Improving the Convergence Rate of One-Point Zeroth–Order Optimization using Residual Feedback

Yan Zhang*
Dept. of Mechanical Eng. and Material Science
Duke University
Durham, NC 27705
yan.zhang2@duke.edu

Yi Zhou*
Dept. of Electrical & Computer Eng.
The University of Utah
Salt Lake City, UT 84112
yi.zhou@utah.edu

Kaiyi Ji
Dept. of Electrical & Computer Eng.
The Ohio State University
Columbus, OH 43210
ji.367@osu.edu

Michael M. Zavlanos
Dept. of Mechanical Eng. and Material Science
Duke University
Durham, NC 27705
michael.zavlanos@duke.edu

Abstract

Many existing zeroth-order optimization (ZO) algorithms adopt two-point feedback schemes due to their fast convergence rate compared to one-point feedback schemes. However, two-point schemes require two evaluations of the objective function at each iteration, which can be impractical in applications where the data are not all available a priori, e.g., in online optimization. In this paper, we propose a novel one-point feedback scheme that queries the function value only once at each iteration and estimates the gradient using the residual between two consecutive feedback points. When optimizing a deterministic Lipschitz function, we show that the query complexity of ZO with the proposed one-point residual feedback matches that of ZO with the existing two-point feedback schemes. Moreover, the query complexity of the proposed algorithm can be improved when the objective function has Lipschitz gradient. Then, for stochastic bandit optimization problems, we show that ZO with one-point residual feedback achieves the same convergence rate as that of ZO with two-point feedback with uncontrollable data samples. We demonstrate the effectiveness of the proposed one-point residual feedback via extensive numerical experiments.

1 Introduction

Zeroth-order optimization algorithms have been widely-used to solve machine learning problems where first or second order information (i.e., gradient or Hessian information) is unavailable, e.g., black-box optimization Nesterov & Spokoiny (2017), adversarial training Chen et al. (2017), reinforcement learning Fazel et al. (2018); Malik et al. (2018) and human-in-the-loop control Kim et al. (2017). In these problems, the goal is to solve the following generic optimization problem

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

where $x \in \mathbb{R}^d$ corresponds to the parameters and f denotes the total loss. Using zeroth-order information, i.e., function evaluations, first-order gradients can be estimated to solve the problem (P).

*Equal contribution

Preprint. Under review.

Table 1: Iteration Complexity of Zeroth-order Methods with One-point, Two-point and Proposed Feedback Schemes

Complexity [2]		Convex $C^{0,0}$	Convex $C^{1,1}$	Nonconvex $C^{0,0}$	Nonconvex $C^{1,1}$
One-point	Gasnikov et al. (2017)	$d^2\epsilon^{-4}$	$d\epsilon^{-3}$	–	–
Two-point	Duchi et al. (2015)	$d\log(d)\epsilon^{-2}$	$d\epsilon^{-2}$	–	–
	Shamir (2017)	$d\epsilon^{-2}$	–	–	–
	Nesterov & Spokoiny (2017)	$d^2\epsilon^{-2}$	$d\epsilon^{-1}$	$d^3\epsilon_f^{-1}\epsilon^{-2}$	$d\epsilon^{-1}$
	Bach & Perchet (2016)	–	$d^2\epsilon^{-3}$ (UN)	–	–
Residual One-point	Deterministic	$d^2\epsilon^{-2}$	$d^3\epsilon^{-1.5}$	$d^4\epsilon_f^{-1}\epsilon^{-2}$	$d^3\epsilon^{-1.5}$
	Stochastic	$d^2\epsilon^{-4}$	$d^2\epsilon^{-3}$	$d^3\epsilon_f^{-3}\epsilon^{-2}$	$d^4\epsilon^{-3}$

Existing zeroth-order optimization (ZO) algorithms can be divided into two categories, namely, ZO with one-point feedback and ZO with two-point feedback. Flaxman et al. (2005) was among the first to propose a ZO algorithm with one-point feedback, i.e., a zeroth-order gradient estimator that queries one function value in each iteration. The corresponding one-point gradient estimator $\widetilde{\nabla} f(x)$ takes the form [2]

$$\text{(One-point feedback):}\ \ \widetilde{\nabla} f(x) = \frac{u}{\delta} f(x + \delta u), \tag{1}$$

where δ is an exploration parameter and $u \in \mathbb{R}^d$ is sampled from the standard normal distribution element-wise. In particular, Flaxman et al. (2005) showed that the above one-point feedback has a large estimation variance and the resulting ZO algorithm achieves a convergence rate of at most $\mathcal{O}(T^{-\frac{1}{4}})$, which is much slower than that of gradient descent algorithms used to solve problem (P). Assuming smoothness and relying on self-concordant regularization, Saha & Tewari (2011); Dekel et al. (2015) further improved this convergence speed. However, the gap in the iteration complexity between one-point feedback and gradient-based methods remained. In order to reduce the estimation variance in the above one-point feedback, Agarwal et al. (2010); Nesterov & Spokoiny (2017); Shamir (2017) introduced the following two-point feedback schemes

$$\text{(Two-point feedback):}\ \ \widetilde{\nabla} f(x) = \frac{u}{\delta} f(x + \delta u) - f(x),\ \text{or}\ \frac{u}{2\delta}\big(f(x + \delta u) - f(x - \delta u)\big), \tag{2}$$

and showed that ZO with these two-point feedbacks that have smaller estimation variance achieves a convergence rate of $\mathcal{O}(\frac{1}{\sqrt{T}})$ (or $\mathcal{O}(\frac{1}{T})$ when the problem is smooth), which is order-wise much faster than the convergence rate achieved by ZO algorithms with one-point feedback. Therefore, as pointed out in Larson et al. (2019), a natural and fundamental question to ask is the following:

- *(Q1): Does there exist a one-point feedback for which zeroth-order optimization can achieve the same query complexity as that of two-point feedback methods?*

