_{\mathbb{R}^d}\left\langle\overline{\nabla f(x^t)}, \overline{s-\mu^t}\right\rangle^2 e^{-\frac{\|s\|^2}{2\varepsilon^2}}\mathrm{d}s \\ &= \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\left\langle\overline{\nabla f(x^t)}, \overline{\varepsilon v-\mu^t}\right\rangle^2 e^{-\frac{\|v\|^2}{2}}\mathrm{d}v. \end{aligned} \tag{52}$$

We apply Laurent-Massart inequality (Laurent & Massart, 2000) (37) with $x = 1$ to justify $\Pr\left(d - \|v\|^2 \geq 2\sqrt{d}\right) \leq e^{-1}$. Then we can bound the integral in (52).

$$\begin{aligned}&\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\left\langle\overline{\nabla f(x^t)}, \overline{\varepsilon v-\mu^t}\right\rangle^2 e^{-\frac{\|v\|^2}{2}}\mathrm{d}u \\ &\geq \frac{1}{(2\pi)^{d/2}}\int_{\|v\|^2\leq 2\sqrt{d}}\left\langle\overline{\nabla f(x^t)}, \overline{\varepsilon v-\mu^t}\right\rangle^2 e^{-\frac{\|v\|^2}{2}}\mathrm{d}u \\ &\geq \frac{1}{(2\pi)^{d/2}}\int_{\|v\|^2\leq 2\sqrt{d}}\cos\left(\frac{\delta}{32d} + L\gamma_x^t + \arccos(1-\delta)\right)^2 e^{-\frac{\|v\|^2}{2}}\mathrm{d}u \\ &\geq \cos\left(\frac{\delta}{32d} + L\gamma_x^t + \arccos(1-\delta)\right)^2(1 - e^{-1}). \end{aligned} \tag{53}$$

It concludes the proof. □

C.1 Initialization with no prior information

Lemma 3.6 naturally leads to two complementary settings. The first is formalized in Lemma C.3 and corresponds to initializing the policy without any prior knowledge of the target function f.

Lemma C.3. *Let Assumptions 3.1, 3.2 be satisfied and $d \geq 16$. Then for iterations of Algorithm 1 with $\delta \leq \cos(\beta^0) \leq 1-\delta$, $\delta \in \left(0; \frac{1}{2}\right)$ and $\varepsilon = \frac{d^{-\frac{3}{2}}\delta M}{960}, \gamma_x^t = \frac{\delta^2}{3\cdot 2^{14}\cdot 5^2 L}, \gamma_\mu^t = \frac{\|\mu^t\|^2}{640}$ the following bound holds:*

$$\sum_{t=0}^{T-1} \frac{1}{T}\mathbb{E}\left[B^t \|\nabla f(x^t)\|^2\right] \leq 2\frac{f(x^0)-f(x^*)}{T\gamma_x^t}, \tag{54}$$

where $B^t = \min\left\{\mathbb{E}[C^0] + t\cdot\frac{\delta^2}{3\cdot 2^{12}\cdot 5^2}; \cos\left(\frac{\delta}{32d} + L\gamma_x^t + \arccos(1-\delta)\right)^2 (1-e^{-1})\right\}$ and $M = \min_t\{\|\mu^t\|\}$.

Proof. We start from the result of Lemma 3.3. The choice of $\gamma_x^t = \frac{\delta^2}{3\cdot 2^{14}\cdot 5^2 L}$ guarantees

$$\gamma_x^t - \frac{L{\gamma_x^t}^2}{2} \geq \frac{\gamma_x^t}{2}. \tag{55}$$

Subsequently, Theorem 3.5 and Lemma 3.6 justifies

$$\mathbb{E}\left[C^t | \mathcal{F}^{t-1}\right] \geq B^t. \tag{56}$$

Substituting (55) and (56) into Lemma 3.3 and rearranging the terms completes the proof. □

Notably, in this setting, substituting $\mathbb{E}\left[C^t \middle| \mathcal{F}^{t-1}\right]$ behaviour described in Lemma 3.6 yields a non-standard convergence criterion, namely convergence in terms of a *weighted* average of squared gradient norms. The weights increase with the iteration index until reaching a saturation level. Such a result may be more transparent for practical scenarios such as memory-efficient LLM fine-tuning, where the algorithm outputs the final iterate rather than an averaged solution.

We also note that Lemma C.3 relies on a uniform-in-iteration lower bound of the form $\cos(\beta^t) > \delta$, which results in a fixed step size γ_x^t and a constant additive increase of $\mathbb{E}\left[C^t \middle| \mathcal{F}^{t-1}\right]$ per iteration. However, as the policy improves, $\cos(\beta^t)$ increases, which simultaneously accelerates the growth of $\mathbb{E}\left[C^t \middle| \mathcal{F}^{t-1}\right]$ and allows for larger admissible step sizes γ_x^t. Consequently, the uniform bound becomes loose after a certain number of iterations and does not fully capture these acceleration effects in the theoretical rate. Along similar lines, bounds of the type used in the proof of Lemma 3.4 yield average values on the order of $\frac{1}{d}$ both for δ^2 (consequently for γ_x^t) and $\mathbb{E}[C^0]$. It introduces a factor d^2 in the rate of Lemma C.3 during the initial phase and results in growth of $\mathbb{E}\left[C^t \middle| \mathcal{F}^{t-1}\right]$ on the order of $\frac{1}{d}$ per iteration.

C.2 Initialization leveraging gradient direction at the initial point

Lemma 3.7. *Let Assumptions 3.1, 3.2 be satisfied, $d \geq 16$, and $\cos(\beta^0) \geq \frac{1}{4}$. Then for iterations of Algorithm 1 with $\varepsilon = \frac{d^{-\frac{3}{2}}M}{3840}, \gamma_x^t = \frac{1}{3\cdot 2^{18}\cdot 5^2 L}, \gamma_\mu^t = \frac{\|\mu^t\|^2}{640}$, where $M = \min_t\{\|\mu^t\|\}$, the following bound holds:*

$$\mathbb{E}[\|\nabla f(\overline{x}^T)\|^2] \quad \leq \quad 10\frac{f(x^0)-f(x^*)}{T\gamma_x}.$$

Proof. We first note that (55) remains true. Additionally, given $\cos(\beta^0) \geq \frac{1}{4}$ in the setting of Lemma 3.6 with $\delta = \frac{1}{4}$ Algorithm 1 is already in the regime where $\mathbb{E}\left[C^t \middle| \mathcal{F}^{t-1}\right]$ fluctuates in a neighbourhood of 1 and attains values no smaller than $\cos\left(\frac{1}{128d} + L\gamma_x^t + \arccos\left(\frac{3}{4}\right)\right)^2 (1-e^{-1})$, which yields

$$\mathbb{E}\left[C^t \middle| \mathcal{F}^{t-1}\right] \geq \frac{1}{5}, \text{ for all } t. \tag{57}$$

Substituting (55) with (57) into the Lemma 3.3 and rearranging the terms we derive the desired result. □