# Zeroth-Order Optimization is Secretly Single-Step Policy Optimization

**Junbin Qiu** [1] **Zhengpeng Xie** [1] **Xiangda Yan** [2] **Yongjie Yang** [2] **Yao Shu** [1]

## Abstract

Zeroth-Order Optimization (ZOO) provides powerful tools for optimizing functions where explicit gradients are unavailable or expensive to compute. However, the underlying mechanisms of popular ZOO methods, particularly those employing randomized finite differences, and their connection to other optimization paradigms like Reinforcement Learning (RL) are not fully elucidated. This paper establishes a fundamental and previously unrecognized connection: ZOO with finite differences is equivalent to a specific instance of single-step Policy Optimization (PO). We formally unveil that the implicitly smoothed objective function optimized by common ZOO algorithms is identical to a single-step PO objective. Furthermore, we show that widely used ZOO gradient estimators, are mathematically equivalent to the REINFORCE gradient estimator with a specific baseline function, revealing the variance-reducing mechanism in ZOO from a PO perspective.Built on this unified framework, we propose ZoAR (*Zeroth-Order Optimization with Averaged Baseline and Query Reuse*), a novel ZOO algorithm incorporating PO-inspired variance reduction techniques: an averaged baseline from recent evaluations and query reuse analogous to experience replay. Our theoretical analysis further substantiates these techniques reduce variance and enhance convergence. Extensive empirical studies validate our theory and demonstrate that ZoAR significantly outperforms other methods in terms of convergence speed and final performance. Overall, our work provides a new theoretical lens for understanding ZOO and offers practical algorithmic improvements derived from its connection to PO.

[1]Hong Kong University of Science and Technology (Guangzhou) [2]Xiaomi Inc. Correspondence to: Yao Shu <yaoshu@hkust-gz.edu.cn>.

*Proceedings of the Workshop on Tiny Titans: The next wave of On-Device Learning for Foundational Models @ ICML 2025*. 

## 1. Introduction

Zeroth-Order Optimization (ZOO) addresses the task of optimizing objectives $F(\boldsymbol{\theta}) = \mathbb{E}_{\xi}[f(\boldsymbol{\theta}; \xi)]$ using only function evaluations, bypassing the need for explicit gradients (Nesterov & Spokoiny, 2017; Ghadimi & Lan, 2013). This paradigm is essential in numerous domains where gradients are intractable, computationally prohibitive, or simply unavailable, such as hyperparameter optimization (Gu et al., 2021), derivative-free reinforcement learning (Salimans et al., 2017), communication-efficient federated learning (Shu et al., 2024), black-box adversarial attacks (Shu et al., 2023; 2025b), prompt optimization (Hu et al., 2024; Zhan et al., 2024), and memory-efficient finetuning for large language models (LLMs) (Malladi et al., 2023; Zhang et al., 2024). A dominant strategy within ZOO involves estimating gradients via randomized finite differences, which implicitly optimize a smoothed surrogate $F_{\mu}(\boldsymbol{\theta})$ of the original objective $F(\boldsymbol{\theta})$ (Nesterov & Spokoiny, 2017; Shu et al., 2025b). A thorough discussion on the most related works of ZOO is in Appx. A. While foundational, these methods often suffer from high variance in their gradient estimates, potentially impeding convergence speed and solution quality. Furthermore, a deep theoretical understanding connecting these ZOO techniques to established principles in related fields like Reinforcement Learning (RL) remains underdeveloped. In parallel, Policy Optimization (PO) forms the bedrock of modern RL, seeking policy parameters $\boldsymbol{\theta}$ to maximize expected cumulative rewards $J(\boldsymbol{\theta})$ (Sutton et al., 1999; Sutton & Barto, 2018). Policy Gradient (PG) algorithms like REINFORCE (Williams, 1992) estimate $\nabla J(\boldsymbol{\theta})$ from trajectory rollouts. A crucial technique for stabilizing PG methods is baseline subtraction, which provably reduces gradient estimate variance and thereby accelerates learning (Sutton & Barto, 2018).

As the **first** primary contribution, this paper establishes a fundamental and *previously unrecognized* connection: *smoothed Zeroth-Order Optimization (ZOO) with finite differences is formally equivalent to a specific instance of single-step Policy Optimization (PO)*. We bridge these two fields, providing theoretical clarification for ZOO mechanisms (Sec. 3): First, we formally unveil that the smoothed objective $F_{\mu}(\boldsymbol{\theta})$ implicitly targeted by common ZOO methods is *identical* to a single-step PO objective $J(\boldsymbol{\theta})$ under a specific reward definition (Thm. 3.1). Second, we prove for

the first time that the standard Gaussian-smoothed ZOO gradient estimator is mathematically *equivalent* to the single-step REINFORCE estimator using the function value $f(\boldsymbol{\theta};\xi)$ as a baseline (Thm. 3.2). This novel interpretation recasts the standard ZOO baseline subtraction not merely as a finite-difference artifact, but as a principled variance reduction technique rooted in PO theory, revealing the variance-reducing mechanism in ZOO from a PO lens. Third, we further extend this foundational equivalence using importance sampling (Thm. 3.3), clarifying how ZOO estimators with alternative sampling distributions relate to weighted REINFORCE and optimize distinct smoothed objectives.

Building upon this newly established unified PO framework, our **second** primary contribution is ZoAR (*Zeroth-Order Optimization with Averaged Baseline and Query Reuse*) proposed in Sec. 4. ZoAR is the first to integrate two PO-inspired variance reduction techniques directly into conventional ZOO (see Sec. 4.1): *(a) Averaged Baseline:* Instead of the high-variance single-point estimate $f(\boldsymbol{\theta};\xi)$, ZoAR introduces an averaged baseline from recent function evaluations in a history buffer. This novel ZOO adaptation of the value function estimation in PO provides a more stable Monte Carlo estimate of the smoothed objective $F_\mu(\boldsymbol{\theta})$. *(b) Query Reuse:* ZoAR computes gradient estimates using all samples in the history buffer (analogous to the experience replay in PO), effectively increasing the batch size for gradient estimation without new queries per iteration, thus enhancing sample efficiency and mitigating variance. We further provide rigorous theoretical analysis in Appx. B to support the variance reduction effect of these two newly introduced PO-inspired techniques from the lens of ZOO theory and show the potentially improved convergence of ZoAR when variance dominates.

Our **third** contribution lies in comprehensive empirical validation (Sec. 5). We benchmark ZoAR against other ZOO baselines, across standard synthetic functions, a black-box adversarial attack task, and memory-efficient finetuning of LLMs. The results consistently show that ZoAR achieves significant improvements in convergence rate and final performance, validating the practical efficacy of leveraging these newly connected PO techniques for ZOO. Notably, substantial gains are observed even with our novel averaged baseline alone, highlighting its distinct effectiveness.

## 2. Preliminaries

This section introduces the necessary background on Zeroth-Order Optimization (ZOO) and Policy Optimization (PO) in Reinforcement Learning (RL), establishing the notation and core concepts used throughout the paper.

**Problem Setup.** We focus on the problem of minimizing a potentially non-convex objective function $F(\boldsymbol{\theta})$ defined as an expectation over a random variable $\xi$:

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^d} F(\boldsymbol{\theta}) \triangleq \mathbb{E}_\xi\left[f(\boldsymbol{\theta};\xi)\right] . \tag{1}$$

Here, $\boldsymbol{\theta}\in\mathbb{R}^d$ represents the $d$-dimensional parameter vector we aim to optimize, $f(\boldsymbol{\theta};\xi)$ is a scalar-valued loss function whose evaluation depends on both the parameters $\boldsymbol{\theta}$ and a random variable $\xi$. The defining characteristic of the Zeroth-Order (ZO) setting is the constraint that we can only access stochastic evaluations of the function value, $f(\boldsymbol{\theta};\xi)$, through a black-box oracle. Importantly, direct access to the gradient $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta};\xi)$ is assumed to be unavailable or computationally prohibitive. Throughout this paper, we use $\nabla$ to denote the gradient with respect to the parameters $\boldsymbol{\theta}$, i.e., $\nabla \equiv \nabla_{\boldsymbol{\theta}}$.

**Zeroth-Order Optimization.** To optimize (1) without explicit gradients, ZOO algorithms employ gradient estimators constructed solely from function evaluations. A prevalent technique is randomized finite differences. A common form of such an estimator, averaged over $K$ directions is:

$$\hat{\nabla} F(\boldsymbol{\theta}) \triangleq \frac{1}{K}\sum_{k=1}^{K} \frac{f(\boldsymbol{\theta}+\mu\mathbf{u}_k;\xi) - f(\boldsymbol{\theta};\xi)}{\mu}\mathbf{u}_k . \tag{2}$$

where $\{\mathbf{u}_k\}_{k=1}^K$ are i.i.d. random direction vectors, $\mu > 0$ is a small smoothing radius parameter, and $K \geq 1$ dictates the number of function queries used per gradient estimate (beyond the baseline evaluation $f(\boldsymbol{\theta};\xi)$). Standard choices for the distribution of $\mathbf{u}_k$ include:

(I) The standard multivariate Gaussian distribution $\mathbf{u}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ (Nesterov & Spokoiny, 2017).

(II) The uniform distribution over the unit sphere $\mathbf{u}_k \sim \mathrm{Unif}(\mathbb{S}^{d-1})$ (Flaxman et al., 2005a).

(III) The uniform distribution over the standard basis vectors $\mathbf{u}_k \sim \mathrm{Unif}(\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\})$ (Lian et al., 2016).

It is well-established that (2) is an unbiased gradient estimation of a smoothed approximation $F_\mu$ (defined as below) for the original objective $F(\boldsymbol{\theta})$ (Nesterov & Spokoiny, 2017; Shu et al., 2025b). This means that ZOO with estimator (2) is in fact implicitly optimizing the smoothed objective $F_\mu$.

$$F_\mu(\boldsymbol{\theta}) \triangleq \mathbb{E}_{\mathbf{u}}\left[F(\boldsymbol{\theta}+\mu\mathbf{u})\right] = \mathbb{E}_{\mathbf{u}}\left[\mathbb{E}_\xi\left[f(\boldsymbol{\theta}+\mu\mathbf{u};\xi)\right]\right] . \tag{3}$$

**Policy Optimization and REINFORCE.** In policy optimization, the objective is typically to find the parameters $\boldsymbol{\theta}$ of a stochastic policy $\pi_{\boldsymbol{\theta}}(a|s)$ that maximize the expected cumulative reward. Let us consider the standard episodic setting. The objective function, $J(\boldsymbol{\theta})$, is the expected total discounted reward obtained by executing the policy $\pi_{\boldsymbol{\theta}}$

starting from an initial state distribution $\rho_0(s_0)$:

$$J(\boldsymbol{\theta}) \triangleq \mathbb{E}_{\tau\sim p_{\boldsymbol{\theta}}(\tau)}\left[\sum_{t=0}^{T-1}\gamma^t R(S_t,A_t)\right]$$
$$=\mathbb{E}_{S_0\sim\rho_0,A_t\sim\pi_{\boldsymbol{\theta}}(\cdot|S_t),S_{t+1}\sim P(\cdot|S_t,A_t)}\left[\sum_{t=0}^{T-1}\gamma^t R(S_t,A_t)\right]. \quad (4)$$

Here, $\tau = (S_0, A_0, R_0, \ldots, S_{T-1}, A_{T-1}, R_{T-1}, S_T)$ represents a trajectory (or episode) of states $S_t$, actions $A_t$, and rewards $R_t = R(S_t, A_t)$. The trajectory distribution $p_{\boldsymbol{\theta}}(\tau)$ is induced by the policy $\pi_{\boldsymbol{\theta}}$ and the transition dynamics $P(S_{t+1}|S_t, A_t)$ of environment. $\gamma \in [0, 1]$ is the discount factor, and $T$ is the episode horizon (which can be finite or infinite). Note that while policy optimization typically involves maximization, we can frame it as minimization by considering the negative reward (cost), i.e., minimizing $-J(\boldsymbol{\theta})$, to align with the optimization setup in (1).

Policy Gradient methods are a class of algorithms designed to optimize $J(\boldsymbol{\theta})$ by estimating its gradient $\nabla J(\boldsymbol{\theta})$ and performing gradient ascent (or descent on $-J(\boldsymbol{\theta})$). The Policy Gradient Theorem (Sutton et al., 1999) provides the analytical form of this gradient and a widely used policy gradient is derived from the REINFORCE (w/ baseline) algorithm (Williams, 1992):

$$\nabla J(\boldsymbol{\theta}) = \mathbb{E}_{\tau\sim p_{\boldsymbol{\theta}}(\tau)}\left[\sum_{t=0}^{T-1}\nabla \ln \pi_{\boldsymbol{\theta}}(A_t|S_t)\,(G_t - b(S_t))\right] \quad (5)$$

where $G_t = \sum_{t'=t}^{T-1}\gamma^{t'-t}R(S_{t'}, A_{t'})$ represents the discounted return-to-go from time step $t$ and the state-dependent baseline $b(S_t)$ is applied for variance reduction.

## 3. A Policy Optimization Framework for Zeroth-Order Optimization

Building on the preliminaries in Sec. 2, this section formally establishes the connection between Zeroth-Order Optimization (ZOO) and Policy Optimization (PO). We demonstrate that the ZOO problem can be precisely framed as a single-step PO problem (Sec. 3.1). Furthermore, we show that common ZOO gradient estimators are equivalent to specific instances of the REINFORCE algorithm with a baseline (Sec. 3.2 & Sec. 3.3).

### 3.1. Equivalence of Objectives in ZOO and PO

We begin by demonstrating the equivalence between the objective function implicitly optimized by many ZOO methods, i.e., $F_\mu(\boldsymbol{\theta})$ in (3), and a specific instance of the PO objective. Formally, consider the standard PO objective from (4) in a simplified, single-step episodic setting (i.e., $T = 1, \gamma = 0$). In this scenario, the agent takes a single action $\mathbf{x}$ sampled from a policy $\pi_{\boldsymbol{\theta}}(\mathbf{x})$, and receives a reward based on this action. To align with the minimization problem (1), we define the reward as the negative function value, $R_0 = -F(\mathbf{x})$. The PO objective is then to minimize the expected negative reward:

$$J(\boldsymbol{\theta}) \triangleq \mathbb{E}_{\mathbf{x}\sim\pi_{\boldsymbol{\theta}}(\mathbf{x})}\left[F(\mathbf{x})\right] = \mathbb{E}_{\mathbf{x}\sim\pi_{\boldsymbol{\theta}}(\mathbf{x})}\left[\mathbb{E}_\xi\left[f(\boldsymbol{\theta};\xi)\right]\right]. \quad (6)$$

The connection between the ZOO smoothed objective $F_\mu(\boldsymbol{\theta})$ defined in (3) and this single-step PO objective $J(\boldsymbol{\theta})$ defined in (6) is formalized below (proof in Appx. C.1).

**Theorem 3.1** (Objective Equivalence)**.** *Let the policy $\pi_{\boldsymbol{\theta}}(\mathbf{x})$ be defined via the reparameterization $\mathbf{x} = \boldsymbol{\theta} + \mu\mathbf{u}$, where $\mathbf{u}$ is a random vector drawn from a distribution $p(\mathbf{u})$ independent of $\boldsymbol{\theta}$. Then, the single-step PO objective $J(\boldsymbol{\theta})$ defined in (6) is identical to the ZOO smoothed objective $F_\mu(\boldsymbol{\theta})$ defined in (3) using the same distribution $p(\mathbf{u})$, i.e.,*

$$J(\boldsymbol{\theta}) = F_\mu(\boldsymbol{\theta}).$$

**Remark.** Thm. 3.1 establishes that optimizing the smoothed function $F_\mu(\boldsymbol{\theta})$, a standard practice in ZOO theory, is equivalent to optimizing a single-step RL objective $J(\boldsymbol{\theta})$ where the policy samples perturbations around the current parameters $\boldsymbol{\theta}$. This equivalence allows us to leverage concepts and algorithms from PO to understand and potentially improve ZOO methods (see Sec. 4). The choice of the smoothing distribution $p(\mathbf{u})$ in ZOO corresponds to the choice of the exploration strategy (policy structure) in this PO context. To the best of our knowledge, this is the first to explicitly interpret the ZOO smoothed objective through this specific PO lens.

### 3.2. Gaussian Smoothing as Single-Step REINFORCE w/ Baseline

We now demonstrate that the widely used Gaussian-smoothed ZOO gradient estimator is equivalent to a specific instance of the REINFORCE w/ baseline algorithm. Let the smoothing distribution be the standard multivariate Gaussian, $p(\mathbf{u}) = \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. The corresponding policy $\pi_{\boldsymbol{\theta}}(\mathbf{x})$ samples $\mathbf{x} = \boldsymbol{\theta} + \mu\mathbf{u}$, which means $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\theta}, \mu^2\mathbf{I}_d)$. To minimize $F_\mu(\boldsymbol{\theta}) = J(\boldsymbol{\theta})$, We apply the REINFORCE w/ baseline algorithm using the policy gradient theorem (5). For our single-step case ($T = 1$), the policy gradient gives:

$$\nabla J(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}\sim\pi_{\boldsymbol{\theta}}(\mathbf{x})}\left[\nabla \ln \pi_{\boldsymbol{\theta}}(\mathbf{x})\,(\mathbb{E}_\xi[f(\mathbf{x};\xi)] - b)\right], \quad (7)$$

where $b$ is a baseline that is independent of the specific sample $\mathbf{x}$. Particularly, for the Gaussian policy $\pi_{\boldsymbol{\theta}}(\mathbf{x}) = \mathcal{N}(\boldsymbol{\theta}, \mu^2\mathbf{I}_d)$, we have $\nabla \ln \pi_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{\mathbf{x}-\boldsymbol{\theta}}{\mu^2}$. Substituting this into (7) gives:

$$\nabla J(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}\sim\pi_{\boldsymbol{\theta}}(\mathbf{x})}\left[\frac{\mathbf{x}-\boldsymbol{\theta}}{\mu^2}\,(\mathbb{E}_\xi[f(\mathbf{x};\xi)] - b)\right]. \quad (8)$$

In practice, the expectations are approximated using Monte Carlo sampling. Let $b = \mathbb{E}_\xi[b(\xi)]$, we sample $\mathbf{x}_k$ from $\pi_{\boldsymbol{\theta}}(\mathbf{x})$ to estimate the outer expectation and $\xi$ to estimate the inner expectation. A common stochastic gradient estimator based on $K$ samples is then:

$$\hat{\nabla}_{\text{GS}} J(\boldsymbol{\theta}) \triangleq \frac{1}{K}\sum_{k=1}^{K} \frac{\mathbf{x}_k - \boldsymbol{\theta}}{\mu^2}\left(f(\mathbf{x}_k;\xi) - b(\xi)\right) . \quad (9)$$

The connection between the standard Gaussian-smoothed ZOO gradient estimator from (2) and the REINFORCE gradient estimator (9) is formalized below (proof in Appx. C.2).

**Theorem 3.2** (Gradient Estimator Equivalence for Gaussian)**.** *Let $\pi_{\boldsymbol{\theta}}(\mathbf{x}) = \mathcal{N}(\boldsymbol{\theta}, \mu^2\mathbf{I}_d)$ and $b(\xi) = f(\boldsymbol{\theta};\xi)$ in (9). Then, the REINFORCE gradient estimator (9) is identical to the Gaussian-smoothed ZOO gradient estimator (2), i.e.,*

$$\hat{\nabla}_{\text{GS}} J(\boldsymbol{\theta}) = \hat{\nabla} F(\boldsymbol{\theta}) .$$

**Remark.** Thm. 3.2 provides the first explicit interpretation of the common ZOO gradient estimator (2) from a novel PO lens. Specifically, it reveals that Gaussian-smoothed ZOO estimator can be interpreted as REINFORCE gradient estimator with gaussian policy. Moreover, it unveils that the subtraction of $f(\boldsymbol{\theta};\xi)$ in conventional ZOO is not merely a result from the first-order Taylor polynomial but corresponds precisely to using a baseline in the REINFORCE algorithm. This baseline is known to reduce the variance of the gradient estimate without introducing bias (Sutton & Barto, 2018). This perspective not only aligns with but also provides a theoretical support for observations in works like (Salimans et al., 2017) where similar estimators were used in the context of evolution strategies, highlighting the variance reduction benefit without explicitly linking it to the REINFORCE w/ baseline mechanism.

### 3.3. Generalization Through Importance Sampling

The previous section only established the equivalence for Gaussian smoothing, whereas ZOO methods can also apply other sampling distributions for $\mathbf{u}_k$, like the uniform distribution over the unit sphere or coordinate directions mentioned in Sec. 2. We hence generalize our PO perspective to encompass these cases using importance sampling (IS) in this section.

Suppose we still consider the objective $J(\boldsymbol{\theta})$ with the Gaussian policy $\pi_{\boldsymbol{\theta}}(\mathbf{x}) = \mathcal{N}(\boldsymbol{\theta}, \mu^2\mathbf{I}_d)$, but we want to estimate its gradient using samples drawn from a different proposal distribution $p(\mathbf{x})$. The policy gradient using importance sampling becomes:

$$\nabla J(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x})}\left[\frac{\pi_{\boldsymbol{\theta}}(\mathbf{x})}{p(\mathbf{x})}\nabla \ln \pi_{\boldsymbol{\theta}}(\mathbf{x})\left(\mathbb{E}_\xi[f(\mathbf{x};\xi)] - b\right)\right] . \quad (10)$$

Similar to (9), by substituting $\nabla \ln \pi_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{\mathbf{x}-\boldsymbol{\theta}}{\mu^2}, b = \mathbb{E}_\xi[b(\xi)]$ and using Monte Carlo approximation with samples $\mathbf{x}_k \sim p(\mathbf{x})$, we get the stochastic gradient estimator:

$$\hat{\nabla}_{\text{IS}} J(\boldsymbol{\theta}) \triangleq \frac{1}{K}\sum_{k=1}^{K} \frac{\pi_{\boldsymbol{\theta}}(\mathbf{x}_k)}{p(\mathbf{x}_k)}\frac{\mathbf{x}_k - \boldsymbol{\theta}}{\mu^2}\left(f(\mathbf{x}_k;\xi) - b(\xi)\right) . \quad (11)$$

The connection between the ZOO gradient estimator under various sampling distributions from (2) and the IS-based REINFORCE gradient estimator (11) is formalized below (proof in Appx. C.3).