The literature discussed above focuses on deterministic optimization problems (P). Nevertheless, in practice, many problems involve randomness in the environment and parameters, giving rise to the following stochastic optimization problem

$$\min_{x \in \mathbb{R}^d} f(x) = \mathbb{E}_\xi[F(x, \xi)], \tag{Q}$$

where only a noisy function evaluation $F(x, \xi)$ with a random data sample ξ is available. Zeroth-order algorithms have also been developed to solve the above problem (Q), e.g., Ghadimi & Lan (2013); Duchi et al. (2015); Hu et al. (2016); Bach & Perchet (2016); Gasnikov et al. (2017). In particular, Ghadimi & Lan (2013) consider the following widely-used stochastic two-point feedback

$$\widetilde{\nabla} f(x) = \frac{u}{\delta}\big(F(x + \delta u, \xi) - F(x, \xi)\big) \tag{3}$$

[2] In Flaxman et al. (2005), the estimator is $\widetilde{\nabla} f(x) = \frac{du}{\delta} f(x + \delta u)$ where u is uniformly sampled from a unit sphere. In this paper, we follow Nesterov & Spokoiny (2017) and sample u from the standard normal distribution.

[2] In convex setting, the accuracy is meaured by $f(x) - f(x^*) \le \epsilon$, while in the non-convex setting, it is measured by $\|\nabla f(x)\|^2 \le \epsilon$ when the objective function is smooth. When the objective function is non-smooth, we enforce two optimality measures, $|f(x) - f_\delta(x)| \le \epsilon_f$ and $\|\nabla f_\delta(x)\|^2 \le \epsilon$ together. (UN) means the oracle considers uncontrollable data samples.

and show that ZO with this stochastic two-point feedback has the same convergence rate as ZO with the two-point feedback scheme in (2) for deterministic problems (P). Similarly, Duchi et al. (2015) further analyzed the above oracle in a mirror descent framework and showed a similar convergence speed. Stochastic one-point and two-point feedback schemes with improved convergence rates have also been studied in Gasnikov et al. (2017). However, these stochastic two-point feedback schemes assume that the data sample ξ is controllable, i.e., one can fix the data sample ξ and evaluate the function value at two distinct points x and $x + \delta u$. This assumption is unrealistic in many applications. For example, in reinforcement learning, controlling the sample ξ requires applying the same sequence of noises to the dynamical system and reward function. Hence, two-point feedback schemes with fixed data sample can be impractical. To address this challenge, Hu et al. (2016); Bach & Perchet (2016) proposed a more practical noisy two-point feedback method that replaces the fixed sample ξ in eq. (3) with two independent samples ξ, ξ'. Its convergence rate was shown to match that of the stochastic one-point feedback $\widetilde{\nabla} f(x) = \frac{u}{\delta} F(x + \delta u, \xi)$. Still though, this two-point feedback method with independent data samples produces gradient estimates with lower variance. Therefore, an additional fundamental question we seek to address in this paper is the following:

- *(Q2): Can we develop a stochastic one-point feedback that achieves the same practical performance as that of the noisy two-point feedback?*

Contributions: In this paper, we provide positive answers to these open questions by introducing a new one-point residual feedback scheme and theoretically analyzing the convergence of zeroth-order optimization using this feedback scheme. Specifically, our contributions are as follows. We propose a new one-point feedback scheme which requires a single function evaluation at each iteration. This feedback scheme estimates the gradient using the residual between two consecutive feedback points and we refer to it as residual feedback. We show that our residual feedback induces a smaller estimation variance than the one-point feedback (1) considered in Flaxman et al. (2005); Gasnikov et al. (2017). Specifically, in deterministic optimization where the objective function is Lipschitz-continuous, we show that ZO with our residual feedback achieves the same convergence rate as existing ZO with two-point feedback schemes. To the best of our knowledge, this is the first one-point feedback scheme with provably comparable performance to two-point feedback schemes in ZO. Moreover, when the objective function has an additional smoothness structure, we further establish an improved convergence rate of ZO with residual feedback. In the stochastic case where only noisy function values are available, we show that the convergence rate of ZO with residual feedback matches the state-of-the-art result of ZO with two-point feedback under uncontrollable data samples. Hence, our residual feedback bridges the theoretical gap between ZO with one-point feedback and ZO with two-point feedback. A summary of the complexity results for the proposed residual-feedback scheme can be found in Table 1.

Related work: Zeroth-order methods have been extended to solve more challenging optimization problems compared to (P) or (Q). For example, Balasubramanian & Ghadimi (2018) apply ZO to solve a set-constrained optimization problem where the projection onto the constraint set is non-trivial. Gorbunov et al. (2018); Ji et al. (2019) apply a variance-reduced technique and acceleration schemes to achieve better convergence speed in ZO. Wang et al. (2018) improves the dependence of the iteration complexity on the dimension of the problem under an additional sparsity assumption on the gradient of the objective function. Finally, Hajinezhad & Zavlanos (2018); Tang & Li (2019) applied zeroth-order oracles to distributed optimization problems when only bandit feedbacks are available at each local agents. Our proposed residual feedback oracle can be extended as in the above methods to deal with more challenging optimization problems.

2 Preliminaries

In this section, we present definitions and preliminary results needed throughout our analysis. Following Nesterov & Spokoiny (2017); Bach & Perchet (2016), we introduce the following classes of Lipschitz and smooth functions.

Definition 2.1 (Lipschitz functions). *The class of Lipschtiz-continuous functions $C^{0,0}$ satisfy: for any $f \in C^{0,0}$, $|f(x) - f(y)| \leq L_0 \|x - y\|$, $\forall x, y \in \mathbb{R}^d$, for some Lipschitz parameter $L_0 > 0$. The class of smooth functions $C^{1,1}$ satisfy: for any $f \in C^{1,1}$, $\|\nabla f(x) - \nabla f(y)\| \leq L_1 \|x - y\|$, $\forall x, y \in \mathbb{R}^d$, for some Lipschitz parameter $L_1 > 0$.*

In ZO, the objective is to estimate the first-order gradient of a function using zeroth-order oracles. Necessarily, we need to perturb the function around the current point along all the directions uniformly in order to estimate the gradient. This motivates us to consider the Gaussian-smoothed version of the function f as introduced in Nesterov & Spokoiny (2017), $f_\delta(x) := \mathbb{E}_{u\sim\mathcal{N}(0,1)}[f(x+\delta u)]$, where the coordinates of the vector u are i.i.d standard Gaussian random variables. The following bounds on the approximation error of the function $f_\delta(x)$ have been developed in Nesterov & Spokoiny (2017).

Lemma 2.2 (Gaussian approximation)**.** *Consider a function f and its Gaussian-smoothed version f_δ. It holds that*

$$|f_\delta(x) - f(x)| \leq \begin{cases} \delta L_0 \sqrt{d}, \ \textit{if } f \in C^{0,0}, \\ \delta^2 L_1 d, \ \textit{if } f \in C^{1,1}, \end{cases} \quad \textit{and } \|\nabla f_\delta(x) - \nabla f(x)\| \leq \delta L_1 (d+3)^{3/2}, \ \textit{if } f \in C^{1,1}.$$

Moreover, the smoothed function $f_\delta(x)$ has the following nice geometrical property as proved in Nesterov & Spokoiny (2017).

Lemma 2.3. *If function $f \in C^{0,0}$ is L_0-Lipschitz, then its Gaussian-smoothed version f_δ belongs to $C^{1,1}$ with Lipschitz constant $L_1 = \sqrt{d}\delta^{-1} L_0$.*

We also introduce the following notions of convexity.

Definition 2.4 (Convexity)**.** *A continuously differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is called convex if for all $x, y \in \mathbb{R}^d$, $f(x) \geq f(y) + \langle x - y, \nabla f(y) \rangle$.*

3 Deterministic ZO with Residual Feedback

In this section, we consider the problem (P), where the objective function evaluation is fully deterministic. To solve this problem, we propose a zeroth-order estimate of the gradient based on the following one-point residual feedback scheme

$$\text{(Residual feedback):} \quad \widetilde{g}(x_t) := \frac{u_t}{\delta}\big(f(x_t + \delta u_t) - f(x_{t-1} + \delta u_{t-1})\big), \tag{4}$$

where u_{t-1} and u_t are independent random vectors sampled from the standard multivariate Gaussian distribution. To elaborate, the gradient estimate in (4) evaluates the function value at one perturbed point $x_t + \delta u_t$ at each iteration t and the other function value evaluation $f(x_{t-1} + \delta u_{t-1})$ is inherited from the previous iteration. Therefore, it is a one-point feedback scheme based on the residual between two consecutive feedback points, and we name it *one-point residual feedback*. Next, we show that this estimator is an unbiased gradient estimate of the smoothed function $f_\delta(x)$ at x_t.

Lemma 3.1. *We have that $\mathbb{E}\big[\tilde{g}(x_t)\big] = \nabla f_\delta(x_t)$ for all $x_t \in \mathbb{R}^d$.*

Proof. Taking the expectation of $\tilde{g}(x_t)$ in (4) with respect to the random vectors u_t and u_{t-1}, we obtain that

$$\mathbb{E}\big[\tilde{g}(x_t)\big] = \frac{1}{\delta}\mathbb{E}\big[f(x_t + \delta u_t)u_t - f(x_{t-1} + \delta u_{t-1})u_t\big] = \frac{1}{\delta}\mathbb{E}\big[f(x_t + \delta u_t)u_t\big] = \nabla f_\delta(x_t),$$

where the second equality follows from the fact that u_t is independent from u_{t-1} and x_{t-1} and has zero mean. The last equality is a fact proved in Nesterov & Spokoiny (2017). ☐

Since $\tilde{g}(x_t)$ is an unbiased estimate of $\nabla f_\delta(x_t)$, we can use it in Stochastic Gradient Descent (SGD) as follows

$$x_{t+1} = x_t - \eta \tilde{g}(x_t), \tag{5}$$

where η is the stepsize. To analyze the convergence of the above ZO algorithm with residual feedback, we need to bound the variance of the gradient estimate under proper choices of the exploration parameter δ in (4) and the stepsize η. In the following result, we present the bounds on the second moment of the gradient estimate $\mathbb{E}[\|\tilde{g}(x_t)\|^2]$, which will be used in our analysis later.

Lemma 3.2. *Consider a function $f \in C^{0,0}$ with Lipschitz constant L_0. Then, under the SGD update rule in* (5)*, the second moment of the residual feedback satisfies*

$$\mathbb{E}[\|\tilde{g}(x_t)\|^2] \leq \frac{2dL_0^2\eta^2}{\delta^2}\mathbb{E}[\|\tilde{g}(x_{t-1})\|^2] + 8L_0^2(d+4)^2.$$

Furthermore, if $f(x)$ also belongs to $C^{1,1}$ with constant L_1, then the second moment of the residual feedback satisfies

$$\mathbb{E}[\|\tilde{g}(x_t)\|^2] \leq \frac{2dL_0^2\eta^2}{\delta^2}\mathbb{E}[\|\tilde{g}(x_{t-1})\|^2] \ + 8(d+4)^2\|\nabla f(x_{t-1})\|^2 + 4L_1^2(d+6)^3\delta^2.$$

Lemma 3.2 shows that the second moment of the residual feedback $\mathbb{E}[\|\tilde{g}(x_t)\|^2]$ can be bounded by a perturbed contraction under the SGD update rule. This perturbation term is crucial to establish the iteration complexity of ZO with our residual feedback. In particular, with the traditional one-point feedback, the perturbation term is in the order of $O(\delta^{-2})$ and significantly degrades the convergence speed Hu et al. (2016). In comparison, our residual feedback induces a much smaller perturbation term. Specifically, when $f \in C^{0,0}$, the perturbation is the order of $O(L_0^2 d^2)$ that is independent of δ, and when $f \in C^{1,1}$, the perturbation is in the order of $O(d^2\|\nabla f(x_{t-1})\|^2 + L_1^2 d^3 \delta^2)$. Therefore, it is expected that ZO with our residual feedback can achieve a better iteration complexity than that of ZO with the traditional one-point feedback.