**Theorem 3.3** (Extended Gradient Estimator Equivalence)**.** *Let $\pi_{\boldsymbol{\theta}}(\mathbf{x}) = \mathcal{N}(\boldsymbol{\theta}, \mu^2\mathbf{I}_d)$, $p(\mathbf{x}) = p(\boldsymbol{\theta} + \mu\mathbf{u})$, and $b(\xi) = f(\boldsymbol{\theta};\xi)$ in (11). IS-based REINFORCE gradient estimator (11) is identical to a scaled ZOO gradient estimator (2) for the three different distributions of $\mathbf{u}_k$ in Sec. 2:*

$$\hat{\nabla}_{\text{IS}} J(\boldsymbol{\theta}) = \gamma \hat{\nabla} F(\boldsymbol{\theta}) .$$

*Particularly, let $\Gamma(\cdot)$ be the Gamma function,* (I) if $\mathbf{u}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, $\gamma = 1$; (II) if $\mathbf{u}_k \sim \text{Unif}(\mathbb{S}^{d-1})$, $\gamma = \frac{2^{1-d/2}\exp(-1/2)}{\mu\Gamma(d/2)}$; (III) if $\mathbf{u}_k \sim \text{Unif}(\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\})$, $\gamma = \frac{d\exp(-1/2)}{(2\pi\mu^2)^{d/2}}$.

**Remark.** Thm. 3.3 reveals that ZOO estimators employing non-Gaussian sampling distributions for $\mathbf{u}_k$ (e.g., uniform on sphere or coordinate-wise) can also be interpreted as REINFORCE gradient estimators through the lens of importance sampling. Specifically, the ZOO gradient $\hat{\nabla} F(\boldsymbol{\theta})$ (unbiased for its own smoothed objective $F_\mu(\boldsymbol{\theta})$ with the non-Gaussian $p(\mathbf{u})$) remains equivalent to an IS-based REINFORCE estimator for $J(\boldsymbol{\theta})$ with the Gaussian policy scaled by $\gamma$. This scaling factor $\gamma$ arises from the implicit importance weights between the Gaussian policy for $J(\boldsymbol{\theta})$ and the ZOO proposal distribution $p(\mathbf{u})$. This perspective unifies diverse ZOO sampling strategies under the REINFORCE lens, provides a principled reason for the learning rate adjustments in Cor. 3.4, and further solidifies the fundamental equivalence between the convergence of ZOO and single-step PO.

**Corollary 3.4** (Convergence Equivalence)**.** *Under the same condition in Thm. 3.3, let baseline $b(\xi)$ and update rule (e.g. gradient descent algorithm and Adam (Kingma & Ba, 2015) algorithm) be the same for ZOO and REINFORCE, they achieve identical convergence when*

$$\eta_R = \eta_Z/\gamma .$$

*Here, $\gamma$ is from Thm. 3.3, $\eta_Z$ and $\eta_R$ are the learning rates of ZOO and REINFORCE, respectively.*

## 4. Zeroth-Order Optimization with Averaged Baseline and Query Reuse

**Algorithm 1** ZOO with Averaged Baseline and Query Reuse

**Input:** objective function $f$, learning rate $\eta$, moment decay rates $\beta_1, \beta_2$, number of queries $K$ and histories $N$
**Initialize:** $\boldsymbol{\theta}_0, \boldsymbol{m}_0, \boldsymbol{v}_0, \mathcal{H}_0 = \varnothing$
**for** iteration $t \in [T]$ **do**
  Sample $\{\mathbf{u}_k\}_{k=1}^K$
  Query $\{y_k | y_k = f(\boldsymbol{\theta}_{t-1} + \mu\mathbf{u}_k; \xi)\}_{k=1}^K$
  $\mathcal{H}_t \leftarrow \mathcal{H}_{t-1} \setminus \mathcal{H}_{t-N} \cup \left\{ (\mathbf{u}_k, f(\boldsymbol{\theta}_{t-1} + \mu\mathbf{u}_k; \xi)) \right\}_{k=1}^K$
  $b_t \leftarrow \frac{1}{|\mathcal{H}_t|} \sum_{(\mathbf{u},y)\in\mathcal{H}_t} y$
  $\boldsymbol{g}_t \leftarrow \frac{1}{|\mathcal{H}_t|-1} \sum_{(\mathbf{u},y)\in\mathcal{H}_t} \frac{y - b_t}{\mu}\mathbf{u}$
  $\boldsymbol{m}_t \leftarrow \beta_1 \boldsymbol{m}_{t-1} + (1-\beta_1)\boldsymbol{g}_t$
  $\boldsymbol{v}_t \leftarrow \beta_2 \boldsymbol{v}_{t-1} + (1-\beta_2)\boldsymbol{m}_t^2$
  $\boldsymbol{\theta}_t \leftarrow \boldsymbol{\theta}_{t-1} - \eta \frac{\boldsymbol{m}_t}{\sqrt{\boldsymbol{v}_t + \zeta}}$
**end for**
**Output:** $\boldsymbol{\theta}_T$

Leveraging the Policy Optimization (PO) framework established in Sec. 3, this section introduces ZoAR (Algo. 1), an improved Zeroth-Order Optimization (ZOO) algorithm. We illustrate in Sec. 4.1 how ZoAR incorporates PO-inspired variance reduction techniques, including an averaged baseline and query reuse, for enhanced efficiency. While Algo. 1 demonstrates these techniques using the update rule from $\mathcal{R}$-AdaZO (Shu et al., 2025b), their core design is general and readily adaptable to other update rules like ZO-SGD (Ghadimi & Lan, 2013) and ZO-AdaMM (Chen et al., 2019). Furthermore, we provide theoretical analyses in ZOO theory to validate these PO-derived improvements in Appx. B.

### 4.1. Algorithm Design

We introduce the two key PO-inspired techniques in ZoAR (line 5 of Algo. 1), namely the averaged baseline and query reuse, below.

**Averaged Baseline.** As established in Thm. 3.2, the standard Gaussian-smoothed ZOO gradient estimator (2) implicitly uses $f(\boldsymbol{\theta}; \xi)$ as a baseline, corresponding to $b(\xi) = f(\boldsymbol{\theta}; \xi)$ in the REINFORCE framework (9). While this baseline helps reduce variance compared to no baseline, it may not be the most effective choice. In the single-step REINFORCE algorithm, the baseline that minimizes the variance of the gradient estimate $\nabla \ln \pi_{\boldsymbol{\theta}}(\mathbf{x})(R(\mathbf{x}) - b)$ is given by $b^* = \frac{\mathbb{E}_{\mathbf{x}\sim\pi_{\boldsymbol{\theta}}(\mathbf{x})}\left[(\nabla \ln \pi_{\boldsymbol{\theta}}(\mathbf{x}))^2 R(\mathbf{x})\right]}{\mathbb{E}_{\mathbf{x}\sim\pi_{\boldsymbol{\theta}}(\mathbf{x})}\left[(\nabla \ln \pi_{\boldsymbol{\theta}}(\mathbf{x}))^2\right]}$. A simpler and widely used near-optimal baseline is the expected reward itself, $b = \mathbb{E}_{\mathbf{x}\sim\pi_{\boldsymbol{\theta}}(\mathbf{x})}[R(\mathbf{x})]$. In our ZOO context, where $R(\mathbf{x}) = -F(\mathbf{x}) = -\mathbb{E}_\xi[f(\mathbf{x}; \xi)]$ and $\mathbf{x} = \boldsymbol{\theta} + \mu\mathbf{u}$, this corresponds to choosing the baseline as $b = \mathbb{E}_{\mathbf{x}\sim\pi_{\boldsymbol{\theta}}(\mathbf{x})}[F(\mathbf{x})] = F_\mu(\boldsymbol{\theta})$. The standard ZOO baseline $f(\boldsymbol{\theta}; \xi)$ can be seen as a single-sample, zero-order approximation of $F_\mu(\boldsymbol{\theta})$ evaluated at the center point. Algo. 1 proposes using a more robust estimate of this expected value. Specifically, it computes the baseline $b_t$ as the empirical average of function values obtained from recent queries stored in a history buffer $\mathcal{H}_t$:

$$b_t \triangleq \frac{1}{|\mathcal{H}_t|} \sum_{(\mathbf{u},y)\in\mathcal{H}_t} y\,, \tag{12}$$

where $y = f(\boldsymbol{\theta}_{t'} + \mu\mathbf{u}; \xi)$ for some past iteration $t' \le t-1$. This average in fact serves as a Monte Carlo estimate of the expected function value $F_\mu(\boldsymbol{\theta})$, potentially providing a lower-variance baseline compared to the single evaluation $f(\boldsymbol{\theta}; \xi)$ used implicitly in (2), which we will verify in Appx. B.

**Query Reuse.** To further enhance sample efficiency and reduce variance, Algo. 1 incorporates query reuse. This mirrors the concept of using off-policy data, common in algorithms like Proximal Policy Optimization (PPO) (Schulman et al., 2017), where experiences gathered under previous policies are reused to improve the current policy update, thereby increasing data efficiency. In our ZOO context, Algo. 1 maintains a history buffer $\mathcal{H}_t$ containing the $N \times K$ most recent query results (pairs of perturbation vectors $\mathbf{u}$ and corresponding function values $y$). At iteration $t$, $K$ new queries based on $\boldsymbol{\theta}_{t-1}$ are performed, added to the buffer, and the oldest $K$ queries are discarded. Crucially, the gradient estimate $\boldsymbol{g}_t = \hat{\nabla} F(\boldsymbol{\theta}_{t-1})$ is then computed using all samples currently in the history $\mathcal{H}_t$:

$$\hat{\nabla} F(\boldsymbol{\theta}_{t-1}) \triangleq \frac{1}{|\mathcal{H}_t| - 1} \sum_{(\mathbf{u},y)\in\mathcal{H}_t} \frac{y - b_t}{\mu}\mathbf{u}\,. \tag{13}$$

This approach uses all $|\mathcal{H}_t| = N \times K$ samples, effectively increasing the gradient estimation batch size without additional queries beyond the initial $K$. The resulting averaging over a larger set is expected to produce a gradient estimate with significantly lower variance (verified in Appx. B).

**Advantages.** The proposed ZoAR algorithm offers several compelling advantages. *(a)* It provides significant *variance reduction* by employing an averaged baseline $b_t$ and reusing historical queries from $\mathcal{H}_t$ (see Appx. B) compared to conventional ZOO with finite difference (Nesterov & Spokoiny, 2017). *(b)* Compared to (Cheng et al., 2021; Wang et al., 2024), the algorithm maintains compelling *computational and memory efficiency*, as the overhead for managing the history buffer (using only random seeds like (Malladi et al., 2023; Shu et al., 2025a)) and performing the averaging calculations is generally modest, which is scaling linearly with history size. *(c)* ZoAR benefits from *ease of Implementation*, representing a straightforward modification to standard ZOO procedures by incorporating a buffer and simple averaging steps. *(d)* It offers enhanced *sample efficiency and flexibility* by leveraging the accumulated information in $\mathcal{H}_t$: a meaningful gradient estimate $\boldsymbol{g}_t$ can be computed even if only a small number of new queries (potentially $K = 1$) are

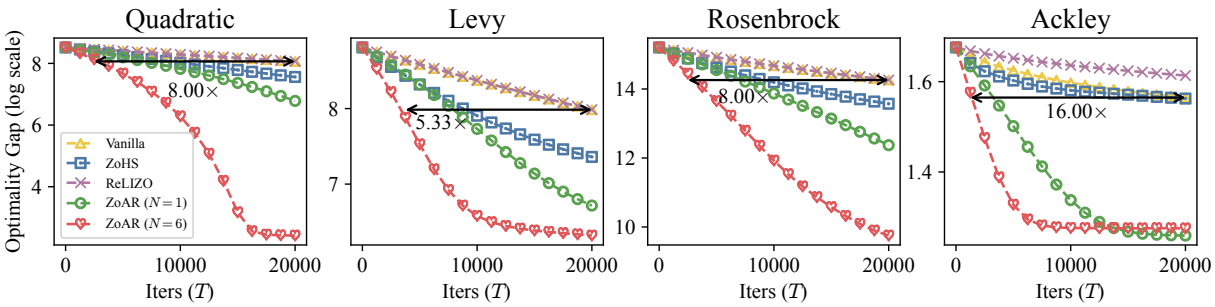


*Figure 1.* Comparison of convergence among different zeroth-order optimization algorithms on four synthetic functions. All curves are averaged over 5 independent runs.

*Table 1.* Comparison of the minimal number of iterations to achieve a successful attack for different ZOO methods. Results are averaged over 5 runs. The speedup is compared against the Vanilla ZOO.

| | Metric | Vanilla | ZoHS | ZoAR w/o history | ZoAR |
|---|---|---|---|---|---|
| $\mathcal{R}$-AdaZO | # Iters ($\times 10^2$) | 23.3±5.4 | 23.3±2.6 | 12.4±1.0 | **8.56**±2.2 |
| | Speedup | 1.0× | 1.0× | 1.87× | **2.72**× |
| ZO-AdaMM | # Iters ($\times 10^2$) | 65.3±12.7 | 36.9±5.8 | 32.6±8.0 | **11.0**±2.8 |
| | Speedup | 1.0× | 1.8× | 2.0× | **5.92**× |

performed at each iteration. These advantages make ZoAR a practical approach for improving ZOO performance, particularly in optimization settings where variance control and query efficiency is crucial.

## 5. Experiments

In this section, we conduct extensive experiments on synthetic functions (Sec. 5.1) and black-box adversarial attack (Sec. 5.2). More results, e.g., the equivalence between ZOO and REINFORCE, memory-efficient LLM fine-tuning, are in Appx. E.

### 5.1. Synthetic Functions

**The Superiority of ZoAR.** We subsequently evaluate the convergence rate and final performance of ZoAR against several baselines on four synthetic functions of dimensionality $d = 10^4$ (detailed in Appx. D.2). The compared methods include Vanilla ZOO (Nesterov & Spokoiny, 2017), ReLIZO (Wang et al., 2024), and ZoHS (details in Appx. D.1). Fig. 1 presents the results using the ZO-AdaMM (Chen et al., 2019) update rule, while corresponding results under the $\mathcal{R}$-AdaZO (Shu et al., 2025b) update rule are available in Appx. E.2. The results in Fig. 1 show that ZoAR consistently outperforms all baseline algorithms in both convergence speed and final optimization performance. Notably, ZoAR with $N = 6$ achieves an $8\times$ speedup over Vanilla ZOO on the Quadratic and Rosenbrock functions, and a $16\times$ speedup on the Ackley function. Moreover, comparing ZoAR with $N = 6$ (utilizing query reuse) against ZoAR with $N = 1$ (using only the averaged baseline) illustrates the significant additional benefit of historical information.

### 5.2. Black-box Adversarial Attack

We further evaluate the performance of ZoAR in the domain of black-box adversarial attacks, a prominent application of zeroth-order optimization (Cheng et al., 2021; Shu et al., 2023). In this scenario, the goal is to identify an optimal perturbation $\boldsymbol{\delta}$ for a given input image $x$ such that a target black-box model misclassifies $x+\boldsymbol{\delta}$. Our experimental setup follows that introduced by (Shu et al., 2025b), targeting a convolutional neural network (CNN) trained on the MNIST dataset (Lecun et al., 1998) (more details in Appx. D.3). We assess algorithm efficiency by the minimum number of iterations required to achieve a successful attack. The comparison includes Vanilla ZOO and ZoHS, with each evaluated under both the ZO-AdaMM (Chen et al., 2019) and $\mathcal{R}$-AdaZO (Shu et al., 2025b) update rules. ReLIZO is omitted from this comparison as it failed to achieve a successful attack within the maximum query budget. The results are summarized in Tab. 1, showing that ZoAR achieves the fastest attack success across both update rules. Specifically, under the ZO-AdaMM setting, ZoAR represents a $5.92\times$ speedup compared to Vanilla ZOO. The less pronounced speedup of ZoAR with $\mathcal{R}$-AdaZO (versus ZO-AdaMM) is likely due to the inherent gradient variance reduction of $\mathcal{R}$-AdaZO (Shu et al., 2025b), which may reduce the marginal impact of additional variance mitigation from ZoAR.

## 6. Conclusion

This paper established a novel and fundamental equivalence between zeroth-order optimization (ZOO) with finite differences and single-step policy optimization (PO). Leveraging this PO framework, we introduced ZoAR, an algorithm incorporating PO-inspired variance reduction techniques (an averaged baseline and query reuse) that demonstrably enhance performance. Our theoretical and empirical results highlight the benefits of this unified perspective, offering new insights into ZOO and providing a principled path for future algorithmic advancements.

## References


Chen, X., Liu, S., Xu, K., Li, X., Lin, X., Hong, M., and Cox, D. Zo-adamm: Zeroth-order adaptive momentum method for black-box optimization. In Proc. NeurIPS, 2019.

Cheng, S., Wu, G., and Zhu, J. On the convergence of prior-guided zeroth-order optimization algorithms. In Proc. NeurIPS, 2021.

Flaxman, A., Kalai, A. T., and McMahan, H. B. Online convex optimization in the bandit setting: Gradient descent without a gradient. In Proc. SODA, 2005a.

Flaxman, A. D., Kalai, A. T., and McMahan, H. B. Online convex optimization in the bandit setting: gradient descent without a gradient. In Proc. SODA, 2005b.

Ghadimi, S. and Lan, G. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. SIAM J. Optim., 23(4):2341–2368, 2013.

Ghadimi, S., Lan, G., and Zhang, H. Mini-batch stochastic approximation methods for nonconvex stochastic composite optimization. Math. Program., 155(1-2):267–305, 2016.

Gu, B., Liu, G., Zhang, Y., Geng, X., and Huang, H. Optimizing large-scale hyperparameters via automated learning algorithm. arXiv:2102.09026, 2021.

Hu, W., Shu, Y., Yu, Z., Wu, Z., Lin, X., Dai, Z., Ng, S.-K., and Low, B. K. H. Localized zeroth-order prompt optimization. In Proc. NeurIPS, 2024.

Jiang, S., Chen, Q., Pan, Y., Xiang, Y., Lin, Y., Wu, X., Liu, C., and Song, X. Zo-adamu optimizer: Adapting perturbation by the momentum and uncertainty in zeroth-order optimization. In Proc. AAAI, 2024.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In Proc. ICLR, 2015.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. Proceedings of the IEEE, pp. 2278–2324, 1998.

Lian, X., Zhang, H., Hsieh, C., Huang, Y., and Liu, J. A comprehensive linear speedup analysis for asynchronous stochastic parallel optimization from zeroth-order to first-order. In Proc. NIPS, 2016.

Liu, S., Kailkhura, B., Chen, P., Ting, P., Chang, S., and Amini, L. Zeroth-order stochastic variance reduction for nonconvex optimization. In Proc. NeurIPS, 2018a.

Liu, S., Li, X., Chen, P., Haupt, J. D., and Amini, L. Zeroth-order stochastic projected gradient descent for nonconvex optimization. In Proc. GlobalSIP, 2018b.

Malladi, S., Gao, T., Nichani, E., Damian, A., Lee, J. D., Chen, D., and Arora, S. Fine-tuning language models with just forward passes. In Proc. NeurIPS, 2023.

Nazari, P., Tarzanagh, D. A., and Michailidis, G. Adaptive first-and zeroth-order methods for weakly convex stochastic optimization problems. arXiv:2005.09261, 2020.

Nesterov, Y. E. and Spokoiny, V. G. Random gradient-free minimization of convex functions. Found. Comput. Math., 17(2):527–566, 2017.

Salimans, T., Ho, J., Chen, X., and Sutskever, I. Evolution strategies as a scalable alternative to reinforcement learning. arXiv:1703.03864, 2017.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. Proximal policy optimization algorithms. arXiv:1707.06347, 2017.

Shu, Y., Dai, Z., Sng, W., Verma, A., Jaillet, P., and Low, B. K. H. Zeroth-order optimization with trajectory-informed derivative estimation. In Proc. ICLR, 2023.

Shu, Y., Lin, X., Dai, Z., and Low, B. K. H. Federated zeroth-order optimization using trajectory-informed surrogate gradients. In Workshop on Differentiable Almost Everything (ICML), 2024.

Shu, Y., Hu, W., Ng, S.-K., Low, B. K. H., and Yu, F. R. Ferret: Federated full-parameter tuning at scale for large language models. In Proc. ICML, 2025a.

Shu, Y., Zhang, Q., He, K., and Dai, Z. Refining adaptive zeroth-order optimization at ease. In Proc. ICML, 2025b.

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

Sutton, R. S. and Barto, A. G. Reinforcement learning - an introduction, 2nd Edition. MIT Press, 2018.

Sutton, R. S., McAllester, D. A., Singh, S., and Mansour, Y. Policy gradient methods for reinforcement learning with function approximation. In Proc. NIPS, 1999.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., and Bowman, S. R. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In Proc. ICLR, 2019. In the Proceedings of ICLR.

Wang, X., Qin, X., Yang, X., and Yan, J. Relizo: Sample reusable linear interpolation-based zeroth-order optimization. In Proc. NeurIPS, 2024.

Williams, R. J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Mach. Learn., 8:229–256, 1992.