3.1 Convergence Analysis

We first consider the case where the objective function f is nonconvex. When f is differentiable, we say a solution x is ϵ-accurate if $\mathbb{E}[\|\nabla f(x)\|^2] \leq \epsilon$. However, when f is nonsmooth, we follow the convention adopted in Nesterov & Spokoiny (2017) and define a solution x to be ϵ-accurate if $\mathbb{E}[\|\nabla f_\delta(x)\|^2] \leq \epsilon$ holds for the Gaussian-smoothed function. In addition, we also require f_δ to be ϵ_f-close to the original f, which requires $\delta \leq \frac{\epsilon_f}{L_0\sqrt{d}}$ according to Lemma 2.2. Under this setup, the convergence rate of ZO with residual feedback is presented below. For simplicity, all the complexity results in this paper are presented in $\mathcal{O}$ notations. The explicit form of the constant terms can be found in the proof in the supplementary material.

Theorem 3.3. *Assume that $f \in C^{0,0}$ with Lipschitz constant L_0 and that f is also bounded below by f^*. Moreover, assume that SGD in* (5) *with residual feedback is run for $T > 1/\epsilon_f$ iterations and that $\tilde{x}$ is selected from the T iterates uniformly at random. Let also $\eta = \frac{\sqrt{\epsilon_f}}{2dL_0^2\sqrt{T}}$ and $\delta = \frac{\epsilon_f}{L_0 d^{\frac{1}{2}}}$. Then, we have that $\mathbb{E}\big[\|\nabla f_\delta(\tilde{x})\|^2\big] = \mathcal{O}(d^2\epsilon_f^{-0.5}T^{-0.5})$.*

Based on the above convergence rate result, the required iteration complexity to achieve a point x that satisfies $|f(x) - f_\delta(x)| \leq \epsilon_f$ as well as $\mathbb{E}[\|\nabla f(\tilde{x})\|^2] \leq \epsilon$ is of the order $\mathcal{O}(\frac{d^4}{\epsilon_f\epsilon^2})$. This complexity result is close to the complexity result $\mathcal{O}(\frac{d^3}{\epsilon_f\epsilon^2})$ of ZO with two-point feedback in Nesterov & Spokoiny (2017). When $f(x) \in C^{1,1}$ is a smooth function, we obtain the following convergence rate result for ZO with residual feedback.

Theorem 3.4. *Assume that $f(x) \in C^{0,0}$ with Lipschitz constant L_0 and that $f(x) \in C^{1,1}$ with Lipschitz constant L_1. Moreover, assume that SGD in* (5) *with residual feedback is run for T iterations and that $\tilde{x}$ is selected from the T iterates uniformly at random. Let also $\eta = \frac{1}{\widetilde{L}(d+4)^2T^{\frac{1}{3}}}$, and $\delta = \frac{1}{\sqrt{d}T^{\frac{1}{3}}}$, where $\widetilde{L} = \max(2L_0, 32L_1)$. Then, we have that $\mathbb{E}\big[\|\nabla f(\tilde{x})\|^2\big] = \mathcal{O}(d^2T^{-\frac{2}{3}})$.*

In particular, to achieve a point x that satisfies $\mathbb{E}\big[\|\nabla f(\tilde{x})\|^2\big] \leq \epsilon$, the required iteration complexity is of the order $\mathcal{O}(d^3\epsilon^{-\frac{3}{2}})$. To the best of our knowledge, the best complexity result for ZO with two-point feedback is of the order $\mathcal{O}(d\epsilon^{-1})$, which is established in Nesterov & Spokoiny (2017). Next, we consider the case where the objective function f is convex. In this case, the optimality of a solution x is measured via the loss gap $f(x) - f(x^*)$, where x^* is the global optimum of f.

Theorem 3.5. *Assume that $f(x) \in C^{0,0}$ is convex with Lipschitz constant L_0. Moreover, assume that SGD in* (5) *with residual feedback is run for T iterations and define the running average $\bar{x} = \frac{1}{T}\sum_{t=0}^{T-1} x_t$. Let also $\eta = \frac{1}{2dL_0\sqrt{T}}$ and $\delta = \frac{1}{\sqrt{T}}$. Then, we have that $f(\bar{x}) - f(x^*) = \mathcal{O}(dT^{-0.5})$.*

Moreover, assume that additionally $f(x) \in C^{1,1}$ with Lipschitz constant L_1, and let $\eta = \frac{1}{2\tilde{L}(d+4)^2T^{\frac{1}{3}}}$ and $\delta = \frac{\sqrt{d}}{T^{\frac{1}{3}}}$, where $\tilde{L} = \max\{L_0, 16L_1\}$. Then, we have that $f(\bar{x}) - f(x^*) = \mathcal{O}(d^2T^{-\frac{2}{3}})$.

To elaborate, to achieve a solution x that satisfies $f(\bar{x}) - f(x^*) \leq \epsilon$, the required iteration complexity is of the order $\mathcal{O}(d^2\epsilon^{-2})$ when $f \in C^{0,0}$. Such a complexity result significantly improves the complexity $\mathcal{O}(d^2\epsilon^{-4})$ of ZO with the traditional one-point feedback and is slightly worse than the best complexity $\mathcal{O}(d\epsilon^{-2})$ of ZO with two-point feedback. On the other hand, when $f(x) \in C^{1,1}$, the required iteration complexity of ZO with residual feedback further reduces to $\mathcal{O}(d^3\epsilon^{-1.5})$, which is better than the complexity $\mathcal{O}(d\epsilon^{-3})$ of ZO with the traditional one-point feedback whenever $\epsilon < d^{-4/3}$.

4 Online ZO with Stochastic Residual Feedback

In this section, we study the Problem (Q) where the objective function takes the form $f(x) := \mathbb{E}[F(x,\xi)]$ and only noisy samples of the function value $F(x,\xi)$ are available. Specifically, we propose the following stochastic residual feedback

$$\tilde{g}(x_t) := \frac{u_t}{\delta}\big(F(x_t+\delta u_t,\xi_t) - F(x_{t-1}+\delta u_{t-1},\xi_{t-1})\big), \tag{6}$$

where ξ_{t-1} and ξ_t are independent random samples that are sampled in iterations $t-1$ and t, respectively. We note that our stochastic residual feedback is more practical than most existing two-point feedback schemes, which require the data samples to be controllable, i.e., one can query the function value at two different variables using the same data sample. This assumption is unrealistic in applications where the environment is dynamic. For example, in reinforcement learning Malik et al. (2018), these data samples can correspond to random initial states, noises added to the dynamical system, and reward functions. Therefore, controlling the data samples requires to hard reset the system to the exact same initial state and apply the same sequence of noises, which is impossible when the data is collected from a real-world system. Our stochastic residual feedback scheme in (6) does not suffer from the same issue since it does not restrict the data sampling procedure. Instead, it simply takes the residual between two consecutive stochastic feedback points. In particular, it is straightforward to show that (6) is an unbiased gradient estimate of the objective function $f_\delta(x)$. Next, we present some assumptions that are used in our analysis later.

Assumption 4.1. *(Bounded Variance) We assume that for any $x \in \mathbb{R}^d$ there exists $\sigma > 0$ such that*

$$\mathbb{E}\big[\big(F(x,\xi) - f(x)\big)^2\big] \le \sigma^2.$$

Assumption 4.1 implies that $\mathbb{E}\big[\big(f(x,\xi_1) - f(x,\xi_2)\big)^2\big] \le 4\sigma^2$. Furthermore, we make the following smoothness assumption in the stochastic setting.

Assumption 4.2. *Let the function $F(x,\xi) \in C^{0,0}$ with Lipschitz constant $L_0(\xi)$. We assume that $L_0(\xi) \le L_0$ for all $\xi \in \Xi$. In addition, let the function $F(x,\xi) \in C^{1,1}$ with Lipschitz constant $L_1(\xi)$. We assume that $L_1(\xi) \le L_1$ for all $\xi \in \Xi$.*

The following lemma provides an upper bound of $\mathbb{E}[\|\tilde{g}(x_t)\|^2]$ in this stochastic setting.

Lemma 4.3. *Let Assumptions 4.1 and 4.2 hold and assume $F(x,\xi) \in C^{0,0}$ with Lipschitz constant $L_0(\xi)$. We have that*

$$\mathbb{E}[\|\tilde{g}(x_t)\|^2] \le \frac{4L_0^2 d\eta^2}{\delta^2}\mathbb{E}[\|\tilde{g}(x_{t-1})\|^2] + 16L_0^2(d+4)^2 + \frac{8\sigma^2 d}{\delta^2}.$$

If we assume that $F(x,\xi) \in C^{1,1}$, the upper bound on the above second moment can be further improved (see supplementary material for the details). However, this improvement does not yield a better iteration complexity due to the uncontrollable samples ξ_t and ξ_{t-1}. More specifically, the uncontrollable samples lead to an additional term $\frac{8\sigma^2 d}{\delta^2}$ in the above second moment bound. According to the analysis in Hu et al. (2016), such a term can significantly degrade the iteration complexity.

4.1 Convergence Analysis

Next, we analyze the iteration complexity of ZO with stochastic residual feedback for both non-convex and convex problems.

Theorem 4.4. *Let Assumptions 4.1 and 4.2 hold and assume also that $F(x,\xi) \in C^{0,0}$. Moreover, assume that SGD in* (5) *with residual feedback is run for $T > 1/(d\epsilon_f)$ iterations and that $\tilde{x}$ is selected from the T iterates uniformly at random. Let also $\eta = \frac{\epsilon_f^{1.5}}{2\sqrt{2}L_0^2 d^{1.5}\sqrt{T}}$ and $\delta = \frac{\epsilon_f}{L_0\sqrt{d}}$. Then, we have that $\mathbb{E}\big[\|\nabla f_\delta(\tilde{x})\|^2\big] = \mathcal{O}(d^{1.5}\epsilon_f^{-1.5}T^{-0.5})$.*

Furthermore, assume that additionally $F(x,\xi) \in C^{1,1}$, and that SGD in (5) *with residual feedback is run for $T > 2$ iterations. Let also $\eta = \frac{1}{2L_0 d^{\frac{4}{3}} T^{\frac{2}{3}}}$ and $\delta = \frac{1}{d^{\frac{5}{6}} T^{\frac{1}{6}}}$. Then, the output $\tilde{x}$ that is sampled uniformly from the T iterates satisfies $\mathbb{E}\big[\|\nabla f(\tilde{x})\|^2\big] = \mathcal{O}(d^{\frac{4}{3}}T^{-\frac{1}{3}})$.*

Based on the above convergence rate results, when $F(x,\xi)$ is non-smooth, to achieve the $\epsilon-$stationary point $\mathbb{E}\big[\|\nabla f_\delta(\tilde{x})\|^2\big] \le \epsilon$ and $|f(x) - f_\delta(x)| \le \epsilon_f$, $\mathcal{O}(\frac{d^3}{\epsilon_f^3\epsilon^2})$ iterations are needed. In addition, if the function $F(x,\xi)$ also satisfies $F(x,\xi) \in C^{1,1}$, then $\mathcal{O}(\frac{d^4}{\epsilon^3})$ iterations are needed to find the $\epsilon-$stationary point of the original function $f(x)$. Next, we provide the iteration complexity results when the Problem (Q) is convex.

Theorem 4.5. *Let Assumptions 4.1 and 4.2 hold and assume that the function $F(x,\xi) \in C^{0,0}$ is also convex. Moreover, assume that SGD in* (5) *with residual feedback is run for T iterations and define the running average $\bar{x} = \frac{1}{T}\sum_{t=0}^{T-1} x_t$. Let also $\eta = \frac{1}{2\sqrt{2}L_0\sqrt{d}T^{\frac{3}{4}}}$ and $\delta = \frac{1}{T^{\frac{1}{4}}}$. Then, we have that $f(\bar{x}) - f(x^*) = \mathcal{O}(\sqrt{d}T^{-\frac{1}{4}})$. Moreover, assume that additionally $F(x,\xi) \in C^{1,1}$, and let $\eta = \frac{1}{2\sqrt{2}L_0 d^{\frac{2}{3}} T^{\frac{2}{3}}}$ and $\delta = \frac{1}{d^{\frac{1}{6}} T^{\frac{1}{6}}}$. Then, we have that $f(\bar{x}) - f(x^*) = \mathcal{O}(d^{\frac{2}{3}} T^{-\frac{1}{3}})$.*

According to Theorem 4.5, $\mathcal{O}(\frac{d^2}{\epsilon^4})$ iterations are needed to achieve $f(\bar{x}) - f(x^*) \leq \epsilon$ with a nonsmooth objective function. On the other hand, if $f(x) \in C^{1,1}$, the iteration complexity is improved to $\mathcal{O}(\frac{d^2}{\epsilon^3})$.