Zhan, H., Chen, C., Ding, T., Li, Z., and Sun, R. Unlocking black-box prompt tuning efficiency via zeroth-order optimization. In Proc. EMNLP (Findings), 2024.

Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., Dewan, C., Diab, M., Li, X., Lin, X. V., Mihaylov, T., Ott, M., Shleifer, S., Shuster, K., Simig, D., Koura, P. S., Sridhar, A., Wang, T., and Zettlemoyer, L. Opt: Open pre-trained transformer language models, 2022.

Zhang, Y., Li, P., Hong, J., Li, J., Zhang, Y., Zheng, W., Chen, P., Lee, J. D., Yin, W., Hong, M., Wang, Z., Liu, S., and Chen, T. Revisiting zeroth-order optimization for memory-efficient LLM fine-tuning: A benchmark. In Proc. ICML, 2024.

Zhou, J., Yang, Y., Zhen, K., Liu, Z., Zhao, Y., Banijamali, E., Mouchtaris, A., Wong, N., and Zhang, Z. Quzo: Quantized zeroth-order fine-tuning for large language models, 2025.

## A. Related Work

Zeroth-Order (ZO) optimization research has primarily advanced along two interconnected fronts: the design of gradient estimators and the development of update rules or full algorithms.

**ZO Gradient Estimation.** A cornerstone of ZOO is the estimation of gradients using only function evaluations, typically through finite difference approximations. Seminal work introduced Gaussian random perturbations for smooth objectives, establishing theoretical convergence (Nesterov & Spokoiny, 2017). Other perturbation strategies include uniform sampling from the unit sphere (Flaxman et al., 2005a) or coordinate-wise perturbations (Lian et al., 2016). A primary challenge with these methods is the high variance in their gradient estimates. To address this, several approaches have been developed. E.g., prior-guided gradient estimation leverages historical estimates to denoise current ones (Cheng et al., 2021). Recently, methods have explored learning surrogate models of the objective function using past queries to derive more stable gradient estimates (Shu et al., 2023; 2024). Another line of work has focused on linear interpolation strategies for more accurate estimates by reusing queries from prior iterations to reduce complexity while maintaining sample quality (Wang et al., 2024). While these methods offer valuable improvements, the underlying connection between the widely-used finite difference ZOO gradient estimators and principles from Reinforcement Learning (RL), particularly Policy Optimization (PO), has remained largely unelucidated. Our work bridges this gap by reinterpreting these estimators through a PO lens, which not only reveals inherent variance reduction mechanisms but also inspires new ones. Leveraging this novel PO framework, this paper introduces new PO-inspired variance reduction techniques, specifically an averaged baseline and query reuse, which are central to our proposed ZoAR algorithm and aim to significantly improve the stability and efficiency of ZO gradient estimation.

**ZO Update Rules and Algorithms.** Given a ZO gradient estimate, many ZOO algorithms directly adopt update rules from first-order (FO) optimization. A significant body of work employs Stochastic Gradient Descent (SGD) or its variants (Ghadimi & Lan, 2013; Ghadimi et al., 2016; Nesterov & Spokoiny, 2017; Liu et al., 2018b;a; Cheng et al., 2021; Shu et al., 2023). Recognizing the potential benefits of adaptive step sizes, some research has integrated adaptive methods like Adam (Kingma & Ba, 2015) into the ZOO setting (Chen et al., 2019; Nazari et al., 2020; Jiang et al., 2024). Further advancing these adaptive methods, recent work such as $\mathcal{R}$-AdaZO (Shu et al., 2025b) has focused on refining the utilization of moment information, demonstrating how careful handling of first and second moment estimates can lead to significant variance reduction in the gradient estimates and a more accurate capture of the optimization landscape, thereby improving convergence. Notably, this paper does not aim to introduce a new update rule, but focus on unveiling the fundamental connection between ZOO and PO, and developing advanced gradient estimation method that is applicable to all these existing update rules and algorithms.

## B. Theoretical Analysis

This section provides a theoretical underpinning for our ZoAR (Algo. 1). We analyze the bias of its gradient estimator, the optimality of its baseline, the bias-variance trade-off, and its convergence. To ease our proof, we follow the common practice in (Shu et al., 2025b) to prove under $\mathbf{u} \sim \mathrm{Unif}(\mathbb{S}^{d-1})$ and the following commonly used assumptions.

**Assumption B.1** (Bounded Continuity and Smoothness)**.** *$\forall \boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathbb{R}^d$ and $i \in [d]$,*

$$\begin{aligned} |f(\boldsymbol{\theta}, \xi)| &\leq C \,, \\ |F(\boldsymbol{\theta}) - F(\boldsymbol{\theta}')| &\leq L_0 \|\boldsymbol{\theta} - \boldsymbol{\theta}'\| \,, \\ |\nabla_i F(\boldsymbol{\theta}) - \nabla_i F(\boldsymbol{\theta}')| &\leq L_1 \|\boldsymbol{\theta} - \boldsymbol{\theta}'\| \,. \end{aligned} \tag{14}$$

**Assumption B.2** (Bounded Variance)**.** *$\forall \boldsymbol{\theta} \in \mathbb{R}^d$,*

$$\begin{aligned} \mathbb{E}_\xi[|f(\boldsymbol{\theta}, \xi) - F(\boldsymbol{\theta})|^2] &\leq \sigma_\xi^2 \,, \\ \mathbb{E}_{\mathbf{u}}[|F(\boldsymbol{\theta} + \mu \mathbf{u}) - F_\mu(\boldsymbol{\theta})|^2] &\leq \sigma_\mu^2 \,. \end{aligned} \tag{15}$$

**Theorem B.3** (Bias Analysis)**.** *For every iteration $t$ of ZoAR (Algo. 1) with history depth $N \geq 1$ and $K$ queries per step, the expected value of the gradient estimator $\hat{\nabla} F(\boldsymbol{\theta}_{t-1})$ is:*

$$\mathbb{E}\left[\hat{\nabla} F(\boldsymbol{\theta}_{t-1})\right] = \frac{1}{N} \sum_{n=1}^{N} \nabla F_\mu(\boldsymbol{\theta}_{t-n}) \,.$$

**Remark.** Its proof is in Appx. C.4. Thm. B.3 reveals that $\hat{\nabla} F(\boldsymbol{\theta}_{t-1})$ in (13) of ZoAR is secretly an unbiased estimator for the average of smoothed gradients from the current and $N-1$ previous parameters. This implies that (13) implicitly targets this historically averaged groundtruth, a mechanism that shall potentially reduce the gradient estimation variance at $\boldsymbol{\theta}_{t-1}$ by effectively increasing the number of queries contributing to the estimate (see Thm. B.5). However, (13) is biased with respect to the current smoothed gradient $\nabla F_\mu(\boldsymbol{\theta}_{t-1})$ when $N \geq 2$, which emerges because these historical parameters $\boldsymbol{\theta}_{t-n}$ have diverged from $\boldsymbol{\theta}_{t-1}$. This is an inherent consequence of leveraging historical queries for variance reduction, creating a bias-variance trade-off detailed in Thm. B.5. Notably, if $N=1$ (no query reuse beyond the current batch), the estimator becomes unbiased for $\nabla F_\mu(\boldsymbol{\theta}_{t-1})$.

**Theorem B.4** (Optimal Baseline)**.** *Let* $\mathbf{u} \sim \mathrm{Unif}(\mathbb{S}^{d-1})$*, for every* $t$ *of ZoAR (Algo. 1) with* $N \geq 1$*, the optimal* $b_t$ *to minimize* $\mathrm{Var}\left(\hat{\nabla} F(\boldsymbol{\theta}_{t-1})\right) = \mathbb{E}\left[\left\|\hat{\nabla} F(\boldsymbol{\theta}_{t-1}) - \frac{1}{N}\sum_{n=1}^{N} \nabla F_\mu(\boldsymbol{\theta}_{t-n})\right\|^2\right]$ *is*

$$b_t^* = \frac{1}{N}\sum_{n=1}^{N} F_\mu(\boldsymbol{\theta}_{t-n})\,.$$

**Remark.** Its proof is in Appx. C.5. Thm. B.4 provides strong theoretical support for the averaged baseline in ZoAR. It demonstrates that for gradient estimator (13) of ZoAR under $\mathbf{u} \sim \mathrm{Unif}(\mathbb{S}^{d-1})$, the baseline $b_t$ defined in (12) is in fact a practical Monte Carlo approximation of the variance-minimizing $b_t^*$. This result formalizes the intuition that averaging recent function evaluations provides a more effective baseline than a single point estimate (like $f(\boldsymbol{\theta};\xi)$ implicitly used in vanilla ZOO, or no baseline at all), thereby contributing to the overall variance reduction of the gradient estimate from a pure perspective of ZOO theory. Crucially, the structural similarity between the optimal $b_t^*$ above and the variance-minimizing baseline $b = \mathbb{E}_{\mathbf{x}\sim\pi_{\boldsymbol{\theta}}(\mathbf{x})}[R(\mathbf{x})]$ in the REINFORCE algorithm further underscores the principled PO foundation and validates the practical efficacy of our $b_t$ approximation.

**Theorem B.5** (Bias-Variance Decomposition)**.** *Let* $\mathbf{u} \sim \mathrm{Unif}(\mathbb{S}^{d-1})$ *and* $b_t$ *in* (13) *be the optimal* $b_t^*$ *in Thm. B.4, under Assump. B.1 and B.2, for every* $t$ *of ZoAR (Algo. 1) with* $N \geq 1$*,*

$$\mathbb{E}\left[\left\|\hat{\nabla} F(\boldsymbol{\theta}_{t-1}) - \nabla F_\mu(\boldsymbol{\theta}_{t-1})\right\|^2\right] \leq \underbrace{\frac{\sigma_\xi^2 + \sigma_\mu^2}{NK\mu^2}}_{\text{Variance} \triangleq V} + \underbrace{\frac{\eta^2 L_0^2 d\left(N^2-1\right)}{3\left(1-\beta_2\right)N^2 K\mu^2} + \frac{\eta^2 L_1^2 d^2\left(N-1\right)}{2\left(1-\beta_2\right)}}_{\text{Squared Bias}}\,.$$

**Remark.** Its proof is in Appx. C.6. Thm. B.5 explicitly quantifies the fundamental trade-off inherent in the query reuse mechanism of ZoAR. The first component *variance* $V$ is what ZoAR primarily targets for reduction through its PO-inspired techniques: the averaged baseline (justified by Thm. B.4) and the reuse of historical samples. The second component *squared bias* arises because the gradient estimator, as shown in Thm. B.3, is an average of historical smoothed gradients, which may differ from the current target gradient $\nabla F_\mu(\boldsymbol{\theta}_{t-1})$. Consequently, while increasing the history depth $N$ can substantially decrease variance, it may simultaneously inflate the bias. Fortunately, this bias can be small with a small learning rate $\eta$, and is completely avoided when $N=1$.

**Theorem B.6** (Variance-Aware Convergence, Informal)**.** *Let* $\mathbf{u} \sim \mathrm{Unif}(\mathbb{S}^{d-1})$ *and* $b_t$ *in* (13) *be the optimal* $b_t^*$ *in Thm. B.4, under Assump. B.1 and B.2, when* $1-\beta_2 \sim \mathcal{O}\left(\epsilon^2\right)$*,* $\eta \sim \mathcal{O}\left(\epsilon^2\right)$*,* $T \sim \mathcal{O}\left(\epsilon^{-4}\right)$*,* $\beta_1 \leq \sqrt{\beta_2}, \beta_2 > 1/2, \boldsymbol{m}_{0,i} = 0, \boldsymbol{v}_{0,i} > 0\ (\forall i \in [d])$*, the following holds for ZoAR (Algo. 1) under certain constants* $B_1$ *and* $B_2$ *that are independent of* $\epsilon$*,*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\nabla F(\boldsymbol{\theta}_t)\|] \leq \sqrt{\frac{2}{\beta_1\left(1-\beta_2\right)}\left(1+\beta_1\right)\epsilon^2} + \left(\sqrt[4]{\zeta} + \sqrt{\Xi}\right)\epsilon + \mu L_1\sqrt{d} + B_2\,,$$

*where* $\Xi \triangleq B_1 + \sqrt{\frac{2(1+\beta_1)(1-\beta_1)^2\beta_2}{\left(\beta_2-\beta_1^2\right)(1-\beta_2)}V}$ *and* $V = \frac{\sigma_\xi^2+\sigma_\mu^2}{NK\mu^2}$*.*

**Remark.** Its proof is in Appx. C.7. Thm. B.6 presents the convergence of ZoAR (Algo.1). Notably, it highlights that the convergence rate is directly influenced by the variance $V$ (occurred in $\Xi$) of our gradient estimator (13), which corresponds to the variance term in Thm. B.5. When $V$ dominates the convergence, minimizing $V$ is hence the key to achieving better optimization performance, which provides a strong theoretical support for the variance reduction techniques used in ZoAR (averaged baseline and query reuse). The $\mathcal{O}(\mu)$ term is standard in conventional ZOO, reflecting the inherent discrepancy from optimizing the smoothed objective $F_\mu$ instead of $F$. While the additional $B_2$ term results from the bias introduced by our (13) as revealed in Thm. B.3, it can be small with a small $\eta$.

# C. Proofs

## C.1. Proof of Thm. 3.1

*Proof.* By definition, $J(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}\sim\pi_{\boldsymbol{\theta}}(\mathbf{x})}[F(\mathbf{x})]$. Substituting the reparameterization $\mathbf{x} = \boldsymbol{\theta} + \mu\mathbf{u}$ where $\mathbf{u} \sim p(\mathbf{u})$, the expectation over $\mathbf{x} \sim \pi_{\boldsymbol{\theta}}(\mathbf{x})$ becomes an expectation over $\mathbf{u} \sim p(\mathbf{u})$:

$$J(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{u}\sim p(\mathbf{u})}\left[F(\boldsymbol{\theta} + \mu\mathbf{u})\right] , \tag{16}$$

which is precisely the definition of the ZOO smoothed objective $F_\mu(\boldsymbol{\theta})$ in (3). □

## C.2. Proof of Thm. 3.2

*Proof.* Substituting $\mathbf{x}_k = \boldsymbol{\theta} + \mu\mathbf{u}_k$ and $b(\xi) = f(\boldsymbol{\theta};\xi)$ into the REINFORCE estimator (9):

$$\hat{\nabla}_{\text{GS}} J(\boldsymbol{\theta}) = \frac{1}{K}\sum_{k=1}^{K} \frac{(\boldsymbol{\theta} + \mu\mathbf{u}_k) - \boldsymbol{\theta}}{\mu^2}\left(f(\boldsymbol{\theta} + \mu\mathbf{u}_k;\xi) - f(\boldsymbol{\theta};\xi)\right) = \frac{1}{K}\sum_{k=1}^{K} \frac{f(\boldsymbol{\theta} + \mu\mathbf{u}_k;\xi) - f(\boldsymbol{\theta};\xi)}{\mu}\mathbf{u}_k \, . \tag{17}$$

This is exactly the Gaussian-smoothed ZOO gradient estimator $\hat{\nabla}F(\boldsymbol{\theta})$ in (2). □

## C.3. Proof of Thm. 3.3

*Proof.* Initially, inserting $\mathbf{x}_k = \boldsymbol{\theta} + \mu\mathbf{u}_k$ into $\pi_{\boldsymbol{\theta}}(\mathbf{x}_k) = \mathcal{N}(\boldsymbol{\theta}, \mu^2\mathbf{I}_d)$:

$$\pi_{\boldsymbol{\theta}}(\mathbf{x}_k) = \frac{e^{-\frac{\|\mathbf{x}_k - \boldsymbol{\theta}\|}{2\mu^2}}}{(2\pi\mu^2)^{\frac{d}{2}}} = \frac{1}{\mu^d}\frac{e^{-\frac{\|\mathbf{u}_k\|}{2}}}{(2\pi)^{\frac{d}{2}}} \, . \tag{18}$$

We next consider the transformation of $p(\mathbf{x}_k)$ under three different distributions separately:

(I) If $\boldsymbol{u}_k$ follows the standard Gaussian distribution, $\boldsymbol{x}_k$ follows the Gaussian distribution $\mathcal{N}(\boldsymbol{\theta}, \mu^2\mathbf{I}_d)$, then consequently:

$$p(\mathbf{x}_k) = \frac{e^{-\frac{\|\mathbf{x}_k - \boldsymbol{\theta}\|}{2\mu^2}}}{(2\pi\mu^2)^{\frac{d}{2}}} = \frac{1}{\mu^d}\frac{e^{-\frac{\|\mathbf{u}_k\|}{2}}}{(2\pi)^{\frac{d}{2}}} \, , \tag{19}$$

which is the same as $\pi_{\boldsymbol{\theta}}(\mathbf{x}_k)$.

(II) If $\boldsymbol{u}_k$ follows the uniform distribution over the unit sphere $\text{Unif}(\mathbb{S}^{d-1})$, $\mathbf{x}_k$ follows the uniform distribution over the sphere with radius $\mu$ $\text{Unif}(\mathbb{S}^{d-1}(\mu))$, then consequently:

$$p(\mathbf{x}_k) = \frac{1}{\text{Area}\left(\mathbb{S}^{d-1}(\mu)\right)} = \frac{\Gamma\left(\frac{d}{2}\right)}{2\pi^{\frac{d}{2}}\mu^{d-1}} \, , \tag{20}$$

with constraint $\|\mathbf{u}\| = 1$, where $\Gamma(\cdot)$ denotes Gamma function.

(III) If $\boldsymbol{u}_k$ follows the uniform distribution over standard basis vectors $\text{Unif}(\{\boldsymbol{e}_i\}_{i=1}^d)$, $\mathbf{x}_k$ follows the uniform distribution over the orthonormal basis vectors $\text{Unif}(\{\boldsymbol{\theta} + \mu\boldsymbol{e}_i\}_{i=1}^d)$, then consequently:

$$p(\mathbf{x}_k) = \frac{1}{d} \, , \tag{21}$$

with constraint $\mathbf{u} \in \{\boldsymbol{e}_i\}_{i=1}^d$.

Afterthat, let $\gamma \triangleq \frac{\pi_{\boldsymbol{\theta}}(\mathbf{x}_k)}{p(\mathbf{x}_k)}$, and substitute $\mathbf{x}_k = \boldsymbol{\theta} + \mu\mathbf{u}_k$, $b(\xi) = f(\boldsymbol{\theta};\xi)$ into IS-based REINFORCE gradient estimator (11):

$$\hat{\nabla}_{\text{IS}} J(\boldsymbol{\theta}) = \gamma\frac{1}{K}\sum_{k=1}^{K} \frac{f(\boldsymbol{\theta} + \mu\mathbf{u}_k;\xi) - f(\boldsymbol{\theta};\xi)}{\mu}\mathbf{u}_k = \gamma\hat{\nabla}F(\boldsymbol{\theta}) \, ; \tag{22}$$

where (I) for $\mathbf{u}_k \sim \mathcal{N}(\mathbf{0}, \mu^2\mathbf{I}_d)$, $\gamma = 1$; (II) for $\mathbf{u}_k \sim \text{Unif}(\mathbb{S}^{d-1})$, $\gamma = \frac{2^{1-\frac{d}{2}}e^{-\frac{1}{2}}}{\mu\Gamma(\frac{d}{2})}$; (III) for $\mathbf{u}_k \sim \text{Unif}(\{\boldsymbol{e}_i\}_{i=1}^d)$, $\gamma = \frac{de^{-\frac{1}{2}}}{(2\pi\mu^2)^{\frac{d}{2}}}$.

□

### C.4. Proof of Thm. B.3

*Proof.* By inserting the average baseline (12), the expectation of the gradient estimator (13) can be expressed as follows:

$$
\begin{aligned}
\mathbb{E}\left[\hat{\nabla} F(\boldsymbol{\theta}_{t-1})\right] =& \frac{1}{NK-1}\mathbb{E}_{\mathbf{u}}\mathbb{E}_{\xi}\left[\sum_{n,k=1}^{N,K}\frac{f\left(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k};\xi\right)-\frac{1}{NK}\sum_{n',k'=1}^{N,K} f\left(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n',k'};\xi\right)}{\mu}\mathbf{u}_{t-n,k}\right] \\
=& \frac{1}{NK-1}\mathbb{E}_{\mathbf{u}}\left[\sum_{n,k=1}^{N,K}\frac{F\left(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k}\right)-\frac{1}{NK}\sum_{n',k'=1}^{N,K} F\left(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n',k'}\right)}{\mu}\mathbf{u}_{t-n,k}\right] \\
=& \frac{1}{NK-1}\mathbb{E}_{\mathbf{u}}\left[\sum_{n,k=1}^{N,K}\frac{\frac{NK-1}{NK}F\left(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k}\right)-\frac{1}{NK}\sum_{\substack{n',k'=1\\ n'\neq n,k'\neq k}}^{N,K} F\left(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n',k'}\right)}{\mu}\mathbf{u}_{t-n,k}\right] \\
\overset{(a)}{=}& \frac{1}{NK}\sum_{n,k=1}^{N,K}\mathbb{E}_{\mathbf{u}_{t-n,k}}\left[\frac{F\left(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k}\right)}{\mu}\mathbf{u}_{t-n,k}\right] \\
\overset{(b)}{=}& \frac{1}{NK}\sum_{n,k=1}^{N,K}\mathbb{E}_{\mathbf{u}_{t-n,k}}\left[\nabla F\left(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k}\right)\right] \\
=& \frac{1}{N}\sum_{n=1}^{N}\nabla F_{\mu}\left(\boldsymbol{\theta}_{t-n}\right) ,
\end{aligned}
\tag{23}
$$

where $\mathbb{E}_{\mathbf{u}}$ denote the expectation over $\mathbf{u}_{t-1,1},\cdots,\mathbf{u}_{t-n,K}$ for simplicity. Besides, $(a)$ follows from the fact that $\mathbf{u}_{t-n',k'}$ within the summation over $n',k'$ is uncorrelated with $\mathbf{u}_{t-n,k}$. When $\mathbf{u}\sim\mathcal{N}(\mathbf{0},\mathbf{I}^d)$, $(b)$ is a direct consequence of Stein's Lemma in (Stein, 1981). Alternatively, when $\mathbf{u}\sim\mathrm{Unif}(\mathbb{S}^{d-1})$, $(b)$ follows from Lemma 2.1 in (Flaxman et al., 2005b), utilizing the definition $F_{\mu}(\boldsymbol{\theta})\triangleq\mathbb{E}_{\mathbf{u}\sim\mathrm{Unif}(\mathbb{B}^d)}\left[F(\boldsymbol{\theta}+\mu\mathbf{u})\right]$. □

**Remark.** Note that in the step $(a)$, the baseline term $\frac{1}{NK}\sum_{\substack{n',k'=1\\ n'\neq n,k'\neq k}}^{N,K} F\left(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n',k'}\right)$ vinishes because it is independent of the random variable $\mathbf{u}_{t-n,k}$. Similar to the REINFORCE estimator in (6), incorporating the baseline $b(\xi)$ does not alter the expected value of the gradient estimator, which further supports the connection between the REINFORCE and ZOO gradient estimators.