5 ZO with Mini-batch Stochastic Residual Feedback

When applying zeroth-order oracles to practical applications, instead of directly using the oracle (6), a mini-batch scheme can be implemented to further reduce the variance of the gradient estimate, as discussed in Fazel et al. (2018). To be more specific, consider the gradient estimate with batch size b:

$$\tilde{g}_b(x_t) = \frac{u_t}{b\delta}\big(F(x_t + \delta u_t, \xi_{1:b}) - F(x_{t-1} + \delta u_{t-1}, \xi'_{1:b})\big),$$

where $F(x_t + \delta u_t, \xi_{1:b}) = \sum_{j=1}^{b} F(x_t + \delta u_t, \xi_j)$. It is straightforward to see that the variance of $\tilde{g}_b(x_t)$ is b^2 times smaller than that of the oracle (6). This is particularly useful when the problem is sensitive to a bad search direction. For example, in the policy optimization problem Fazel et al. (2018), when the gradient has a large variance, it can easily drive the policy parameter to divergence and result in infinite cost. A Mini-batch scheme can reduce the variance of the policy gradient (search direction) estimate and therefore is of particular interest in this scenario. In this paper, we show that using the oracle (6) in a mini-batch scheme can achieve the same query complexity as standard SGD. Its analysis is provided in the supplementary material.

6 Numerical Experiments

In this section, we demonstrate the effectiveness of the residual one-point feedback scheme for both deterministic and stochastic problems. In the deterministic case, we compare the performance of the proposed oracle with the original one-point feedback and two-point feedback schemes, for the quadratic programming (QP) example considered in Shamir (2013). In the stochastic case, we employ the stochastic variants of above oracles to optimize the policy parameters in a Linear Quadratic Regulation (LQR) problem considered in Fazel et al. (2018); Malik et al. (2018). It is shown that the proposed residual one-point feedback significantly outperforms the traditional one-point feedback and its convergence rate matches that of the two-point oracles in both deterministic and stochastic cases. All experiments are conducted using Matlab R2018b on a 2018 Macbook Pro with a 2.3 GHz Quad-Core Intel Core i5 and 8GB 2133MHz memory.

6.1 A Deterministic Scenario: QP Problem

As in Shamir (2013), consider the QP example $\min \frac{1}{2}(x-c)^T M(x-c)$, where $x, c \in \mathbb{R}^{30}$ and $M \in \mathbb{R}^{30\times 30}$ is a positive semi-definite matrix. This constitutes a convex and smooth problem. The vector c is randomly generated from a uniform distribution in $[0, 2]$. The matrix $M = PP^T$, where each entry in $P \in \mathbb{R}^{30\times 29}$ is sampled from a uniform distribution in $[0, 1]$. The initial point is set to be the origin. For every algorithm tested, we manually optimize the selection of the exploration parameter δ and stepsize η and run it 100 times. The convergence of the function value $f(x) - f(x^*)$ is presented in Figure 1(a). We observe that the proposed oracle converges as fast as the two-point oracle (2) when the iterates are far from the optimizer but achieve less accuracy in the end. Both methods find the optimal function value much faster than the one-point feedback studied in Flaxman et al. (2005); Gasnikov et al. (2017). These observations validate our theoretical results in Section 3.

6.2 A Stochastic Scenario: Policy Optimization

We use the proposed residual feedback to optimize the policy parameters in a LQR problem, as in Fazel et al. (2018); Malik et al. (2018). Specifically, consider a system whose state $x_k \in \mathbb{R}^{n_x}$ at time k

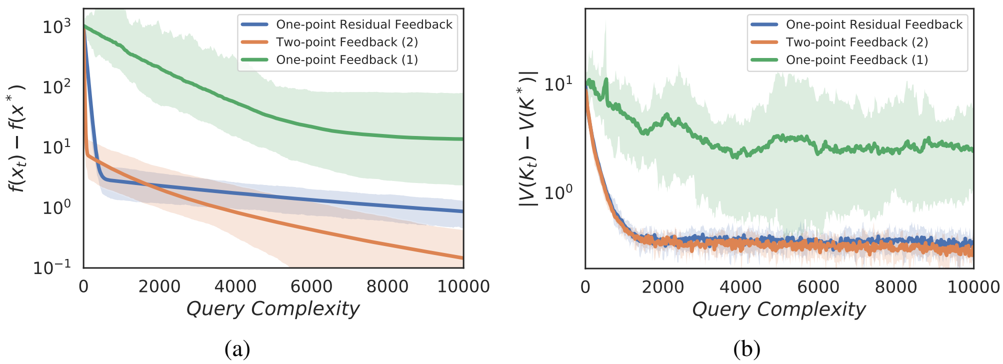

Figure 1: The convergence rate of applying the proposed residual one-point feedback (4) (blue), the two-point oracle (2) in Nesterov & Spokoiny (2017) (orange) and the one-point oracle (1) in Flaxman et al. (2005) (green) to two problems. In (a), the convergence of $f(x_t) - f(x^*)$ in a deterministic QP problem is presented. In (b), the convergence of the costs of policies in the stochastic LQR problem is presented. The horizontal axis shows the number of queries to the value of the objective function.

is subject to the dynamical equation $x_{k+1} = Ax_k + Bu_k + w_k$, where $u_k \in \mathbb{R}^{n_u}$ is the control input at time k, $A \in \mathbb{R}^{n_x \times n_x}$ and $B \in \mathbb{R}^{n_x \times n_u}$ are dynamical matrices that are unknown, and w_k is the noise on the state transition. Moreover, consider a state feedback policy $u_k = Kx_k$, where $K \in \mathbb{R}^{n_u \times n_x}$ is the policy parameter. Policy optimization essentially aims to find the optimal policy parameter K so that the discounted accumulated cost function $V(K) := \mathbb{E}\big[\sum_{t=0}^{\infty} \gamma^t (x_k^T Q x_k + u_k^T R u_k)\big]$ is minimized, where $\gamma \leq 1$ is the discount factor.