### C.5. Proof of Thm. B.4

*Proof.* Beginning with the definition of $\mathrm{Var}\left(\hat{\nabla} F(\boldsymbol{\theta}_{t-1})\right)$:

$$
\begin{aligned}
\mathrm{Var}\left(\hat{\nabla} F(\boldsymbol{\theta}_{t-1})\right) =& \mathbb{E}\left[\left\|\hat{\nabla} F(\boldsymbol{\theta}_{t-1})-\frac{1}{N}\sum_{n=1}^{N}\nabla F_{\mu}\left(\boldsymbol{\theta}_{t-n}\right)\right\|^2\right] \\
=& \mathbb{E}\left[\left\|\hat{\nabla} F(\boldsymbol{\theta}_{t-1})\right\|^2\right]-2\left\langle\mathbb{E}\left[\hat{\nabla} F(\boldsymbol{\theta}_{t-1})\right],\frac{1}{N}\sum_{n=1}^{N}\nabla F_{\mu}\left(\boldsymbol{\theta}_{t-n}\right)\right\rangle+\left\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_{\mu}\left(\boldsymbol{\theta}_{t-n}\right)\right\|^2 \\
\overset{(a)}{=}& \mathbb{E}\left[\left\|\frac{1}{NK}\sum_{n,k=1}^{N,K}\frac{f(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k};\xi)-b_t}{\mu}\mathbf{u}_{t-n,k}\right\|^2\right]-\left\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_{\mu}\left(\boldsymbol{\theta}_{t-n}\right)\right\|^2 ,
\end{aligned}
\tag{24}
$$

where $(a)$ comes from Thm. B.3.

It is obvious that (24) is actually a quadratic function w.r.t $b_t$, and hence the optimal $b_t^*$ is derived by setting its derivative to

zero $\frac{\partial}{\partial b_t}\operatorname{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{t-1})\right)=0$, i.e.

$$\begin{aligned}\frac{\partial}{\partial b_t}\operatorname{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{t-1})\right)=&2\mathbb{E}\left[\left\langle\frac{1}{NK}\sum_{n,k=1}^{N,K}\frac{f(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k};\xi)-b_t}{\mu}\mathbf{u}_{t-n,k},-\frac{1}{NK\mu}\sum_{n,k=1}^{N,K}\mathbf{u}_{t-n,k}\right\rangle\right]\\ \overset{(a)}{=}&-\frac{2}{N^2K^2\mu^2}\sum_{n,k=1}^{N,K}\mathbb{E}\left[\left\langle\left(f(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k};\xi)-b_t\right)\mathbf{u}_{t-n,k},\mathbf{u}_{t-n,k}\right\rangle\right]\\ =&-\frac{2}{N^2K^2\mu^2}\sum_{n,k=1}^{N,K}\mathbb{E}_{\mathbf{u}_{t-n,k}}\left[\left\langle\left(F(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k})-b_t\right)\mathbf{u}_{t-n,k},\mathbf{u}_{t-n,k}\right\rangle\right]\\ =&-\frac{2}{N^2K^2\mu^2}\sum_{n,k=1}^{N,K}\left(\mathbb{E}_{\mathbf{u}_{t-n,k}}\left[F(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k})\left\|\mathbf{u}_{t-n,k}\right\|^2\right]-b_t\mathbb{E}_{\mathbf{u}_{t-n,k}}\left[\left\|\mathbf{u}_{t-n,k}\right\|^2\right]\right),\end{aligned}\tag{25}$$

where $(a)$ is due to the independence of $\mathbf{u}_{t-n,k}$ across different iterations $t$ and queries $k$.

Setting $\frac{\partial}{\partial b_t}\operatorname{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{t-1})\right)=0$, we have:

$$b_t^*=\frac{\frac{1}{NK}\sum_{n,k=1}^{N,K}\mathbb{E}_{\mathbf{u}_{t-n,k}}\left[F(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k})\left\|\mathbf{u}_{t-n,k}\right\|^2\right]}{\frac{1}{NK}\sum_{n,k=1}^{N,K}\mathbb{E}_{\mathbf{u}_{t-n,k}}\left[\left\|\mathbf{u}_{t-n,k}\right\|^2\right]}=\frac{\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}_{\mathbf{u}}\left[F(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u})\left\|\mathbf{u}\right\|^2\right]}{\mathbb{E}_{\mathbf{u}}\left[\left\|\mathbf{u}\right\|^2\right]}.\tag{26}$$

If $\mathbf{u}_k\sim\operatorname{Unif}(\mathbb{S}^{d-1})$ or $\mathbf{u}_k\sim\operatorname{Unif}(\{\boldsymbol{e}_i\}_{i=1}^d)$, it follows that $\|\mathbf{u}\|^2=1$. Consequently:

$$b_t^*=\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}_{\mathbf{u}}\left[F(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u})\right]=\frac{1}{N}\sum_{n=1}^{N}F_\mu\left(\boldsymbol{\theta}_{t-n}\right),\tag{27}$$

which completes the proof. □

### C.6. Proof of Thm. B.5

*Proof.* To begin with, we let $b_t$ be the optimal value obtained from Thm. B.4, and proceed with the calculation of $\operatorname{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{t-1})\right)$ from (24):

$$\operatorname{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{t-1})\right)\le\mathbb{E}\left[\left\|\frac{1}{NK}\sum_{n,k=1}^{N,K}\frac{f(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k};\xi)-b_t}{\mu}\mathbf{u}_{t-n,k}\right\|^2\right]-\left\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_\mu\left(\boldsymbol{\theta}_{t-n}\right)\right\|^2.\tag{28}$$

For the first term of (28):

$$\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{NK}\sum_{n,k=1}^{N,K}\frac{f(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k};\xi)-b_t}{\mu}\mathbf{u}_{t-n,k}\right\|^2\right]\\
\overset{(a)}{=}&\frac{1}{N^2K^2\mu^2}\sum_{n,k=1}^{N,K}\mathbb{E}\left[\left\|\left(f(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k};\xi)-b_t\right)\mathbf{u}_{t-n,k}\right\|^2\right]\\
\overset{(b)}{=}&\frac{1}{N^2K^2\mu^2}\sum_{n,k=1}^{N,K}\mathbb{E}\left[\left|f(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k};\xi)-b_t\right|^2\right]\\
\overset{(c)}{\leq}&\frac{1}{N^2K^2\mu^2}\sum_{n,k=1}^{N,K}\mathbb{E}_{\mathbf{u}}\left[\sigma_\xi^2+\left|F(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k})-b_t\right|^2\right]\\
\overset{(d)}{\leq}&\frac{1}{N^2K^2\mu^2}\left(NK\sigma_\xi^2+\sum_{n,k=1}^{N,K}\mathbb{E}_{\mathbf{u}}\left[\left|F(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k})-F_\mu(\boldsymbol{\theta}_{t-n})\right|^2\right]+K\sum_{n=1}^{N}\left|F_\mu(\boldsymbol{\theta}_{t-n})-b_t\right|^2\right)\\
\overset{(e)}{\leq}&\frac{1}{NK\mu^2}\left(\sigma_\xi^2+\sigma_\mu^2+\frac{L_0^2}{N^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\left\|\boldsymbol{\theta}_{t-n}-\boldsymbol{\theta}_{t-n'}\right\|^2\right),
\end{aligned}\tag{29}$$

where $(a)$ is derived from the independence of $\mathbf{u}_{t-n,k}$ across different iterations $t$ and queries $k$, $(b)$ comes from the fact that $\mathbf{u}\sim\mathrm{Unif}(\mathbb{S}^{d-1})$ and hence $\|\mathbf{u}\|^2=1$, $(c)$ is obtained from Assump. B.2, and $(d)$ results from $\mathbb{E}_{\mathbf{u}}\left[F(\boldsymbol{\theta}_{t-n}+\mu\mathbf{u}_{t-n,k})-F_\mu(\boldsymbol{\theta}_{t-n})\right]=0$. In step $(e)$, we apply Assump. B.2 and the following inequality:

$$\begin{aligned}
\left|F_\mu(\boldsymbol{\theta}_{t-n})-b_t\right|^2=&\left|\frac{1}{N}\sum_{n'=1}^{N}\left(F_\mu(\boldsymbol{\theta}_{t-n})-F_\mu(\boldsymbol{\theta}_{t-n'})\right)\right|^2\overset{(a)}{\leq}\frac{1}{N}\sum_{n'=1}^{N}\left|F_\mu(\boldsymbol{\theta}_{t-n})-F_\mu(\boldsymbol{\theta}_{t-n'})\right|^2\\
\overset{(b)}{\leq}&\frac{L_0^2}{N}\sum_{n'=1}^{N}\left\|\boldsymbol{\theta}_{t-n}-\boldsymbol{\theta}_{t-n'}\right\|^2,
\end{aligned}\tag{30}$$

where $(a)$ follows from the Jensen's inequality and $(b)$ is derived from Assump. B.1.

Inserting the results of 29 into 28, we have

$$\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{t-1})\right)\leq\frac{1}{NK\mu^2}\left(\sigma_\xi^2+\sigma_\mu^2+\frac{L_0^2}{N^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\left\|\boldsymbol{\theta}_{t-n}-\boldsymbol{\theta}_{t-n'}\right\|^2\right)-\left\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_\mu\left(\boldsymbol{\theta}_{t-n}\right)\right\|^2.\tag{31}$$

Finally, the MSE of the gradient estimator $\hat{\nabla}F(\boldsymbol{\theta}_{t-1})$ with respect to $\nabla F_\mu(\boldsymbol{\theta}_{t-1})$ can be bounded as below:

$$\begin{aligned}
&\mathbb{E}\left[\left\|\hat{\nabla}F(\boldsymbol{\theta}_{t-1})-\nabla F_\mu(\boldsymbol{\theta}_{t-1})\right\|^2\right]\\
\overset{(a)}{\leq}&\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{t-1})\right)+\left\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_\mu\left(\boldsymbol{\theta}_{t-n}\right)-\nabla F_\mu(\boldsymbol{\theta}_{t-1})\right\|^2\\
\overset{(b)}{\leq}&\frac{1}{NK\mu^2}\left(\sigma_\xi^2+\sigma_\mu^2+\frac{L_0^2}{N^2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\left\|\boldsymbol{\theta}_{t-n}-\boldsymbol{\theta}_{t-n'}\right\|^2\right)+\frac{L_1^2d}{N}\sum_{n=1}^{N}\left\|\boldsymbol{\theta}_{t-n}-\boldsymbol{\theta}_{t-1}\right\|^2-\left\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_\mu\left(\boldsymbol{\theta}_{t-n}\right)\right\|^2,
\end{aligned}\tag{32}$$

where $(a)$ follows from the fact that $\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{t-1})\right)$ and $\left\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_\mu\left(\boldsymbol{\theta}_{t-n}\right)-\nabla F_\mu(\boldsymbol{\theta}_{t-1})\right\|^2$ with respect to $\{\xi_\tau\}_\tau^t$ are independent, while $(b)$ is derived from the subsequent inequality using Jensen's inequality and Assump. B.1:

$$\left\|\frac{1}{N}\sum_{n=1}^{N}\nabla F_\mu\left(\boldsymbol{\theta}_{t-n}\right)-\nabla F_\mu(\boldsymbol{\theta}_{t-1})\right\|^2\leq\frac{1}{N}\sum_{n=1}^{N}\left\|\nabla F_\mu\left(\boldsymbol{\theta}_{t-n}\right)-\nabla F_\mu(\boldsymbol{\theta}_{t-1})\right\|^2\leq\frac{L_1^2d}{N}\sum_{n=1}^{N}\left\|\boldsymbol{\theta}_{t-n}-\boldsymbol{\theta}_{t-1}\right\|^2.\tag{33}$$

Considering the update rule of $\mathcal{R}$-AdaZO, (32) can be further simplified as:

$$\mathbb{E}\left[\left\|\hat{\nabla}F(\boldsymbol{\theta}_{t-1})-\nabla F_\mu(\boldsymbol{\theta}_{t-1})\right\|^2\right]\leq\frac{\sigma_\xi^2+\sigma_\mu^2}{NK\mu^2}+\frac{L_0^2\eta^2d\left(N^2-1\right)}{3\left(1-\beta_2\right)N^2K\mu^2}+\frac{L_1^2\eta^2d^2\left(N-1\right)}{2\left(1-\beta_2\right)}\,, \tag{34}$$

where we perform $\frac{\boldsymbol{m}_{t,i}}{\sqrt{\boldsymbol{v}_{t,i}}}\leq\frac{1}{\sqrt{1-\beta_2}}$ in the following inequality:

$$\left\|\boldsymbol{\theta}_{t-n}-\boldsymbol{\theta}_{t-n'}\right\|^2\leq\sum_{n''=\min(n,n')}^{\max(n,n')-1}\left\|\frac{\eta\boldsymbol{m}_{t-n''}}{\sqrt{\boldsymbol{v}_{t-n''}+\zeta}}\right\|^2\leq\sum_{n''=\min(n,n')}^{\max(n,n')-1}\frac{\eta^2d}{1-\beta_2}=\frac{\eta^2d}{1-\beta_2}\left|n-n'\right|\,. \tag{35}$$

□

### C.7. Proof of Thm. B.6

To ease the proof of Thm. B.6, we first prove the smoothness of $F_\mu$ (Lemma C.1), then the upper bound of the squared first moment $\boldsymbol{m}_{t,i}^2$ (Lemma C.2) and the second moment $\boldsymbol{v}_{t,i}$ (Lemma C.3).

**Lemma C.1.** *$\forall\boldsymbol{\theta},\boldsymbol{\theta}'\in\mathbb{R}^d$, we have*

$$\left|\nabla_iF_\mu(\boldsymbol{\theta})-\nabla_iF_\mu(\boldsymbol{\theta}')\right|\leq L_1\left\|\boldsymbol{\theta}-\boldsymbol{\theta}'\right\|\,.$$

*Proof.*

$$\begin{aligned}\left|\nabla_iF_\mu(\boldsymbol{\theta})-\nabla_iF_\mu(\boldsymbol{\theta}')\right|&=\left|\mathbb{E}_{\mathbf{u}}\left[\nabla_iF(\boldsymbol{\theta}+\mu\mathbf{u})-\nabla_iF(\boldsymbol{\theta}'+\mu\mathbf{u})\right]\right|\\&\overset{(a)}{\leq}\mathbb{E}_{\mathbf{u}}\left[\left|\nabla_iF(\boldsymbol{\theta}+\mu\mathbf{u})-\nabla_iF(\boldsymbol{\theta}'+\mu\mathbf{u})\right|\right]\\&\overset{(b)}{\leq}L_1\left\|\boldsymbol{\theta}-\boldsymbol{\theta}'\right\|\,,\end{aligned} \tag{36}$$

where $(a)$ comes from the Jensen's inequality and $(b)$ follows from Assump. B.2. □

**Lemma C.2.** *$\forall\boldsymbol{\theta}\in\mathbb{R}^d$, $i\in[d]$ and $t\in[T]$, if $\boldsymbol{m}_{0,i}=0$, the following inequality holds for ZoAR,*

$$\begin{aligned}\boldsymbol{m}_{t,i}^2\leq&2\left(1+\beta_1\right)\left(1-\beta_1\right)^2\sum_{\tau=1}^t\beta_1^{2(t-\tau)}\left|\hat{\nabla}_if(\boldsymbol{\theta}_{\tau-1};\xi_\tau)-\frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_iF_\mu(\boldsymbol{\theta}_{\tau-n})\right|^2\\&+\frac{2\beta_1(1+\beta_1)^2}{\left(1-\beta_1\right)^2\left(1-\beta_2\right)}\eta^2L_1^2dC_N+2\frac{(1+\beta_1)^2}{\beta_1}\left|\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2\,,\end{aligned}$$

*where $C_N\triangleq\frac{2(1-\beta_1)^2N^2-3(1-\beta_1)(1-3\beta_1)N-\beta_1(2-13\beta_1)+1}{6\beta_1(1+\beta_1)}$ is monotonously increasing in $N$ and satisfies $C_N=1$ when $N=1$.*

*Proof.* First of all, the square of the first moment $\boldsymbol{m}_{t,i}^2$ can be bounded as below:

$$\begin{aligned}\boldsymbol{m}_{t,i}^2=&\left|\boldsymbol{m}_{t,i}-\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})+\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2\\\leq&\left(1+\beta_1\right)\left|\boldsymbol{m}_{t,i}-\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2+\left(1+\frac{1}{\beta_1}\right)\left|\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2\,,\end{aligned} \tag{37}$$

where we apply the inequality $\left(a+b\right)^2\leq\left(1+\beta_1\right)a^2+\left(1+\frac{1}{\beta_1}\right)b^2$.

The first term of (37) can be further bounded:

$$\begin{aligned}&\left|\boldsymbol{m}_{t,i}-\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2\\=&\left|\boldsymbol{m}_{t,i}-\mathbb{E}\left[\boldsymbol{m}_{t,i}\right]+\mathbb{E}\left[\boldsymbol{m}_{t,i}\right]-\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2\\\overset{(a)}{\leq}&2\left|\boldsymbol{m}_{t,i}-\mathbb{E}\left[\boldsymbol{m}_{t,i}\right]\right|^2+2\left|\mathbb{E}\left[\boldsymbol{m}_{t,i}\right]-\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2\\=&2\left|\boldsymbol{m}_{t,i}-\mathbb{E}\left[\boldsymbol{m}_{t,i}\right]\right|^2+2\left|\mathbb{E}\left[\boldsymbol{m}_{t,i}\right]-\left(1-\beta_1^t\right)\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})-\beta_1^t\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2\\\overset{(b)}{\leq}&2\left|\boldsymbol{m}_{t,i}-\mathbb{E}\left[\boldsymbol{m}_{t,i}\right]\right|^2+2\left(1-\beta_1^t\right)\left|\frac{\mathbb{E}\left[\boldsymbol{m}_{t,i}\right]}{1-\beta_1^t}-\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2+2\beta_1^t\left|\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2\,,\end{aligned} \tag{38}$$

where $(a)$ and $(b)$ utilize the inequality $(a+b)^2 \leq (1+k)\,a^2 + \left(1+\frac{1}{k}\right) b^2$ for any $k>0$, with $k=1$ in step $(a)$ and $k=\frac{1-\beta_1^t}{\beta_1^t}$ in step $(b)$.

We next bound the first and second terms of (38) separately. First, by assuming $\boldsymbol{m}_{0,i}=0$, the geometric series of $\boldsymbol{m}_{t,i}$ and $\mathbb{E}\left[\boldsymbol{m}_{t,i}\right]$ are given by:

$$\begin{aligned}\boldsymbol{m}_{t,i} =&(1-\beta_1)\sum_{\tau=1}^{t}\beta_1^{t-\tau}\hat{\nabla}_i f(\boldsymbol{\theta}_{\tau-1};\xi_\tau);\\ \mathbb{E}\left[\boldsymbol{m}_{t,i}\right] =&(1-\beta_1)\sum_{\tau=1}^{t}\beta_1^{t-\tau}\frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})\,.\end{aligned} \tag{39}$$

Therefore, the first term of (38) can be bounded as:

$$\begin{aligned}\left|\boldsymbol{m}_{t,i}-\mathbb{E}\left[\boldsymbol{m}_{t,i}\right]\right|^2 =&\left|(1-\beta_1)\sum_{\tau=1}^{t}\beta_1^{t-\tau}\left(\hat{\nabla}_i f(\boldsymbol{\theta}_{\tau-1};\xi_\tau)-\frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})\right)\right|^2\\ \overset{(a)}{=}&(1-\beta_1)^2\sum_{\tau=1}^{t}\beta_1^{2(t-\tau)}\left|\hat{\nabla}_i f(\boldsymbol{\theta}_{\tau-1};\xi_\tau)-\frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})\right|^2,\end{aligned} \tag{40}$$

where $(a)$ results from the independence of different $\{\xi_\tau\}_{\tau=1}^t$.

Besides, the second term of (38) can be bounded as below:

$$\begin{aligned}&\left|\frac{\mathbb{E}\left[\boldsymbol{m}_{t,i}\right]}{1-\beta_1^t}-\nabla_i F_\mu(\boldsymbol{\theta}_{t-1})\right|^2\\ =&\left|\frac{(1-\beta_1)}{1-\beta_1^t}\sum_{\tau=1}^{t}\beta_1^{t-\tau}\left(\frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})-\nabla_i F_\mu(\boldsymbol{\theta}_{t-1})\right)\right|^2\\ =&\frac{(1-\beta_1)^2}{(1-\beta_1^t)^2}\sum_{\tau,\tau'=1}^{t}\beta_1^{2t-\tau-\tau'}\left(\frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})-\nabla_i F_\mu(\boldsymbol{\theta}_{t-1})\right)\\ &\qquad\times\left(\frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau'-n})-\nabla_i F_\mu(\boldsymbol{\theta}_{t-1})\right)\\ \overset{(a)}{\leq}&\frac{(1-\beta_1)^2}{2\,(1-\beta_1^t)^2}\sum_{\tau,\tau'=1}^{t}\beta_1^{2t-\tau-\tau'}\left(\left|\frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})-\nabla_i F_\mu(\boldsymbol{\theta}_{t-1})\right|^2\right.\\ &\qquad\left.+\left|\frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})-\nabla_i F_\mu(\boldsymbol{\theta}_{t-1})\right|^2\right)\\ =&\frac{1-\beta_1}{1-\beta_1^t}\sum_{\tau=1}^{t}\beta_1^{t-\tau}\left|\frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})-\nabla_i F_\mu(\boldsymbol{\theta}_{t-1})\right|^2\\ \overset{(b)}{\leq}&\frac{1-\beta_1}{1-\beta_1^t}\sum_{\tau=1}^{t}\beta_1^{t-\tau}\frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\left|\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})-\nabla_i F_\mu(\boldsymbol{\theta}_{t-1})\right|^2\\ \overset{(c)}{\leq}&\frac{1-\beta_1}{1-\beta_1^t}L_1^2\sum_{\tau=1}^{t}\beta_1^{t-\tau}\frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\left\|\boldsymbol{\theta}_{\tau-n}-\boldsymbol{\theta}_{t-1}\right\|^2,\end{aligned} \tag{41}$$

where $(a)$ is derived from $ab\leq\frac{1}{2}(a^2+b^2)$, $(b)$ is due to Jensen's inequality, and $(c)$ is obtained from Lemma C.1.

Recalled to the update rule of $\mathcal{R}$-AdaZO, we have:

$$\begin{aligned}\|\boldsymbol{\theta}_{\tau-n}-\boldsymbol{\theta}_{t-1}\|^2=&\sum_i^d|\boldsymbol{\theta}_{\tau-n,i}-\boldsymbol{\theta}_{t-1,i}|^2=\eta^2\sum_i^d\left|\sum_{s=\tau-n+1}^{t-1}\frac{\boldsymbol{m}_{s,i}}{\sqrt{\boldsymbol{v}_{s,i}+\zeta}}\right|^2\\ \overset{(a)}{\leq}&\eta^2\left(t-\tau+n-1\right)\sum_i^d\sum_{s=\tau-n+1}^{t-1}\frac{\boldsymbol{m}_{s,i}^2}{\boldsymbol{v}_{s,i}+\zeta}\overset{(b)}{\leq}\frac{d}{1-\beta_2}\eta^2\left(t-\tau+n-1\right)^2\ ,\end{aligned}\tag{42}$$

where $(a)$ is from Cauchy-Schwarz inequality $\left|\sum_{s=\tau-n+1}^{t-1}a_s\right|^2\leq(t-\tau+n-1)\sum_{s=\tau-n+1}^{t-1}a_s^2$, and $(b)$ follows from $\frac{\boldsymbol{m}_{s,i}^2}{\boldsymbol{v}_{s,i}+\zeta}\leq\frac{1}{1-\beta_2}$.

Putting the result of (42) into (41), we have:

$$\begin{aligned}\left|\frac{\mathbb{E}\left[\boldsymbol{m}_{t,i}\right]}{1-\beta_1^t}-\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2\leq&\frac{(1-\beta_1)\eta^2L_1^2d}{(1-\beta_1^t)(1-\beta_2)}\sum_{\tau=1}^t\frac{\beta_1^{t-\tau}}{\min(N,\tau)}\sum_{n=1}^{\min(N,\tau)}\left(t-\tau+n-1\right)^2\\ \overset{(a)}{\leq}&\frac{\beta_1(1+\beta_1)}{(1-\beta_1^t)(1-\beta_1)^2(1-\beta_2)}\eta^2L_1^2dC_N\ ,\end{aligned}\tag{43}$$

where $(a)$ comes from the geometric series summation over $\tau$ and $n$:

$$\sum_{\tau=1}^t\frac{1}{\min(N,\tau)}\sum_{n=1}^{\min(N,\tau)}\beta_1^{t-\tau}\left(t-\tau+n-1\right)^2\leq\frac{\beta_1(1+\beta_1)}{(1-\beta_1)^3}C_N\ ,\tag{44}$$

where $C_N\triangleq\frac{2(1-\beta_1)^2N^2-3(1-\beta_1)(1-3\beta_1)N-\beta_1(2-13\beta_1)+1}{6\beta_1(1+\beta_1)}$ is monotonously increasing in $N$ and satisfies $C_N=1$ when $N=1$.

Finally, gathering the results of (40), (43) into (38), we obtain:

$$\begin{aligned}|\boldsymbol{m}_{t,i}-\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})|^2\leq&(1-\beta_1)^2\sum_{\tau=1}^t\beta_1^{2(t-\tau)}\left|\hat{\nabla}_if(\boldsymbol{\theta}_{\tau-1};\xi_\tau)-\frac{1}{\min(N,\tau)}\sum_{n=1}^{\min(N,\tau)}\nabla_iF_\mu(\boldsymbol{\theta}_{\tau-n})\right|^2\\ &+\frac{\beta_1(1+\beta_1)}{(1-\beta_1)^2(1-\beta_2)}\eta^2L_1^2dC_N+\beta_1^t\left|\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2\\ &+2\left(m_{\tau,i}-\mathbb{E}\left[\boldsymbol{m}_{\tau,i}\right]\right)\left(\mathbb{E}\left[\boldsymbol{m}_{\tau,i}\right]-\nabla_iF_\mu(\theta_{\tau-1})\right)\ .\end{aligned}\tag{45}$$

Consequently, (37) becomes:

$$\begin{aligned}\boldsymbol{m}_{t,i}^2\leq&2\left(1+\beta_1\right)(1-\beta_1)^2\sum_{\tau=1}^t\beta_1^{2(t-\tau)}\left|\hat{\nabla}_if(\boldsymbol{\theta}_{\tau-1};\xi_\tau)-\frac{1}{\min(N,\tau)}\sum_{n=1}^{\min(N,\tau)}\nabla_iF_\mu(\boldsymbol{\theta}_{\tau-n})\right|^2\\ &+\frac{2\beta_1(1+\beta_1)^2}{(1-\beta_1)^2(1-\beta_2)}\eta^2L_1^2dC_N+2\left(1+\beta_1\right)\beta_1^t\left|\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2+2\left(1+\frac{1}{\beta_1}\right)\left|\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2\\ \leq&2\left(1+\beta_1\right)(1-\beta_1)^2\sum_{\tau=1}^t\beta_1^{2(t-\tau)}\left|\hat{\nabla}_if(\boldsymbol{\theta}_{\tau-1};\xi_\tau)-\frac{1}{\min(N,\tau)}\sum_{n=1}^{\min(N,\tau)}\nabla_iF_\mu(\boldsymbol{\theta}_{\tau-n})\right|^2\\ &+\frac{2\beta_1(1+\beta_1)^2}{(1-\beta_1)^2(1-\beta_2)}\eta^2L_1^2dC_N+\frac{2(1+\beta_1)^2}{\beta_1}\left|\nabla_iF_\mu(\boldsymbol{\theta}_{t-1})\right|^2\ ,\end{aligned}\tag{46}$$

which concludes the proof. □

**Lemma C.3.** *$\forall \boldsymbol{\theta} \in \mathbb{R}^d$, $i \in [d]$ and $t \in [T]$, if $\boldsymbol{v}_{0,i} > 0$, the following inequality holds for ZoAR,*

$$\begin{aligned}\boldsymbol{v}_{t,i} \leq & \beta_2^t \boldsymbol{v}_{0,i} + \frac{2\left(1+\beta_1\right)\left(1-\beta_1\right)^2\left(1-\beta_2\right)}{\beta_2-\beta_1^2} \sum_{\tau=1}^{t}\left(\beta_2^{t+1-\tau}-\beta_1^{2(t+1-\tau)}\right)\left|\hat{\nabla}_i f(\boldsymbol{\theta}_{\tau-1};\xi_\tau) - \frac{1}{\min\left(N,\tau\right)} \sum_{n=1}^{\min(N,\tau)} \nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})\right|^2 \\ & + \frac{2\beta_1(1+\beta_1)^2}{\left(1-\beta_1\right)^2\left(1-\beta_2\right)} L_1^2\eta^2 d C_N + \frac{2\left(1+\beta_1\right)^2\left(1-\beta_2\right)}{\beta_1} \sum_{\tau=1}^{t} \beta_2^{t-\tau} |\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-1})|^2 ,\end{aligned}$$

*where $C_N \triangleq \frac{2(1-\beta_1)^2N^2-3(1-\beta_1)(1-3\beta_1)N-\beta_1(2-13\beta_1)+1}{6\beta_1(1+\beta_1)}$ is monotonously increasing in $N$ and satisfies $C_N = 1$ when $N = 1$.*

*Proof.* Starting the geometric series of $\boldsymbol{v}_{t,i}$, the following inequality holds:

$$\begin{aligned}\boldsymbol{v}_{t,i} &= \beta_2^t \boldsymbol{v}_{0,i} + (1-\beta_2)\sum_{\tau=1}^{t}\beta_2^{t-\tau}\boldsymbol{m}_{\tau,i}^2 \\ &\overset{(a)}{\leq} \beta_2^t \boldsymbol{v}_{0,i} + (1-\beta_2)\sum_{\tau=1}^{t}\beta_2^{t-\tau}\left(2\left(1+\beta_1\right)\left(1-\beta_1\right)^2\sum_{s=1}^{\tau}\beta_1^{2(\tau-s)}\left|\hat{\nabla}_i f(\boldsymbol{\theta}_{s-1};\xi_s) - \frac{1}{\min\left(N,s\right)}\sum_{n=1}^{\min(N,s)}\nabla_i F_\mu(\boldsymbol{\theta}_{s-n})\right|^2\right. \\ &\quad \left. + \frac{2\beta_1(1+\beta_1)^2}{\left(1-\beta_1\right)^2\left(1-\beta_2\right)} L_1^2\eta^2 d C_N + \frac{2\left(1+\beta_1\right)^2}{\beta_1}|\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-1})|^2\right) \\ &\overset{(b)}{=} \beta_2^t \boldsymbol{v}_{0,i} + \frac{2\left(1+\beta_1\right)\left(1-\beta_1\right)^2\left(1-\beta_2\right)}{\beta_2-\beta_1^2}\sum_{\tau=1}^{t}\left(\beta_2^{t+1-\tau}-\beta_1^{2(t+1-\tau)}\right)\left|\hat{\nabla}_i f(\boldsymbol{\theta}_{\tau-1};\xi_\tau) - \frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})\right|^2 \\ &\quad + \frac{2\beta_1(1+\beta_1)^2\left(1-\beta_2^t\right)}{\left(1-\beta_1\right)^2\left(1-\beta_2\right)} L_1^2\eta^2 d C_N + \frac{2\left(1+\beta_1\right)^2\left(1-\beta_2\right)}{\beta_1}\sum_{\tau=1}^{t}\beta_2^{t-\tau}|\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-1})|^2 \\ &\overset{(c)}{\leq} \beta_2^t \boldsymbol{v}_{0,i} + \frac{2\left(1+\beta_1\right)\left(1-\beta_1\right)^2\left(1-\beta_2\right)}{\beta_2-\beta_1^2}\sum_{\tau=1}^{t}\left(\beta_2^{t+1-\tau}-\beta_1^{2(t+1-\tau)}\right)\left|\hat{\nabla}_i f(\boldsymbol{\theta}_{\tau-1};\xi_\tau) - \frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})\right|^2 \\ &\quad + \frac{2\beta_1(1+\beta_1)^2}{\left(1-\beta_1\right)^2\left(1-\beta_2\right)} L_1^2\eta^2 d C_N + \frac{2\left(1+\beta_1\right)^2\left(1-\beta_2\right)}{\beta_1}\sum_{\tau=1}^{t}\beta_2^{t-\tau}|\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-1})|^2 ,\end{aligned} \tag{47}$$

where $(a)$ comes from Lemma C.2, and $(c)$ is due to $\beta_2 \leq 1$. In step $(b)$, we use the following geometric series summation:

$$\begin{aligned}&\sum_{\tau=1}^{t}\sum_{s=1}^{\tau}\beta_2^{t-\tau}\beta_1^{2(\tau-s)}\left|\hat{\nabla}_i f(\boldsymbol{\theta}_{s-1};\xi_s) - \frac{1}{\min\left(N,s\right)}\sum_{n=1}^{\min(N,s)}\nabla_i F_\mu(\boldsymbol{\theta}_{s-n})\right|^2 \\ \overset{(a)}{=}& \sum_{s=1}^{t}\sum_{\tau=s}^{t}\beta_2^{t-\tau}\beta_1^{2(\tau-s)}\left|\hat{\nabla}_i f(\boldsymbol{\theta}_{s-1};\xi_s) - \frac{1}{\min\left(N,s\right)}\sum_{n=1}^{\min(N,s)}\nabla_i F_\mu(\boldsymbol{\theta}_{s-n})\right|^2 \\ =& \frac{1}{\beta_2-\beta_1^2}\sum_{\tau=1}^{t}\left(\beta_2^{t+1-\tau}-\beta_1^{2(t+1-\tau)}\right)\left|\hat{\nabla}_i f(\boldsymbol{\theta}_{\tau-1};\xi_\tau) - \frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})\right|^2 ,\end{aligned} \tag{48}$$

where we exchange the order of summation in step $(a)$. This concludes the proof. □

Here, we give the formal statement of the convergence of ZoAR in Thm. B.6.

**Theorem C.4** (Variance-Aware Convergence of $\mathcal{R}$-AdaZO, Formal)**.** *Let* $\mathbf{u} \sim \mathrm{Unif}(\mathbb{S}^{d-1})$ *and* $b_t$ *in* (13) *be the optimal* $b_t^*$ *in Thm. B.4, under the Assump. B.1, B.2, and the gradient of* $F_\mu$ *being bounded* $|\nabla_i F_\mu(\boldsymbol{\theta}_t)| \le G_\mu$*, when* $\eta = \frac{(1-\beta_1)\left(1-\beta_1/\sqrt{\beta_2}\right)\epsilon^2}{128 L_1 d^{3/2}} \sim \mathcal{O}\left(\epsilon^2\right)$, $1-\beta_2 = \min\left(\frac{(1-\beta_1)^2\mu^2\eta\sqrt{\zeta}\epsilon^2}{64C^2d^3}, \frac{(1-\beta_1)^2\epsilon^2}{4\beta_1^2 d\sqrt{\zeta}}\right) \sim \mathcal{O}\left(\epsilon^2\right)$, $T = \max\left(\frac{64C^2d^3}{(1-\beta_1)^2\mu^2\eta\sqrt{\zeta}\epsilon^2}, \frac{8\left(1-\beta_1/\sqrt{\beta_2}\right)}{(1-\beta_1)\eta\epsilon^2}, \frac{64L_1\sqrt{d}\eta}{(1-\beta_1)\left(1-\beta_1/\sqrt{\beta_2}\right)(1-\beta_2)\epsilon^2}\sum_i^d \ln\left(\frac{\beta_2^T \boldsymbol{v}_{0,i} + 4C^2d^2/\mu^2}{\boldsymbol{v}_{0,i}}\right)\right) \sim \mathcal{O}\left(\epsilon^{-4}\right)$, $\beta_1 \le \sqrt{\beta_2}$, $\beta_2 > 1/2$, $\boldsymbol{m}_{0,i} = 0$, $\boldsymbol{v}_{0,i} > 0$ $(\forall i \in [d])$*, the following convergence holds for ZoAR (Algo. 1),*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\|\nabla F(\boldsymbol{\theta}_t)\|] \le \sqrt{\frac{2}{\beta_1(1-\beta_2)}}(1+\beta_1)\epsilon^2 + \left(\sqrt[4]{\zeta} + \sqrt{\Xi}\right)\epsilon + \mu L_1\sqrt{d} + B_2\,, \tag{49}$$

*where* $B_1 \triangleq \sqrt{d\beta_2\|\boldsymbol{v}_0\| + \frac{2\beta_1(1+\beta_1)^2}{(1-\beta_1)^2(1-\beta_2)}L_1^2\eta^2d^2C_N} + \sqrt{\frac{2(1+\beta_1)(1-\beta_1)^2\beta_2 L_0^2\eta^2 d(N^2-1)}{3\left(\beta_2-\beta_1^2\right)(1-\beta_2)^2N^2K\mu^2}}$, $\Xi \triangleq B_1 + \sqrt{\frac{2(1+\beta_1)(1-\beta_1)^2\beta_2}{\left(\beta_2-\beta_1^2\right)(1-\beta_2)}V}$, $V = \frac{\sigma_\xi^2+\sigma_\mu^2}{NK\mu^2}$, $G \triangleq \frac{2G_\mu}{\sqrt{\zeta}}\sqrt{d\left(V + \frac{L_0^2\eta^2d(N^2-1)}{3(1-\beta_2)N^2K\mu^2} + \frac{L_1^2\eta^2d^2(N-1)}{2(1-\beta_2)}\right)}$, $B_2 \triangleq \sqrt{\frac{2}{\beta_1(1-\beta_2)}}(1+\beta_1)G + \left(\sqrt[4]{\zeta}+\sqrt{\Xi}\right)\sqrt{G}$*, and* $C_N \triangleq \frac{2(1-\beta_1)^2N^2 - 3(1-\beta_1)(1-3\beta_1)N - \beta_1(2-13\beta_1)+1}{6\beta_1(1+\beta_1)}$ *is monotonously increasing in* $N$ *and satisfies* $C_N = 1$ *when* $N = 1$.

*Proof.* We begin by introducing the following transformation:

$$\begin{aligned}
\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_t)\|\right]\right)^2 &= \left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\sqrt[4]{\beta_2\|\boldsymbol{v}_t\|+\zeta}\cdot\frac{\|\nabla F_\mu(\boldsymbol{\theta}_t)\|}{\sqrt[4]{\beta_2\|\boldsymbol{v}_t\|+\zeta}}\right]\right)^2 \\
&\overset{(a)}{\le} \frac{1}{T^2}\left(\sum_{t=1}^{T}\mathbb{E}\left[\sqrt{\beta_2\|\boldsymbol{v}_t\|+\zeta}\right]^{\frac{1}{2}}\cdot\mathbb{E}\left[\frac{\|\nabla F_\mu(\boldsymbol{\theta}_t)\|^2}{\sqrt{\beta_2\|\boldsymbol{v}_t\|+\zeta}}\right]^{\frac{1}{2}}\right)^2 \\
&\overset{(b)}{\le} \underbrace{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\sqrt{\beta_2\|\boldsymbol{v}_t\|+\zeta}\right]}_{\text{Term I}}\cdot\underbrace{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\frac{\|\nabla F_\mu(\boldsymbol{\theta}_t)\|^2}{\sqrt{\beta_2\|\boldsymbol{v}_t\|+\zeta}}\right]}_{\text{Term II}},
\end{aligned} \tag{50}$$

where $(a)$ comes from the Hölder's inequality $\mathbb{E}[|\boldsymbol{a}\boldsymbol{b}|] \le \left(\mathbb{E}\left[|\boldsymbol{a}|^2\right]\right)^{\frac{1}{2}}\left(\mathbb{E}\left[|\boldsymbol{b}|^2\right]\right)^{\frac{1}{2}}$, and $(b)$ results from the Cauchy-Schwarz inequality.

**Calculation of Term I.** Based on Lemma C.3, $\boldsymbol{v}_{0,i} \le \|\boldsymbol{v}_0\|$, and $\beta_2 \le 1$, we have:

$$\begin{aligned}
\boldsymbol{v}_{t,i} \le& \beta_2\|\boldsymbol{v}_0\| + \frac{2(1+\beta_1)(1-\beta_1)^2(1-\beta_2)}{\beta_2-\beta_1^2}\sum_{\tau=1}^{t}\left(\beta_2^{t+1-\tau} - \beta_1^{2(t+1-\tau)}\right)\left|\hat{\nabla}_i f(\boldsymbol{\theta}_{\tau-1};\xi_\tau) - \frac{1}{\min(N,\tau)}\sum_{n=1}^{\min(N,\tau)}\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-n})\right|^2 \\
&+ \frac{2\beta_1(1+\beta_1)^2}{(1-\beta_1)^2(1-\beta_2)}L_1^2\eta^2 dC_N + \frac{2(1+\beta_1)^2(1-\beta_2)}{\beta_1}\sum_{\tau=1}^{t}\beta_2^{t-\tau}|\nabla_i F_\mu(\boldsymbol{\theta}_{\tau-1})|^2\,.
\end{aligned} \tag{51}$$

Therefore, the square root of the summed second moment can be bounded as follows:

$$\begin{aligned}
\sqrt{\sum_i^d \boldsymbol{v}_{t,i}} \le& \sqrt{d\beta_2\|\boldsymbol{v}_0\| + \frac{2\beta_1(1+\beta_1)^2}{(1-\beta_1)^2(1-\beta_2)}L_1^2\eta^2d^2C_N} + \sqrt{\frac{2(1-\beta_2)}{\beta_1}}(1+\beta_1)\sum_{\tau=1}^{t}\beta_2^{\frac{t-\tau}{2}}\|\nabla F_\mu(\theta_{\tau-1})\| \\
&+ \sqrt{\frac{2(1+\beta_1)(1-\beta_2)}{\beta_2-\beta_1^2}}(1-\beta_1)\sum_{\tau=1}^{t}\sqrt{\beta_2^{t+1-\tau}-\beta_1^{2(t+1-\tau)}}\left\|\hat{\nabla} f(\boldsymbol{\theta}_{\tau-1};\xi_\tau) - \frac{1}{\min(N,\tau)}\sum_{n=1}^{\min(N,\tau)}\nabla F_\mu(\boldsymbol{\theta}_{\tau-n})\right\|,
\end{aligned} \tag{52}$$

where we utilize the inequality $\sqrt{\sum_i a_i} \le \sum_i \sqrt{a_i}$.