In our simulation, we select $n_x = 6$, $n_u = 6$ and $\gamma = 0.5$. Therefore, the problem has dimension $d = 36$. When implementing the policy $u_k = K_t x_k$, due to the noise w_k, evaluation of the cost of the policy K_t is noisy. We apply the one-point feedback (1) with noise Gasnikov et al. (2017), two-point feedback with uncontrolled noise Hu et al. (2016); Bach & Perchet (2016) and the residual one-point feedback (6) to solve the above policy optimization problem. To evaluate the cost $V(K_t)$ given the policy parameter K_t at iteration t, we run one episode with a finite horizon length $H = 50$. The dynamical matrices A and B are randomly generated and the noise w_k is sampled from a Gaussian distribution $\mathcal{N}(0, 0.1^2)$. We run each algorithm 10 times. At each trial, all the algorithms start from the same initial guess of the policy parameter K_0, which is generated by perturbing the optimal policy parameter K^* with a random matrix, as in Malik et al. (2018). Each entry in this random perturbation matrix is sampled from a uniform distribution in $[0, 0.2]$. The performance of all the algorithms over 10 trials is measured in terms of $|V(K_t) - V(K^*)|$ and is presented in Figure 1(b). We observe that the residual one-point feedback (6) converges much faster than the one-point oracle in Gasnikov et al. (2017) and has comparable query complexity to the two-point feedback under uncontrolled noises considered in Hu et al. (2016); Bach & Perchet (2016). This corroborates our theoretical analysis in Section 4. Furthermore, we apply our residual-feedback zeroth-order gradient estimate to solve a large-scale stochastic multi-stage decision making problem with problem dimension $d = 576$. The implementation details and results of the simulation can be found in the supplementary material, where it can be seen that our residual feedback achieves similar improvement of the convergence rate over the conventional one-point feedback scheme.

7 Conclusion

In this paper, we proposed a residual one-point feedback oracle for zeroth-order optimization, which estimates the gradient of the objective function using a single query of the function value at each iteration. When the function evaluation is noiseless, we showed that ZO using the proposed oracle can achieve the same iteration complexity as ZO using two-point oracles when the function is non-smooth. When the function is smooth, this complexity of ZO can be further improved. This is the first time that a one-point zeroth-order oracle is shown to match the performance of two-point oracles in ZO. In addition, we considered a more realistic scenario where the function evaluation is corrupted by noise. We showed that the convergence rate of ZO using the proposed oracle matches the best known results using one-point feedback or two-point feedback with uncontrollable data samples. We provided numerical experiments that showed that the proposed oracle outperforms the one-point oracle and is as effective as two-point feedback methods.

Broader Impact

This work develops a new tool for derivative-free optimization and provides supporting theoretical convergence guarantees. Derivative-free methods can significantly impact a variety of practical applications where exact system models are unknown or are known but are too complex to differentiate. An example are decision making problems that incorporate human preference, which usually cannot be modeled using explicit mathematical formulas. Other examples include machine learning and reinforcement learning problems that rely on deep neural networks. Such problems are common in many everyday systems, including robotic, cyber-physical systems, and IoT systems. The zeroth-order oracle proposed in this paper significantly improves the performance of derivative-free methods and, therefore, their applicability in controlling such systems. At the same time, derivative-free optimization methods can also be used to attack black-box cyber or physical systems. Therefore, as the study in this field advances, safety concerns related to zeroth-order attacks must also be addressed.

References

Agarwal, A., Dekel, O., and Xiao, L. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *COLT*, pp. 28–40. Citeseer, 2010.

Bach, F. and Perchet, V. Highly-smooth zero-th order online optimization. In *Conference on Learning Theory*, pp. 257–283, 2016.

Balasubramanian, K. and Ghadimi, S. Zeroth-order (non)-convex stochastic optimization via conditional gradient and gradient updates. In *Advances in Neural Information Processing Systems*, pp. 3455–3464, 2018.

Chen, P.-Y., Zhang, H., Sharma, Y., Yi, J., and Hsieh, C.-J. Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models. In *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*, pp. 15–26, 2017.

Dekel, O., Eldan, R., and Koren, T. Bandit smooth convex optimization: Improving the bias-variance tradeoff. In *Advances in Neural Information Processing Systems*, pp. 2926–2934, 2015.

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

Fazel, M., Ge, R., Kakade, S., and Mesbahi, M. Global convergence of policy gradient methods for the linear quadratic regulator. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, 2018.

Flaxman, A. D., Kalai, A. T., and McMahan, H. B. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 385–394. Society for Industrial and Applied Mathematics, 2005.

Gasnikov, A. V., Krymova, E. A., Lagunovskaya, A. A., Usmanova, I. N., and Fedorenko, F. A. Stochastic online optimization. single-point and multi-point non-linear multi-armed bandits. convex and strongly-convex case. *Automation and remote control*, 78(2):224–234, 2017.

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

Gorbunov, E., Dvurechensky, P., and Gasnikov, A. An accelerated method for derivative-free smooth stochastic convex optimization. *arXiv preprint arXiv:1802.09022*, 2018.

Hajinezhad, D. and Zavlanos, M. M. Gradient-free multi-agent nonconvex nonsmooth optimization. In *2018 IEEE Conference on Decision and Control (CDC)*, pp. 4939–4944. IEEE, 2018.

Hu, X., Prashanth, L., György, A., and Szepesvári, C. (bandit) convex optimization with biased noisy gradient oracles. In *Artificial Intelligence and Statistics*, pp. 819–828, 2016.