Subsequently, the expectation of (52) can be bounded as follows:

$$\begin{aligned}\mathbb{E}\left[\sqrt{\sum_{i}^{d}\boldsymbol{v}_{t,i}}\right] \leq & \sqrt{d\beta_2\left\|\boldsymbol{v}_0\right\| + \frac{2\beta_1(1+\beta_1)^2}{\left(1-\beta_1\right)^2\left(1-\beta_2\right)}L_1^2\eta^2d^2C_N} + \sqrt{\frac{2\left(1-\beta_2\right)}{\beta_1}}\left(1+\beta_1\right)\sum_{\tau=1}^{t}\beta_2^{\frac{t-\tau}{2}}\mathbb{E}\left[\left\|\nabla_i F_\mu(\theta_{\tau-1})\right\|\right] \\ & + \sqrt{\frac{2\left(1+\beta_1\right)\left(1-\beta_2\right)}{\beta_2-\beta_1^2}}\left(1-\beta_1\right)\sum_{\tau=1}^{t}\sqrt{\beta_2^{t+1-\tau}-\beta_1^{2(t+1-\tau)}}\sqrt{\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{\tau-1})\right)},\end{aligned} \tag{53}$$

where we apply the following inequality by Jensen's inequality:

$$\begin{aligned}\mathbb{E}\left[\left\|\hat{\nabla}f(\boldsymbol{\theta}_{\tau-1};\xi_\tau) - \frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla F_\mu(\boldsymbol{\theta}_{\tau-n})\right\|\right] \leq & \sqrt{\mathbb{E}\left[\left\|\hat{\nabla}f(\boldsymbol{\theta}_{\tau-1};\xi_\tau) - \frac{1}{\min\left(N,\tau\right)}\sum_{n=1}^{\min(N,\tau)}\nabla F_\mu(\boldsymbol{\theta}_{\tau-n})\right\|^2\right]} \\ = & \sqrt{\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{\tau-1})\right)},\end{aligned} \tag{54}$$

where the definition of $\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{\tau-1})\right)$ comes from Thm. B.4.

Considering the average over all iterations $t$, we have:

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\sqrt{\sum_{i}^{d}\boldsymbol{v}_{t,i}}\right] \leq & \sqrt{d\beta_2\left\|\boldsymbol{v}_0\right\| + \frac{2\beta_1(1+\beta_1)^2}{\left(1-\beta_1\right)^2\left(1-\beta_2\right)}L_1^2\eta^2d^2C_N} \\ & + \sqrt{\frac{2\left(1+\beta_1\right)\left(1-\beta_2\right)}{\beta_2-\beta_1^2}}\left(1-\beta_1\right)\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{t}\sqrt{\beta_2^{t+1-\tau}-\beta_1^{2(t+1-\tau)}}\sqrt{\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{\tau-1})\right)} \\ & + \sqrt{\frac{2\left(1-\beta_2\right)}{\beta_1}}\left(1+\beta_1\right)\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{t}\beta_2^{\frac{t-\tau}{2}}\mathbb{E}\left[\left\|\nabla F_\mu(\theta_{\tau-1})\right\|\right].\end{aligned} \tag{55}$$

The second and third terms in (55) contain double geometric series summations over $\tau$ and $t$. For the second term in (55),

we have:

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{t}\sqrt{\left(\beta_2^{t+1-\tau}-\beta_1^{2(t+1-\tau)}\right)}\sqrt{\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{\tau-1})\right)}\\
\overset{(a)}{\leq}&\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{t}\left(\beta_2^{\frac{t+1-\tau}{2}}-\beta_1^{t+1-\tau}\right)\sqrt{\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{\tau-1})\right)}\\
\overset{(b)}{\leq}&\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{t}\beta_2^{\frac{t+1-\tau}{2}}\sqrt{\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{\tau-1})\right)}\\
\overset{(c)}{=}&\frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=\tau}^{T}\beta_2^{\frac{t+1-\tau}{2}}\sqrt{\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{\tau-1})\right)}\\
=&\frac{1}{T}\sum_{t=1}^{T}\frac{\sqrt{\beta_2}-\sqrt{\beta_2^{2+T-t}}}{1-\sqrt{\beta_2}}\sqrt{\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{t-1})\right)}\\
\overset{(d)}{\leq}&\frac{\sqrt{\beta_2}}{1-\beta_2}\frac{1}{T}\sum_{t=1}^{T}\sqrt{\mathrm{Var}\left(\hat{\nabla}F(\boldsymbol{\theta}_{t-1})\right)}\\
\overset{(e)}{\leq}&\frac{\sqrt{\beta_2}}{1-\beta_2}\sqrt{V+\frac{L_0^2\eta^2 d\left(N^2-1\right)}{3\left(1-\beta_2\right)N^2K\mu^2}}\\
\overset{(f)}{\leq}&\frac{\sqrt{\beta_2}}{1-\beta_2}\left(\sqrt{V}+\sqrt{\frac{L_0^2\eta^2 d\left(N^2-1\right)}{3\left(1-\beta_2\right)N^2K\mu^2}}\right),
\end{aligned}
\tag{56}
$$

where $(a)$, $(f)$ results from $\sqrt{\sum_i a_i}\leq\sum_i\sqrt{a_i}$, $(b)$, $(d)$ comes from $0\leq\beta_1^2\leq\beta_2\leq 1$, and $(e)$ is due to (34) in Appx. C.6 and $V\triangleq\frac{\sigma_\xi^2+\sigma_\mu^2}{NK\mu^2}$. In step $(c)$ we exchange the order of summation over $t$ and $\tau$.

For the third term in (55), we have:

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{t}\beta_2^{\frac{t-\tau}{2}}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_{\tau-1})\|\right]\overset{(a)}{=}&\frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=\tau}^{T}\beta_2^{\frac{t-\tau}{2}}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_{\tau-1})\|\right]=\frac{1}{T}\sum_{t=1}^{T}\frac{1-\beta_2^{\frac{1+T-t}{2}}}{1-\sqrt{\beta_2}}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_{t-1})\|\right]\\
\overset{(b)}{\leq}&\frac{1}{1-\beta_2}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_{t-1})\|\right],
\end{aligned}
\tag{57}
$$

where $(b)$ is due to the fact that $\beta_2<1$. In step $(a)$, we change the summation order over $t$ and $\tau$.

Therefore, (55) can be rewritten as:

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\sqrt{\sum_i^d\boldsymbol{v}_{t,i}}\right]\leq&\sqrt{d\beta_2\left\|\boldsymbol{v}_0\right\|+\frac{2\beta_1(1+\beta_1)^2}{\left(1-\beta_1\right)^2\left(1-\beta_2\right)}L_1^2\eta^2d^2C_N}\\
&+\sqrt{\frac{2\left(1+\beta_1\right)\left(1-\beta_2\right)}{\beta_2-\beta_1^2}}\left(1-\beta_1\right)\left(\sqrt{V}+\sqrt{\frac{L_0^2\eta^2d\left(N^2-1\right)}{3\left(1-\beta_2\right)N^2K\mu^2}}\right)\\
&+\sqrt{\frac{2}{\beta_1\left(1-\beta_2\right)}}\left(1+\beta_1\right)\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_{t-1})\|\right]\\
=&\Xi+\sqrt{\frac{2}{\beta_1\left(1-\beta_2\right)}}\left(1+\beta_1\right)\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_{t-1})\|\right],
\end{aligned}
\tag{58}
$$

where we let $\Xi\triangleq B_1+\sqrt{\frac{2(1+\beta_1)(1-\beta_1)^2\beta_2}{\left(\beta_2-\beta_1^2\right)(1-\beta_2)}V}$ and $B_1\triangleq\sqrt{d\beta_2\left\|\boldsymbol{v}_0\right\|+\frac{2\beta_1(1+\beta_1)^2}{(1-\beta_1)^2(1-\beta_2)}L_1^2\eta^2d^2C_N}+\sqrt{\frac{2(1+\beta_1)(1-\beta_1)^2\beta_2L_0^2\eta^2d(N^2-1)}{3\left(\beta_2-\beta_1^2\right)(1-\beta_2)^2N^2K\mu^2}}$.

Therefore, we can bound Term I as follows:

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\sqrt{\beta_2\|\boldsymbol{v}_t\|+\zeta}\right] &\overset{(a)}{\leq}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\sqrt{\beta_2\sum_i^d \boldsymbol{v}_{t,i}+\zeta}\right] \overset{(b)}{\leq} \frac{1}{T}\sum_{t=1}^{T}\left(\sqrt{\zeta}+\sqrt{\beta_2}\mathbb{E}\left[\sqrt{\sum_i^d \boldsymbol{v}_{t,i}}\right]\right)\\ &\overset{(c)}{\leq}\sqrt{\zeta}+\Xi+\sqrt{\frac{2}{\beta_1\left(1-\beta_2\right)}}\left(1+\beta_1\right)\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_{t-1})\|\right]\ ,\end{aligned} \tag{59}$$

where $(a)$ and $(b)$ are obtained by $\sqrt{\sum_i a_i}\leq\sum_i\sqrt{a_i}$, and $(c)$ is due to (58) and $\beta_2\leq 1$.

**Calculation of Term II.** Following a similar approach as in (Shu et al., 2025b), we first introduce the following auxiliary variable:

$$\boldsymbol{x}_t\triangleq\frac{\boldsymbol{\theta}_t-\beta_1/\sqrt{\beta_2}\boldsymbol{\theta}_{t-1}}{1-\beta_1/\sqrt{\beta_2}}=\frac{\boldsymbol{\theta}_t-\kappa\boldsymbol{\theta}_{t-1}}{1-\kappa}\ , \tag{60}$$

where $\kappa\triangleq\beta_1/\sqrt{\beta_2}$.

Based on the definition of $\boldsymbol{x}_t$, the following relationships hold:

$$\boldsymbol{x}_{t+1}-\boldsymbol{x}_t=\frac{\boldsymbol{\theta}_{t+1}-\boldsymbol{\theta}_t-\kappa\left(\boldsymbol{\theta}_t-\boldsymbol{\theta}_{t-1}\right)}{1-\kappa}=\frac{1}{1-\kappa}\left(-\frac{\eta\boldsymbol{m}_{t+1}}{\sqrt{\boldsymbol{v}_{t+1}+\zeta}}+\kappa\frac{\eta\boldsymbol{m}_t}{\sqrt{\boldsymbol{v}_t+\zeta}}\right)\ , \tag{61}$$

and,

$$\boldsymbol{x}_t-\boldsymbol{\theta}_t=\frac{\kappa}{1-\kappa}\left(\boldsymbol{\theta}_t-\boldsymbol{\theta}_{t-1}\right)=-\frac{\kappa}{1-\kappa}\frac{\eta\boldsymbol{m}_t}{\sqrt{\boldsymbol{v}_t+\zeta}}\ . \tag{62}$$

Starting from Lemma C.1:

$$F_\mu(\boldsymbol{x}_{t+1})-F_\mu(\boldsymbol{x}_t)\leq\langle\nabla F_\mu(\boldsymbol{x}_t),\boldsymbol{x}_{t+1}-\boldsymbol{x}_t\rangle+\frac{L_1}{2}\sqrt{d}\|\boldsymbol{x}_{t+1}-\boldsymbol{x}_t\|^2\ . \tag{63}$$

Firstly, we focus on iteration $t$ and calculate the conditional expectation $\mathbb{E}\left[\cdots|\mathcal{F}_t\right]$ of (63), where $\mathcal{F}_t$ denotes all stochatsics up to $t$. After that:

$$\begin{aligned}&\mathbb{E}\left[F_\mu(\boldsymbol{x}_{t+1})-F_\mu(\boldsymbol{x}_t)|\mathcal{F}_t\right]\leq\mathbb{E}\left[\langle\nabla F_\mu(\boldsymbol{x}_t),\boldsymbol{x}_{t+1}-\boldsymbol{x}_t\rangle|\mathcal{F}_t\right]+\frac{L_1}{2}\sqrt{d}\mathbb{E}\left[\|\boldsymbol{x}_{t+1}-\boldsymbol{x}_t\|^2|\mathcal{F}_t\right]\\ =&\mathbb{E}\left[\langle\nabla F_\mu(\boldsymbol{\theta}_t),\boldsymbol{x}_{t+1}-\boldsymbol{x}_t\rangle|\mathcal{F}_t\right]+\underbrace{\mathbb{E}\left[\langle\nabla F_\mu(\boldsymbol{x}_t)-\nabla F_\mu(\boldsymbol{\theta}_t),\boldsymbol{x}_{t+1}-\boldsymbol{x}_t\rangle|\mathcal{F}_t\right]}_{④}+\underbrace{\frac{L_1}{2}\sqrt{d}\mathbb{E}\left[\|\boldsymbol{x}_{t+1}-\boldsymbol{x}_t\|^2|\mathcal{F}_t\right]}_{⑤}\ .\end{aligned} \tag{64}$$

With the help of (61), the first term of (64) can be separated as below:

$$
\begin{aligned}
&\mathbb{E}\left[\langle \nabla F_\mu(\boldsymbol{\theta}_t), \boldsymbol{x}_{t+1}-\boldsymbol{x}_t\rangle \,|\mathcal{F}_t\right] = \mathbb{E}\left[\left\langle \nabla F_\mu(\boldsymbol{\theta}_t), \frac{1}{1-\kappa}\left(-\frac{\eta \boldsymbol{m}_{t+1}}{\sqrt{\boldsymbol{v}_{t+1}+\zeta}}+\kappa\frac{\eta \boldsymbol{m}_t}{\sqrt{\boldsymbol{v}_t+\zeta}}\right)\right\rangle \middle| \mathcal{F}_t\right] \\
=&\frac{1}{1-\kappa}\mathbb{E}\left[\left\langle \nabla F_\mu(\boldsymbol{\theta}_t), -\frac{\eta \boldsymbol{m}_{t+1}}{\sqrt{\beta_2\boldsymbol{v}_t+\zeta}}+\beta_1\frac{\eta \boldsymbol{m}_t}{\sqrt{\beta_2\boldsymbol{v}_t+\zeta}}-\frac{\eta \boldsymbol{m}_{t+1}}{\sqrt{\boldsymbol{v}_{t+1}+\zeta}}+\frac{\eta \boldsymbol{m}_{t+1}}{\sqrt{\beta_2\boldsymbol{v}_t+\zeta}}\right\rangle \middle| \mathcal{F}_t\right] \\
&+\frac{1}{1-\kappa}\mathbb{E}\left[\left\langle \nabla F_\mu(\boldsymbol{\theta}_t), \beta_1\left(\frac{\eta \boldsymbol{m}_t}{\sqrt{\beta_2\boldsymbol{v}_t+\beta_2\zeta}}-\frac{\eta \boldsymbol{m}_t}{\sqrt{\beta_2\boldsymbol{v}_t+\zeta}}\right)\right\rangle \middle| \mathcal{F}_t\right] \\
=&\underbrace{\frac{1-\beta_1}{1-\kappa}\eta\mathbb{E}\left[\left\langle \nabla F_\mu(\boldsymbol{\theta}_t), -\frac{\hat{\nabla} f(\boldsymbol{\theta}_t;\xi_{t+1})}{\sqrt{\beta_2\boldsymbol{v}_t+\zeta}}\right\rangle \middle| \mathcal{F}_t\right]}_{①} \\
&+\underbrace{\frac{1}{1-\kappa}\eta\mathbb{E}\left[\left\langle \nabla F_\mu(\boldsymbol{\theta}_t), \boldsymbol{m}_{t+1}\left(\frac{1}{\sqrt{\beta_2\boldsymbol{v}_t+\zeta}}-\frac{1}{\sqrt{\boldsymbol{v}_{t+1}+\zeta}}\right)\right\rangle \middle| \mathcal{F}_t\right]}_{②} \\
&+\underbrace{\frac{\beta_1}{1-\kappa}\eta\mathbb{E}\left[\left\langle \nabla F_\mu(\boldsymbol{\theta}_t), \boldsymbol{m}_t\left(\frac{1}{\sqrt{\beta_2\boldsymbol{v}_t+\beta_2\zeta}}-\frac{1}{\sqrt{\beta_2\boldsymbol{v}_t+\zeta}}\right)\right\rangle \middle| \mathcal{F}_t\right]}_{③}.
\end{aligned}
\tag{65}
$$

Thereafter, we would bound term ①, ②, ③, ④ and ⑤ one by one.

For the term ①:

$$
\begin{aligned}
① =&\frac{1-\beta_1}{1-\kappa}\eta\mathbb{E}\left[\left\langle \nabla F_\mu(\boldsymbol{\theta}_t), -\frac{\hat{\nabla} f(\boldsymbol{\theta}_t;\xi_{t+1})}{\sqrt{\beta_2\boldsymbol{v}_t+\zeta}}\right\rangle \middle| \mathcal{F}_t\right] \\
=&-\frac{1-\beta_1}{1-\kappa}\eta\sum_i^d \nabla_i F_\mu(\boldsymbol{\theta}_t)\frac{\mathbb{E}\left[\hat{\nabla}_i f(\boldsymbol{\theta}_t;\xi_{t+1})\middle|\mathcal{F}_t\right]}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}} \\
=&-\frac{1-\beta_1}{1-\kappa}\eta\sum_i^d\left(\frac{|\nabla_i F_\mu(\boldsymbol{\theta}_t)|^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}+\frac{\nabla_i F_\mu(\boldsymbol{\theta}_t)}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}\mathbb{E}\left[\hat{\nabla}_i f(\boldsymbol{\theta}_t;\xi_{t+1})-\nabla_i F_\mu(\boldsymbol{\theta}_t)\middle|\mathcal{F}_t\right]\right).
\end{aligned}
\tag{66}
$$

If we assume that the gradient of $F_\mu$ is bounded, i.e. $|\nabla_i F_\mu(\boldsymbol{\theta}_t)| \le G_\mu$, we can simplify 66 as follows:

$$
① \le -\frac{1-\beta_1}{1-\kappa}\eta\sum_i^d\frac{|\nabla_i F_\mu(\boldsymbol{\theta}_t)|^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}+\frac{(1-\beta_1)\,G_\mu}{(1-\kappa)\sqrt{\zeta}}\eta\sum_i^d\mathbb{E}\left[\left|\hat{\nabla}_i f(\boldsymbol{\theta}_t;\xi_{t+1})-\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|\,|\mathcal{F}_t\right],
\tag{67}
$$

where we utilize $\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta} \le \sqrt{\zeta}$ and $a \le |a|$ for simplicity.

Now turn to the calculation of term ②. Note that:

$$
\begin{aligned}
&\frac{1}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}-\frac{1}{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}} = \frac{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}-\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}} \\
\overset{(a)}{=}&\frac{\boldsymbol{v}_{t+1,i}-\beta_2\boldsymbol{v}_{t,i}}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}\left(\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\right)} \\
=&\frac{(1-\beta_2)\,\boldsymbol{m}_{t,i}^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}\left(\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\right)},
\end{aligned}
\tag{68}
$$

where in step $(a)$, we multiply $\left(\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\right)$ in both the numerator and denominator.

Therefore, we can rewrite the term ② as:

$$
\begin{aligned}
② =& \frac{1}{1-\kappa}\eta\sum_i^d \mathbb{E}\left[\left\langle \nabla_i F_\mu(\boldsymbol{\theta}_t), \boldsymbol{m}_{t+1,i}\left(\frac{1}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}-\frac{1}{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}}\right)\right\rangle\middle|\mathcal{F}_t\right]\\
\overset{(a)}{=}& \frac{1}{1-\kappa}\eta\sum_i^d \mathbb{E}\left[\left\langle \nabla_i F_\mu(\boldsymbol{\theta}_t), \boldsymbol{m}_{t+1,i}\frac{(1-\beta_2)\,\boldsymbol{m}_{t,i}^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}\left(\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\right)}\right\rangle\middle|\mathcal{F}_t\right]\\
\overset{(b)}{\leq}& \frac{1}{1-\kappa}\eta\sum_i^d \mathbb{E}\left[\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|\frac{(1-\beta_2)\,\boldsymbol{m}_{t,i}^2\left|\boldsymbol{m}_{t+1,i}\right|}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}\left(\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\right)}\middle|\mathcal{F}_t\right]\\
\overset{(c)}{\leq}& \frac{1}{1-\kappa}\eta\sum_i^d \mathbb{E}\left[\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|\frac{\sqrt{1-\beta_2}\boldsymbol{m}_{t,i}^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\left(\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\right)}\middle|\mathcal{F}_t\right]\\
=& \frac{1}{1-\kappa}\eta\sum_i^d \frac{1}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|\mathbb{E}\left[\frac{\sqrt{1-\beta_2}\boldsymbol{m}_{t,i}^2}{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}\middle|\mathcal{F}_t\right]\\
\overset{(d)}{\leq}& \frac{1}{1-\kappa}\eta\sum_i^d \frac{1}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}\left(\frac{\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|^2}{2\gamma_0}+\frac{\gamma_0}{2}\left(\mathbb{E}\left[\frac{\sqrt{1-\beta_2}\boldsymbol{m}_{t,i}^2}{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}\middle|\mathcal{F}_t\right]\right)^2\right)\\
\overset{(e)}{\leq}& \frac{1}{1-\kappa}\eta\sum_i^d \left(\frac{1}{2\gamma_0}\frac{\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}+\frac{\gamma_0\mathbb{E}\left[\boldsymbol{m}_{t,i}^2|\mathcal{F}_t\right]}{2\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}\mathbb{E}\left[\frac{\sqrt{1-\beta_2}\boldsymbol{m}_{t,i}^2}{\left(\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\right)^2}\middle|\mathcal{F}_t\right]\right)\ ,
\end{aligned}
\tag{69}
$$

where $(a)$ comes from (68), $(b)$ is due to Cauchy-Schwarz inequality, $(c)$ is due to the fact that $\frac{|\boldsymbol{m}_{s,i}|}{\sqrt{\boldsymbol{v}_{s,i}+\zeta}}\leq\frac{1}{\sqrt{1-\beta_2}}$, $(d)$ is obtained by $ab\leq\frac{1}{\gamma_0}a^2+\frac{\gamma_0}{2}b^2$ for any positive number $\gamma_0$, and $(e)$ results from the Hölder's inequality $\mathbb{E}[|\boldsymbol{ab}|]\leq\left(\mathbb{E}\left[|\boldsymbol{a}|^2\right]\right)^{\frac{1}{2}}\left(\mathbb{E}\left[|\boldsymbol{b}|^2\right]\right)^{\frac{1}{2}}$.

Taking the Cauchy-Schwarz inequality and Assump. B.1 into account, the term $\mathbb{E}\left[\boldsymbol{m}_{t,i}^2|\mathcal{F}_t\right]$ can be bounded by:

$$
\left|\hat{\nabla}_i f(\boldsymbol{\theta};\xi)\right|\leq\frac{d}{NK}\sum_{n,k}^{N,K}\left|\frac{f\left(\boldsymbol{\theta}+\mu\boldsymbol{u}_k;\xi\right)-b_t}{\mu}\right||\boldsymbol{u}_{k,i}|\leq\frac{2Cd}{\mu}\ ,
\tag{70}
$$

$$
|\boldsymbol{m}_{t+1,i}|=\left|(1-\beta_1)\sum_{\tau=1}^t\beta_1^{t-\tau}\hat{\nabla}_i f(\boldsymbol{\theta}_{\tau-1};\xi_\tau)\right|\leq(1-\beta_1)\sum_{\tau=1}^t\beta_1^{t-\tau}\left|\hat{\nabla}_i f(\boldsymbol{\theta}_{\tau-1};\xi_\tau)\right|\leq\frac{2Cd}{\mu}\ .
\tag{71}
$$

Besides:

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\sqrt{1-\beta_2}\boldsymbol{m}_{t,i}^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\left(\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\right)^2}\middle|\mathcal{F}_t\right]\\
\overset{(a)}{\leq}&\mathbb{E}\left[\frac{\boldsymbol{v}_{t+1,i}+\zeta-(\beta_2\boldsymbol{v}_{t,i}+\zeta)}{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\left(\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\right)}\middle|\mathcal{F}_t\right]\\
=&\mathbb{E}\left[\frac{1}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}-\frac{1}{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}}\middle|\mathcal{F}_t\right]\ .
\end{aligned}
\tag{72}
$$

where in step $(a)$ we apply $\left(\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}\right)\leq\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}$.

Hence, substituting (71), (72) into (69):

$$
\begin{aligned}
② \leq& \frac{1}{1-\kappa}\eta\sum_i^d\left(\frac{1}{2\gamma_0}\frac{\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}+\frac{\gamma_0}{2}\frac{4C^2d^2}{\mu}\mathbb{E}\left[\frac{1}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}-\frac{1}{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}}\middle|\mathcal{F}_t\right]\right)\\
=&\frac{1-\beta_1}{4\left(1-\kappa\right)}\eta\sum_i^d\frac{\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}+\frac{4C^2d^2}{\left(1-\beta_1\right)\left(1-\kappa\right)\mu^2}\eta\sum_i^d\mathbb{E}\left[\frac{1}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}-\frac{1}{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}}\middle|\mathcal{F}_t\right]\ ,
\end{aligned}
\tag{73}
$$

where we let $\gamma_0 = \frac{2}{1-\beta_1}$ in the last step.

Next, the term ③ can be bounded as below:

$$
\begin{aligned}
③ =& \frac{\beta_1}{1-\kappa}\eta \sum_i^d \nabla_i F_\mu(\boldsymbol{\theta}_t) \boldsymbol{m}_{t,i} \left( \frac{1}{\sqrt{\beta_2 \boldsymbol{v}_{t,i} + \beta_2 \zeta}} - \frac{1}{\sqrt{\beta_2 \boldsymbol{v}_{t,i} + \zeta}} \right) \\
\overset{(a)}{\leq}& \frac{\beta_1}{1-\kappa}\eta \sum_i^d \left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right| \left|\boldsymbol{m}_{t,i}\right| \left| \frac{1}{\sqrt{\beta_2 \boldsymbol{v}_{t,i} + \beta_2 \zeta}} - \frac{1}{\sqrt{\beta_2 \boldsymbol{v}_{t,i} + \zeta}} \right| \\
=& \frac{\beta_1}{1-\kappa}\eta \sum_i^d \left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right| \left|\boldsymbol{m}_{t,i}\right| \left| \frac{(1-\beta_2)\,\zeta}{\sqrt{\beta_2 \boldsymbol{v}_{t,i} + \beta_2 \zeta}\sqrt{\beta_2 \boldsymbol{v}_{t,i} + \zeta}\left(\sqrt{\beta_2 \boldsymbol{v}_{t,i} + \beta_2 \zeta} + \sqrt{\beta_2 \boldsymbol{v}_{t,i} + \zeta}\right)} \right| \\
\overset{(b)}{\leq}& \frac{\beta_1}{1-\kappa}\eta \sum_i^d \frac{1}{\sqrt{\beta_2 \boldsymbol{v}_{t,i} + \zeta}} \left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right| \left| \frac{\sqrt{1-\beta_2}\zeta}{\sqrt{\beta_2}\left(\sqrt{\beta_2 \boldsymbol{v}_{t,i} + \beta_2 \zeta} + \sqrt{\beta_2 \boldsymbol{v}_{t,i} + \zeta}\right)} \right| \\
\overset{(c)}{\leq}& \frac{\beta_1}{1-\kappa}\eta \sum_i^d \left( \frac{\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|^2}{2\gamma_1 \sqrt{\beta_2 \boldsymbol{v}_{t,i} + \zeta}} + \frac{\gamma_1 (1-\beta_2)\,\zeta^2}{2\beta_2 \sqrt{\beta_2 \boldsymbol{v}_{t,i} + \zeta}\left(\sqrt{\beta_2 \boldsymbol{v}_{t,i} + \beta_2 \zeta} + \sqrt{\beta_2 \boldsymbol{v}_{t,i} + \zeta}\right)^2} \right) \\
\overset{(d)}{\leq}& \frac{\beta_1}{1-\kappa}\eta \sum_i^d \left( \frac{\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|^2}{2\gamma_1 \sqrt{\beta_2 \boldsymbol{v}_{t,i} + \zeta}} + \frac{\gamma_1 (1-\beta_2)\sqrt{\zeta}}{8\beta_2^2} \right) \\
=& \frac{1-\beta_1}{4(1-\kappa)}\eta \sum_i^d \frac{\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|^2}{\sqrt{\beta_2 \boldsymbol{v}_{t,i} + \zeta}} + \frac{\beta_1^2 (1-\beta_2)}{4\beta_2^2 (1-\beta_1)(1-\kappa)}\eta d\sqrt{\zeta} \\
\overset{(e)}{\leq}& \frac{1-\beta_1}{4(1-\kappa)}\eta \sum_i^d \frac{\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|^2}{\sqrt{\beta_2 \boldsymbol{v}_{t,i} + \zeta}} + \frac{\beta_1^2 (1-\beta_2)}{(1-\beta_1)(1-\kappa)}\eta d\sqrt{\zeta}\,,
\end{aligned}
\tag{74}
$$

where $(a)$ comes from Cauchy-Schwarz inequality, $(b)$ results from the fact that $\frac{\boldsymbol{m}_{t,i}^2}{\boldsymbol{v}_{t,i}+\zeta} \leq \frac{1}{1-\beta_2}$, $(c)$ is because of $ab \leq \frac{1}{2}(a^2+b^2)$, and $(d)$ is due to $\sqrt{\boldsymbol{v}_{t,i}+\zeta} \leq \sqrt{\zeta}$. In step $(e)$, we assume $2\beta_2 \geq 1$ and let $\gamma_1 = \frac{2\beta_1}{1-\beta_1}$.

Term ④ is bounded as below:

$$
\begin{aligned}
④ =& \sum_i^d \mathbb{E}\left[\left(\nabla_i F_\mu(\boldsymbol{x}_t) - \nabla_i F_\mu(\boldsymbol{\theta}_t)\right)\left(\boldsymbol{x}_{t+1,i} - \boldsymbol{x}_{t,i}\right)\right] \\
\overset{(a)}{\leq}& \frac{1}{1-\kappa} \sum_i^d \mathbb{E}\left[\left|\nabla_i F_\mu(\boldsymbol{x}_t) - \nabla_i F_\mu(\boldsymbol{\theta}_t)\right| \left|\boldsymbol{\theta}_{t+1,i} - \boldsymbol{\theta}_{t,i} - \kappa\left(\boldsymbol{\theta}_{t,i} - \boldsymbol{\theta}_{t-1,i}\right)\right| \left|\mathcal{F}_t\right]\right] \\
\overset{(b)}{\leq}& \frac{\kappa}{(1-\kappa)^2} L_1 \sum_i^d \mathbb{E}\left[\left\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\right\| \left|\boldsymbol{\theta}_{t+1,i} - \boldsymbol{\theta}_{t,i} - \kappa\left(\boldsymbol{\theta}_{t,i} - \boldsymbol{\theta}_{t-1,i}\right)\right| \left|\mathcal{F}_t\right]\right] \\
\overset{(c)}{\leq}& \frac{\kappa}{(1-\kappa)^2} L_1 \sum_i^d \mathbb{E}\left[\left\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\right\| \left|\boldsymbol{\theta}_{t+1,i} - \boldsymbol{\theta}_{t,i}\right| + \kappa \left\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\right\| \left|\boldsymbol{\theta}_{t,i} - \boldsymbol{\theta}_{t-1,i}\right| \left|\mathcal{F}_t\right]\right] \\
\overset{(d)}{\leq}& \frac{\kappa}{2(1-\kappa)^2} \sqrt{d} L_1 \eta^2 \sum_i^d \left( (1+2\kappa) \frac{\boldsymbol{m}_{t,i}^2}{\boldsymbol{v}_{t,i} + \zeta} + \mathbb{E}\left[ \frac{\boldsymbol{m}_{t+1,i}^2}{\boldsymbol{v}_{t+1,i} + \zeta} \middle| \mathcal{F}_t \right] \right) \\
\overset{(e)}{\leq}& \frac{1}{2(1-\kappa)^2} \sqrt{d} L_1 \eta^2 \sum_i^d \left( 3\frac{\boldsymbol{m}_{t,i}^2}{\boldsymbol{v}_{t,i} + \zeta} + \mathbb{E}\left[ \frac{\boldsymbol{m}_{t+1,i}^2}{\boldsymbol{v}_{t+1,i} + \zeta} \middle| \mathcal{F}_t \right] \right)\,,
\end{aligned}
\tag{75}
$$

where $(a)$ is due to (61) and Cauchy-Schwarz inequality, $(b)$ is due to Lemma C.1, $(c)$ comes from the fact that $|a-b| \leq$

$|a|+|b|$, and in step $(e)$ we assume $\kappa \leq 1$. In step $(d)$, we apply the following inequality by $ab \leq \frac{a^2}{2\sqrt{d}} + \frac{\sqrt{d}b^2}{2}$:

$$\begin{aligned}
&\sum_i^d \left(\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\| \left|\boldsymbol{\theta}_{t+1,i} - \boldsymbol{\theta}_{t,i}\right| + \kappa \|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\| \left|\boldsymbol{\theta}_{t,i} - \boldsymbol{\theta}_{t-1,i}\right|\right) \\
\leq &\sum_i^d \left(\frac{\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\|^2}{2\sqrt{d}} + \frac{\sqrt{d}\left|\boldsymbol{\theta}_{t+1,i} - \boldsymbol{\theta}_{t,i}\right|^2}{2} + \kappa\left(\frac{\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\|^2}{2\sqrt{d}} + \frac{\sqrt{d}\left|\boldsymbol{\theta}_{t,i} - \boldsymbol{\theta}_{t-1,i}\right|^2}{2}\right)\right) \\
= &\frac{\sqrt{d}}{2}\left((1+2\kappa)\|\boldsymbol{\theta}_t - \boldsymbol{\theta}_{t-1}\|^2 + \|\boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t\|^2\right) .
\end{aligned} \tag{76}$$

Finally, with the help of (61), the term ⑤ is bounded as below:

$$\begin{aligned}
⑤ = &\frac{1}{2(1-\kappa)^2}\sqrt{d}L_1 \sum_i^d \mathbb{E}\left[\left|\boldsymbol{\theta}_{t+1,i} - \boldsymbol{\theta}_{t,i} - \kappa\left(\boldsymbol{\theta}_{t,i} - \boldsymbol{\theta}_{t-1,i}\right)\right|^2 \middle| \mathcal{F}_t\right] \\
\overset{(a)}{\leq} &\frac{1}{(1-\kappa)^2}\sqrt{d}L_1 \sum_i^d \mathbb{E}\left[\left|\boldsymbol{\theta}_{t+1,i} - \boldsymbol{\theta}_{t,i}\right|^2 + \kappa^2\left|\boldsymbol{\theta}_{t,i} - \boldsymbol{\theta}_{t-1,i}\right| \middle| \mathcal{F}_t\right] \\
= &\frac{1}{(1-\kappa)^2}\sqrt{d}L_1\eta^2 \sum_i^d \left(\mathbb{E}\left[\frac{\boldsymbol{m}_{t+1,i}^2}{\boldsymbol{v}_{t+1,i}+\zeta}\middle|\mathcal{F}_t\right] + \kappa^2\frac{\boldsymbol{m}_{t,i}^2}{\boldsymbol{v}_{t,i}+\zeta}\right) \\
\overset{(b)}{\leq} &\frac{1}{(1-\kappa)^2}\sqrt{d}L_1\eta^2 \sum_i^d \left(\mathbb{E}\left[\frac{\boldsymbol{m}_{t+1,i}^2}{\boldsymbol{v}_{t+1,i}+\zeta}\middle|\mathcal{F}_t\right] + \frac{\boldsymbol{m}_{t,i}^2}{\boldsymbol{v}_{t,i}+\zeta}\right) ,
\end{aligned} \tag{77}$$

where $(a)$ results from the inequality $(a-b)^2 \leq 2a^2 + 2b^2$, and in step $(b)$, we assume $\kappa \leq 1$.

Gathering the results of (66), (73), (74), (75) and (77), (64) can be bounded as below:

$$\begin{aligned}
\mathbb{E}\left[F_\mu(\boldsymbol{x}_{t+1}) - F_\mu(\boldsymbol{x}_t)|\mathcal{F}_t\right] \leq &-\frac{1-\beta_1}{2(1-\kappa)}\eta\sum_i^d \frac{\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}} + \frac{\beta_1^2(1-\beta_2)}{(1-\beta_1)(1-\kappa)}\eta d\sqrt{\zeta} \\
&+ \frac{4\eta C^2 d^3}{(1-\beta_1)(1-\kappa)\mu^2}\sum_i^d \mathbb{E}\left[\frac{1}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}} - \frac{1}{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}}\middle|\mathcal{F}_t\right] \\
&+ \frac{5L_1\sqrt{d}}{2(1-\kappa)^2}\eta^2\sum_i^d \frac{\boldsymbol{m}_{t,i}^2}{\boldsymbol{v}_{t,i}+\zeta} + \frac{3L_1\sqrt{d}}{2(1-\kappa)^2}\eta^2\sum_i^d \mathbb{E}\left[\frac{\boldsymbol{m}_{t+1,i}^2}{\boldsymbol{v}_{t+1,i}+\zeta}\middle|\mathcal{F}_t\right] \\
&+ \frac{(1-\beta_1)G_\mu}{(1-\kappa)\sqrt{\zeta}}\eta\sum_i^d \mathbb{E}\left[\left|\hat{\nabla}_i f(\boldsymbol{\theta}_t;\xi_{t+1}) - \nabla_i F_\mu(\boldsymbol{\theta}_t)\right| \middle|\mathcal{F}_t\right] .
\end{aligned} \tag{78}$$

Considering the summation of (78) over all iterations $t$ from 0 to $T-1$:

$$\text{LHS} = \sum_{t=0}^{T-1}\mathbb{E}\left[F_\mu(\boldsymbol{x}_{t+1}) - F_\mu(\boldsymbol{x}_t)\right] = \mathbb{E}\left[F_\mu(\boldsymbol{x}_T)\right] - F_\mu(\boldsymbol{x}_0) \triangleq -\Delta , \tag{79}$$

$$\begin{aligned}
\text{RHS} \overset{(a)}{\leq} &-\frac{1-\beta_1}{2(1-\kappa)}\eta\sum_{t=0}^{T-1}\sum_i^d \frac{\left|\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}} + \frac{\beta_1^2(1-\beta_2)}{(1-\beta_1)(1-\kappa)}Td\sqrt{\zeta} \\
&+ \frac{4C^2d^3}{(1-\beta_1)(1-\kappa)\mu^2}\left(\frac{1}{\sqrt{\zeta}} + \frac{T(1-\beta_2)}{\sqrt{\zeta}}\right) + \frac{(1-\beta_1)}{2(1-\kappa)}\eta TG \\
&+ \frac{4L_1\sqrt{d}}{(1-\kappa)^2}\eta^2\sum_i^d\left(\frac{1}{1-\beta_2}\ln\left(\frac{\beta_2^T\boldsymbol{v}_{0,i} + 4C^2d^2/\mu^2}{\boldsymbol{v}_{0,i}}\right) + 2T\right) ,
\end{aligned} \tag{80}$$

where $G \triangleq \frac{2G_\mu}{\sqrt{\zeta}}\sqrt{d\left(V+\frac{L_0^2\eta^2 d(N^2-1)}{3(1-\beta_2)N^2K\mu^2}+\frac{L_1^2\eta^2d^2(N-1)}{2(1-\beta_2)}\right)}$ is a constant number. In step $(a)$, we apply the following three inequalities. The first one is:

$$\begin{aligned}
&\sum_{t=0}^{T}\sum_{i}^{d}\mathbb{E}\left[\frac{1}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}-\frac{1}{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}}\right]\\
=&\sum_{i}^{d}\left(\frac{1}{\sqrt{\beta_2\boldsymbol{v}_{0,i}+\zeta}}+\sum_{t=0}^{T-2}\mathbb{E}\left[\frac{1}{\sqrt{\beta_2\boldsymbol{v}_{t+1,i}+\zeta}}-\frac{1}{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}}\right]-\mathbb{E}\left[\frac{1}{\sqrt{\boldsymbol{v}_{T,i}+\zeta}}\right]\right)\\
\le&\sum_{i}^{d}\left(\frac{1}{\sqrt{\zeta}}+\sum_{t=0}^{T-2}\mathbb{E}\left[\frac{1}{\sqrt{\beta_2\boldsymbol{v}_{t+1,i}+\zeta}}-\frac{1}{\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}}\right]\right)\\
=&\sum_{i}^{d}\left(\frac{1}{\sqrt{\zeta}}+\sum_{t=0}^{T-2}\mathbb{E}\left[\frac{(1-\beta_2)\,\boldsymbol{v}_{t+1,i}}{\sqrt{\beta_2\boldsymbol{v}_{t+1,i}+\zeta}\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}\left(\sqrt{\beta_2\boldsymbol{v}_{t+1,i}+\zeta}+\sqrt{\boldsymbol{v}_{t+1,i}+\zeta}\right)}\right]\right)\\
\le&\sum_{i}^{d}\left(\frac{1}{\sqrt{\zeta}}+\frac{1-\beta_2}{\sqrt{\zeta}}T\right).
\end{aligned}\tag{81}$$

The second one is:

$$\begin{aligned}
\sum_{t=0}^{T-1}\frac{(1-\beta_2)\,\boldsymbol{m}_{t,i}^2}{\boldsymbol{v}_{t,i}+\zeta}=\sum_{t=0}^{T-1}\frac{\frac{(1-\beta_2)\boldsymbol{m}_{t,i}^2}{\boldsymbol{v}_{t,i}-(1-\beta_2)\boldsymbol{m}_{t,i}^2}}{1+\frac{(1-\beta_2)\boldsymbol{m}_{t,i}^2}{\boldsymbol{v}_{t,i}-(1-\beta_2)\boldsymbol{m}_{t,i}^2}}&\le\sum_{t=0}^{T-1}\ln\left(1+\frac{(1-\beta_2)\,\boldsymbol{m}_{t,i}^2}{\boldsymbol{v}_{t,i}-(1-\beta_2)\,\boldsymbol{m}_{t,i}^2}\right)\\
=\sum_{t=0}^{T-1}\ln\left(\frac{\boldsymbol{v}_{t,i}}{\beta_2\boldsymbol{v}_{t-1,i}}\right)=\ln\left(\frac{\boldsymbol{v}_{T,i}}{\boldsymbol{v}_{0,i}}\right)-T\ln\beta_2\,,&
\end{aligned}\tag{82}$$

where we utilize $\ln(1+a)\le a$.

And the last one is:

$$\begin{aligned}
&\sum_{t=0}^{T-1}\sum_{i}^{d}\mathbb{E}\left[\left|\hat{\nabla}_i f(\boldsymbol{\theta}_t;\xi_{t+1})-\nabla_i F_\mu(\boldsymbol{\theta}_t)\right|\right]\overset{(a)}{\le}\sqrt{d}\sum_{t=0}^{T-1}\mathbb{E}\left[\left\|\hat{\nabla} f(\boldsymbol{\theta}_t;\xi_{t+1})-\nabla F_\mu(\boldsymbol{\theta}_t)\right\|\right]\\
\overset{(b)}{\le}&\sqrt{d}\sum_{t=0}^{T-1}\sqrt{\mathbb{E}\left[\left\|\hat{\nabla} f(\boldsymbol{\theta}_t;\xi_{t+1})-\nabla F_\mu(\boldsymbol{\theta}_t)\right\|^2\right]}\overset{(c)}{\le}T\sqrt{d\left(V+\frac{L_0^2\eta^2d\,(N^2-1)}{3\,(1-\beta_2)\,N^2K\mu^2}+\frac{L_1^2\eta^2d^2\,(N-1)}{2\,(1-\beta_2)}\right)},
\end{aligned}\tag{83}$$

where $(a)$ comes from Cauchy-Schwarz inequality, $(b)$ is due to Jensen's inequality, and $(c)$ comes from Thm. B.5.

Reorganizing (80), we can derive:

$$\begin{aligned}
\frac{1}{T}\sum_{t=0}^{T-1}\sum_{i}^{d}\frac{|\nabla_i F_\mu(\boldsymbol{\theta}_t)|^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}\le&\frac{8C^2d^3}{(1-\beta_1)^2\mu^2\eta T}\left(\frac{1}{\sqrt{\zeta}}+\frac{T\,(1-\beta_2)}{\sqrt{\zeta}}\right)+\frac{\beta_1^2\,(1-\beta_2)}{(1-\beta_1)^2}d\sqrt{\zeta}\\
&+\frac{2\,(1-\kappa)}{(1-\beta_1)\,\eta T}\Delta+\frac{8L_1\sqrt{d}}{(1-\beta_1)\,(1-\kappa)\,T}\eta\sum_{i}^{d}\left(\frac{1}{1-\beta_2}\ln\left(\frac{\beta_2^T\boldsymbol{v}_{0,i}+4C^2d^2/\mu^2}{\boldsymbol{v}_{0,i}}\right)+2T\right)+G\,.
\end{aligned}\tag{84}$$

To simplify the equation, we choose $1-\beta_2=\min\left(\frac{(1-\beta_1)^2\mu^2\eta\sqrt{\zeta}\epsilon^2}{64C^2d^3},\frac{(1-\beta_1)^2\epsilon^2}{4\beta_1^2d\sqrt{\zeta}}\right)\sim\mathcal{O}\left(\epsilon^2\right)$, $T=\max\left(\frac{64C^2d^3}{(1-\beta_1)^2\mu^2\eta\sqrt{\zeta}\eta\epsilon^2},\frac{8(1-\kappa)}{(1-\beta_1)\eta\epsilon^2},\frac{64L_1\sqrt{d}\eta}{(1-\beta_1)(1-\kappa)(1-\beta_2)\epsilon^2}\sum_i^d\ln\left(\frac{\beta_2^T\boldsymbol{v}_{0,i}+4C^2d^2/\mu^2}{\boldsymbol{v}_{0,i}}\right)\right)\sim\mathcal{O}\left(\epsilon^{-4}\right)$, $\eta=\frac{(1-\beta_1)(1-\kappa)\epsilon^2}{128L_1d^{3/2}}\sim\mathcal{O}\left(\epsilon^2\right)$, and then have:

$$\frac{1}{T}\sum_{t=0}^{T-1}\frac{\|\nabla F_\mu(\boldsymbol{\theta}_t)\|^2}{\sqrt{\beta_2\|\boldsymbol{v}_t\|+\zeta}}\le\frac{1}{T}\sum_{t=0}^{T-1}\sum_{i}^{d}\frac{|\nabla_i F_\mu(\boldsymbol{\theta}_t)|^2}{\sqrt{\beta_2\boldsymbol{v}_{t,i}+\zeta}}\le\frac{1}{4}\epsilon^2+\frac{1}{4}\epsilon^2+\frac{1}{4}\epsilon^2+\frac{1}{4}\epsilon^2+G\le\epsilon^2+G\,.\tag{85}$$

Overall, inserting the results of term I (55) and term II (85) into (50):

$$\left(\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_t)\|\right]\right)^2 \leq \left(\sqrt{\zeta}+\Xi+\sqrt{\frac{2}{\beta_1\left(1-\beta_2\right)}}\left(1+\beta_1\right)\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_{t-1})\|\right]\right)\left(\epsilon^2+G\right)\ , \tag{86}$$

which is actually a quadratic inequality. After solving the root of the quadratic equation, we obtain:

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_t)\|\right] \leq & \sqrt{\frac{2}{\beta_1\left(1-\beta_2\right)}}\left(1+\beta_1\right)\left(\epsilon^2+G\right)+\left(\sqrt[4]{\zeta}+\sqrt{\Xi}\right)\sqrt{\epsilon^2+G} \\ \leq & \sqrt{\frac{2}{\beta_1\left(1-\beta_2\right)}}\left(1+\beta_1\right)\left(\epsilon^2+G\right)+\left(\sqrt[4]{\zeta}+\sqrt{\Xi}\right)\left(\epsilon+\sqrt{G}\right) \\ \leq & \sqrt{\frac{2}{\beta_1\left(1-\beta_2\right)}}\left(1+\beta_1\right)\epsilon^2+\left(\sqrt[4]{\zeta}+\sqrt{\Xi}\right)\epsilon+B_2\ ,\end{aligned} \tag{87}$$

where $B_2 \triangleq \sqrt{\frac{2}{\beta_1(1-\beta_2)}}\left(1+\beta_1\right)G+\left(\sqrt[4]{\zeta}+\sqrt{\Xi}\right)\sqrt{G}$.

To derive the convergence guarantee of $F(\boldsymbol{\theta}_t)$, we introduce the bias between $F_\mu(\boldsymbol{\theta}_t)$ and $F(\boldsymbol{\theta}_t)$, which is defined as:

$$\begin{aligned}\mathbb{E}\left[\|\nabla F(\boldsymbol{\theta})-\nabla F_\mu(\boldsymbol{\theta})\|\right] \overset{(a)}{=} & \mathbb{E}\left[\|\mathbb{E}_{\mathbf{u}}\left[\nabla F(\boldsymbol{\theta})-\nabla F(\boldsymbol{\theta}+\mu\mathbf{u})\right]\|\right] \overset{(b)}{\leq} \mathbb{E}\left[\|\nabla F(\boldsymbol{\theta})-\nabla F(\boldsymbol{\theta}+\mu\mathbf{u})\|\right] \\ \overset{(c)}{\leq} & \sqrt{d}L_1\mathbb{E}\left[\|\mu\mathbf{u}\|\right] \overset{(d)}{=} \mu L_1\sqrt{d}\ ,\end{aligned} \tag{88}$$

where $(a)$ comes from the definition of $F_\mu$ (3), $(b)$ results from Jensen's inequility, $(c)$ is due to Assump. B.1, and $(d)$ follows from the fact that $\mathbf{u}\sim\mathrm{Unif}(\mathbb{S}^{d-1})$ and hence $\|\mathbf{u}\|=1$.

Afterthat, we can bound the convergence of $F(\boldsymbol{\theta}_t)$ as below:

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla F(\boldsymbol{\theta}_t)\|\right] \leq & \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_t)\|\right]+\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\nabla F_\mu(\boldsymbol{\theta}_t)-\nabla F(\boldsymbol{\theta}_t)\|\right] \\ \leq & \sqrt{\frac{2}{\beta_1\left(1-\beta_2\right)}}\left(1+\beta_1\right)\epsilon^2+\left(\sqrt[4]{\zeta}+\sqrt{\Xi}\right)\epsilon+\mu L_1\sqrt{d}+B_2\ ,\end{aligned} \tag{89}$$

which completes the proof. □

## D. Experiments Setup

In this section, we first introduce the baselines used in our experiments (Sec. D.1), and then we provide experimental details on synthetic functions (Sec. D.2), black-box adversarial attack (Sec. D.3), and memory-efficient LLM fine-tuning (Sec. D.4).

### D.1. Baselines

First of all, we claims that our experiments compare only the differing gradient estimation methods among all baselines and ZoAR. Consequently, all baselines and ZoAR share the same update rule, such as ZO-AdaMM and $\mathcal{R}$-AdaZO. Below, we introduce the three baselines used in our study.

- **Vanilla ZOO**. This zeroth-order optimization algorithm employs the gradient estimator in (2). When paired with the Adam update rule, it is denoted ZO-AdaMM (Chen et al., 2019); when paired with the $\mathcal{R}$-AdaZO update rule, it is referred to as $\mathcal{R}$-AdaZO (Shu et al., 2025b).
- **ReLIZO** (Wang et al., 2024). ReLIZO is zeroth-order gradient estimation algorithm, which reuses queries from previous iterations through a quadratically constrained linear program, and effectively decouples sample size from variable dimension.

- **ZOO with HiStorical gradient (ZoHS)**. On the basis of the Vanilla ZOO framework, ZoHS integrates historical gradient information into the gradient estimation procedure. Specifically, the gradeint estimator for ZoHS is formally defined as:

$$\hat{\nabla}_{\text{ZoHS}} F(\boldsymbol{\theta}_{t-1}) \triangleq \frac{1}{N} \sum_{n=1}^{N} \hat{\nabla} F(\boldsymbol{\theta}_{t-n}) \,, \tag{90}$$

  where $\hat{\nabla} F(\boldsymbol{\theta}_{t-n})$ is the gradient estimator of Vanilla ZOO at iteration $t-n$.

### D.2. Synthetic Functions

All experiments are conducted in $d = 10000$ dimensions and run for $T = 20000$ iterations. For a fair comparison, all experiments share the same initialization and hyperparameters: the step size $\eta = 0.001$, the number of queries $K = 10$, the smoothing radius parameter $\mu = 0.05$, and the number of histories $N = 6$. The analytical forms of the synthetic functions used in our experiments are as follows:

- **Ackley Function:**

$$f(\boldsymbol{\theta}) = -20 \exp\left(-0.2 \sqrt{\frac{1}{d} \sum_{i=1}^{D} \theta_i^2}\right) - \exp\left(\frac{1}{d} \sum_{i=1}^{d} \cos(2\pi\theta_i)\right) + 20 + e \,. \tag{91}$$

- **Levy Function:**

$$f(\boldsymbol{\theta}) = \sin^2(\pi w_1) + \sum_{i=1}^{d-1} (w_i - 1)^2 \left[1 + 10 \sin^2(\pi w_i + 1)\right] + (w_d - 1)^2 \left[1 + \sin^2(2\pi w_d)\right] \,, \tag{92}$$

  where $w_i = 1 + \frac{\theta_i - 1}{4}$.

- **Quadratic Function:**

$$f(\boldsymbol{\theta}) = \frac{1}{2} \sum_{i=1}^{d} \theta_i^2 \,. \tag{93}$$

- **Rosenbrock Function:**

$$f(\boldsymbol{\theta}) = \sum_{i=1}^{d-1} \left[100(\theta_{i+1} - \theta_i^2)^2 + (1 - \theta_i)^2\right] \,. \tag{94}$$

Note that all four functions have the same optimal solution of zero.

### D.3. Black-box Adversarial Attack

For the black-box adversarial attack, we use the same model as in (Wang et al., 2024): a simple two-layer CNN trained on the MNIST dataset. To ensure a fair comparison, all experiments utilize the same initialization and the following hyperparameters: step size $\eta = 0.01$, number of queries $K = 2$, smoothing parameter $\mu = 0.5$, and number of histories $N = 6$.

### D.4. Memory-Efficient LLM Fine-Tuning

For the memory-efficient fine-tuning of large language models, we select OPT-1.3B and OPT-13B (Zhang et al., 2022) as the pretrained models, and fine-tune them with LoRA adapters on the SST-2 and COPA datasets from the GLUE benchmark (Wang et al., 2019). All experiments are conducted using the same initialization and hyperparameters: step size $\eta = 0.00005$, number of queries $K = 2$, smoothing parameter $\mu = 0.01$, and history lengths $N = \{15, 50\}$. The batch size is fixed at 16 for both datasets.

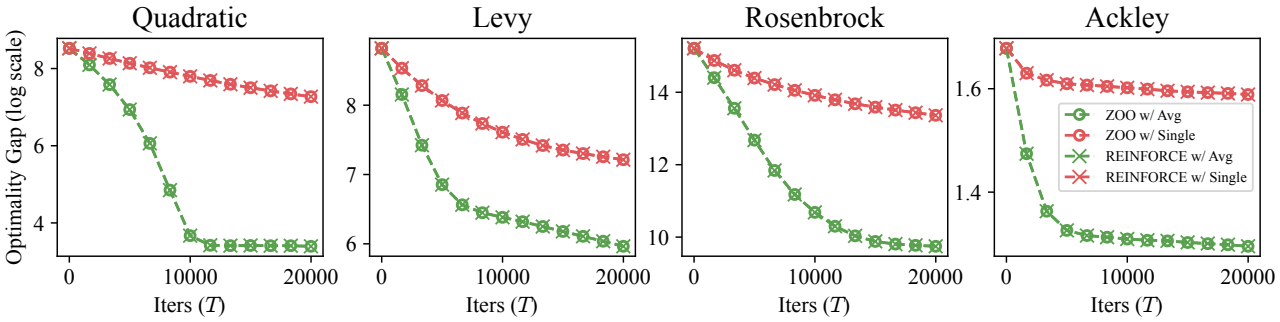


*Figure 2.* Equivalence of ZOO and REINFORCE with two different baselines. The $y$-axis denotes the gap between the current function value and the optimal function value. The green curves denote to the average baseline defined in (12), while the red curves denote to the single-point baseline $b_t = f(\boldsymbol{\theta}_{t-1}; \xi)$. All curves are averaged over 5 independent runs.

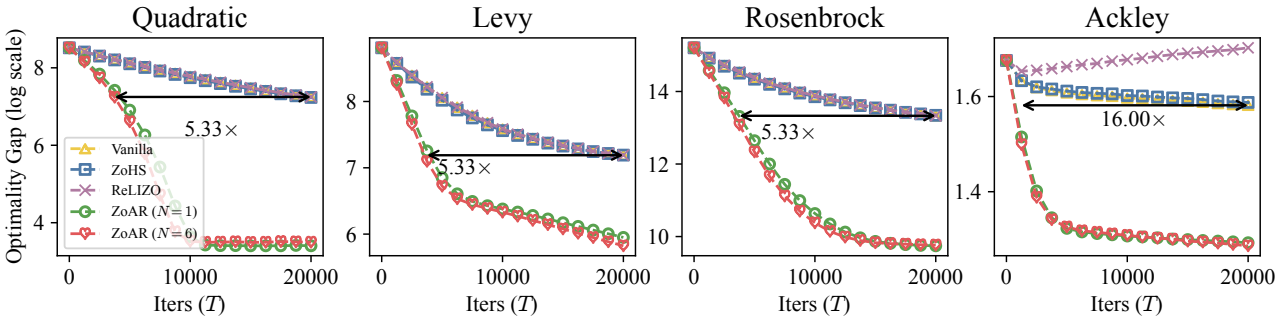


*Figure 3.* Comparison of convergence among different zeroth-order optimization algorithms on four synthetic functions under $\mathcal{R}$-AdaZO setting. The $x$-axis denotes the number of iterations, and the $y$-axis denotes the gap between the current function value and the optimal function value, i.e. $F(\boldsymbol{\theta}) - \min_{\boldsymbol{\theta}'} F(\boldsymbol{\theta}')$. All curves are averaged over 5 independent runs.

## E. Additional Experiments

### E.1. The Equivalence between ZOO and REINFORCE

To empirically validate our core theoretical finding that the Gaussian-smoothed ZOO shares the same convergence as the single-step REINFORCE with baseline (Cor. 3.4), we conduct comparison on four synthetic functions. Fig. 2 illustrates these comparisons using two baselines: the standard ZOO single-point baseline ($b_t = f(\boldsymbol{\theta}_{t-1}; \xi)$, red curves) and an averaged baseline (green curves) proposed for ZoAR in (12). The results in Fig. 2 clearly deliver two key points. First, for any given baseline strategy (either single-point or averaged), the convergence trajectories of ZOO and REINFORCE are virtually indistinguishable across all four synthetic functions. This provides strong numerical evidence supporting our theoretical equivalence. Second, the averaged baseline (green curves) consistently and significantly outperforms the single-point baseline (red curves) for both ZOO and REINFORCE. This manifests as faster convergence and a lower final optimality gap, underscoring the effectiveness of the PO-inspired averaged baseline in reducing variance and improving optimization performance, a central premise of our ZoAR.

### E.2. Synthetic Functions Optimization under $\mathcal{R}$-AdaZO

Consistent with the experiments in Section 5.1, we further conducted evaluations on four synthetic functions—Ackley, Levy, Quadratic, and Rosenbrock—utilizing the $\mathcal{R}$-AdaZO update rule. The results are presented in Figure 3. Notably, the performance of Vanilla ZOO and ZoHS is highly similar, which indicates that ZoHS does not confer any additional advantage within the $\mathcal{R}$-AdaZO setting. Furthermore, the performance of ZoAR w/ and w/o historical information is closely comparable, suggesting that ZoAR w/o history is sufficiently effective for practical application under the $\mathcal{R}$-AdaZO framework.

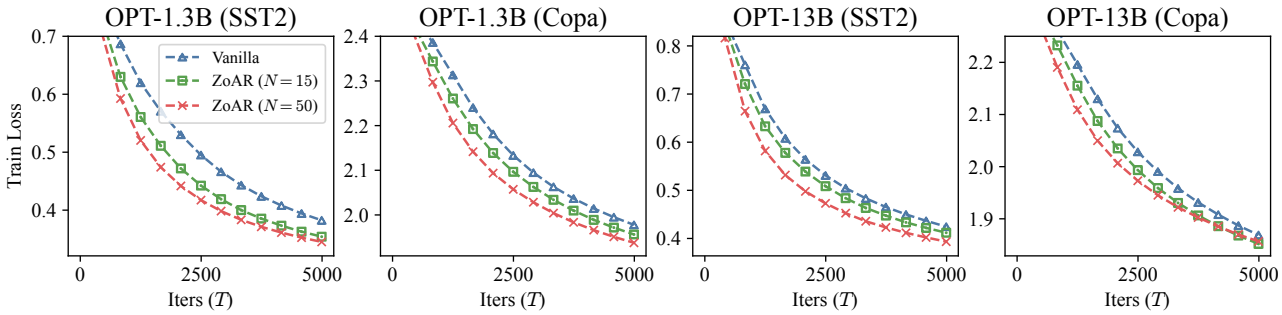


*Figure 4.* Training loss comparison between Vanilla ZOO and ZoAR for the LLM fine-tuning under different model sizes on SST2 and Copa datasets. Each curve is averaged over 3 independent runs.

### E.3. Memory-Efficient LLM Fine-Tuning

The pursuit of memory-efficient fine-tuning for large language models (LLMs) has recently incorporated zeroth-order optimization techniques ((Malladi et al., 2023)). However, conventional zeroth-order optimization methods typically exhibit increased variance in gradient estimation, which can adversely affect the convergence of LLM fine-tuning. To mitigate this variance, ZoAR reuses historical information without incurring the additional cost of new queries. In this section, we fine-tunes the OPT-1.3B and OPT-13B models on the SST2 and COPA datasets, respectively, employing the $\mathcal{R}$-AdaZO update rule (refer to D.4 for more details). ZoAR is compared against the vanilla Zeroth-Order Optimization (ZOO) method, which served as the baseline. The results, presented in Figure 4, demonstrate that ZoAR outperforms the vanilla ZOO method, particularly for the smaller OPT-1.3B model. Furthermore, the convergence rate of ZoAR incorporating historical information surpasses that of the variant without historical information, suggesting the beneficial role of historical data in LLM fine-tuning.

## F. Limitations and Broader Impact

ZoAR is an excellent variance reduction ZOO method, which can not only reduce the memory cost, but also increase the convergence rate. Therefore, ZoAR is suitable for many variance dominate tasks, especially for LLM fine-tuning. Besides, ZoAR works well even with a large smoothing parameter $\mu = 0.01$ or $\mu = 0.1$, which is much larger than the commonly used $\mu = 0.001$ in Vanilla ZOO. This is because ZoAR reuses the queries to smooth the gradient estimation, which is equivalent to using a smaller smoothing parameter $\mu$. This suggests that ZoAR is suitable to some non-smoothness objective function, such as quantized function in quantization aware training (QAT) field. Recent study (Zhou et al., 2025) have combined the ZOO with QAT to avoid the inaccuracy occured by straight through estimator (STE). However, quantized function is actually a multiple step function, where small smoothing parameter $\mu$ would not sufficiently change the quantized function value, especially for ultra-low precision (such as FP4), and often leads to worse convergence. ZoAR would be a good choice for ultra-low precision QAT, since it can use a large smoothing parameter $\mu$ to smooth the quantized function, which is left for future work.

Besides, despite its effectiveness, ZoAR presents several limitations. First, similarly with some variance reduction techniques, such as (Shu et al., 2025b), ZoAR reuse historical queries, which introduce additional bias (Thm. B.5), potentially leading to inaccurate descent directions. However, the extra bias is proptotional to the length of the historical gradient, and hence we can introduce linear schudule to dynamically adjust the history length, aiming to reduce or even eliminate the bias. This can be left for future work. Moreover, ZoAR retrive the historical samples from random seed storage, which may cost extra computation when history length is large. This can be solved by utilizing parallel computing techniques or employing dynamic scheduling of history length to improve computational efficiency.