Ji, K., Wang, Z., Zhou, Y., and Liang, Y. Improved zeroth-order variance reduced algorithms and analysis for nonconvex optimization. *arXiv preprint arXiv:1910.12166*, 2019.

Kim, M., Ding, Y., Malcolm, P., Speeckaert, J., Siviy, C. J., Walsh, C. J., and Kuindersma, S. Human-in-the-loop bayesian optimization of wearable device parameters. *PloS one*, 12(9), 2017.

Larson, J., Menickelly, M., and Wild, S. M. Derivative-free optimization methods. *Acta Numerica*, 28:287–404, 2019.

Malik, D., Pananjady, A., Bhatia, K., Khamaru, K., Bartlett, P. L., and Wainwright, M. J. Derivative-free methods for policy optimization: Guarantees for linear quadratic systems. *arXiv preprint arXiv:1812.08305*, 2018.

Nesterov, Y. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

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

Saha, A. and Tewari, A. Improved regret guarantees for online smooth convex optimization with bandit feedback. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 636–642, 2011.

Shamir, O. On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on Learning Theory*, pp. 3–24, 2013.

Shamir, O. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *Journal of Machine Learning Research*, 18(52):1–11, 2017.

Tang, Y. and Li, N. Distributed zero-order algorithms for nonconvex multi-agent optimization. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 781–786. IEEE, 2019.

Wang, Y., Du, S., Balakrishnan, S., and Singh, A. Stochastic zeroth-order optimization in high dimensions. In *International Conference on Artificial Intelligence and Statistics*, pp. 1356–1365, 2018.

Supplementary Materials

A Zeroth-Order Policy Optimization for A Large-Scale Multi-Stage Decision Making Problem

In this section, we consider a large-scale multi-stage resource allocation problem. Specifically, we consider 16 agents that are located on a 4×4 grid. At agent i, resources are stored in the amount of $m_i(k)$ and there is also a demand for resources in the amount of $d_i(k)$ at instant k. In the meantime, agent i also decides what fraction of resources $a_{ij}(k) \in [0,1]$ it sends to its neighbors $j \in \mathcal{N}_i$ on the grid. The local amount of resources and demands at agent i evolve as $m_i(k+1) = m_i(k) - \sum_{j \in \mathcal{N}_i} a_{ij}(k) m_i(k) + \sum_{j \in \mathcal{N}_i} a_{ji}(k) m_j(k) - d_i(k)$ and $d_i(k) = A_i \sin(\omega_i k + \phi_i) + w_{i,k}$, where $w_{i,k}$ is the noise in the demand. At time k, agent i receives a local reward $r_i(k)$, such that $r_i(k) = 0$ when $m_i(k) \geq 0$ and $r_i(k) = -m_i(k)^2$ when $m_i(k) < 0$. Let agent i makes its decisions according to a parameterized policy function $\pi_{i,\theta_i}(o_i) : \mathcal{O}_i \to [0,1]^{|\mathcal{N}_i|}$, where θ_i is the parameter of the policy function π_i, $o_i \in \mathcal{O}_i$ denotes agent i's observation, and $|\mathcal{N}_i|$ represents the number of agent i's neighbors on the grid.

Our goal is to train a policy that can be executed in a fully distributed way based on agents' local information. Specifically, during the execution of policy functions $\{\pi_{i,\theta_i}(o_i)\}$, we let each agent only observe its local amount of resource $m_i(k)$ and demand and $d_i(k)$, i.e., $o_i(k) = [m_i(k), d_i(k)]^T$. In addition, the policy function $\pi_{i,\theta_i}(o_i)$ is parameterized as the following: $a_{ij} = \exp(z_{ij}) / \sum_j \exp(z_{ij})$, where $z_{ij} = \sum_{p=1}^{9} \psi_p(o_i) \theta_{ij}(p)$ and $\theta_i = [\dots, \theta_{ij}, \dots]^T$. Specifically, the feature function $\psi_p(o_i)$ is selected as $\psi_p(o_i) = \|o_i - c_p\|^2$, where c_p is the parameter of the p-th feature function. The goal for the agents is to find an optimal policy $\pi^* = \{\pi_{i,\theta_i}(o_i)\}$ so that the global accumulated reward

$$J(\theta) = \sum_{i=1}^{16} \sum_{k=0}^{K} \gamma^k r_i(k) \tag{7}$$

is maximized, where $\theta = [\dots, \theta_i, \dots]$ is the global policy parameter, K is the horizon of the problem, and γ is the discount factor. Effectively, the agents need to make decisions on 64 actions, and each action is decided by 9 parameters. Therefore, the problem dimension is $d = 576$. To implement zeroth-order policy gradient estimators (1) and (6) to find the optimal policy, at iteration t, we let all agents implement the policy with parameter $\theta_t + \delta u_t$, collect rewards $\{r_i(k)\}$ at time instants $k = 0, 1, \dots, K$ and compute the noisy policy value according to (7). Then, the zeroth-order policy gradient is estimated using (1) or (6). On the contrary, when the two-point zeroth-order policy gradient estimator (2) is used, at each iteration k, all agents need to evaluate two policies $\theta_t \pm \delta u_t$ to update the policy parameter once. In Figure 2, we present the performance of using zeroth-order policy gradients (1), (2) and (6) to solve this large-scale multi-stage resource allocation problem, where the discount factor is set as $\gamma = 0.75$ and the length of horizon $K = 30$. Each algorithm is run for 10 trials. We observe that policy optimization with the proposed residual-feedback gradient estimate (6) improves the optimal policy parameters with the same learning rate as the two-point zeroth-order gradient estimator (2), where the learning rate is measured by the number of episodes the agents take to evaluate the policy parameter iterates. In the meantime, both estimators perform much better than the one-point policy gradient estimate (1) considered in Fazel et al. (2018); Malik et al. (2018).

B Proof of Lemma 3.2

First, we show the bound when $f(x) \in C^{0,0}$. Recalling the expression of $\tilde{g}(x_t)$ in (4), we have that

$$\begin{aligned}
\mathbb{E}[\|\tilde{g}(x_t)\|^2] &= \mathbb{E}[\frac{1}{\delta^2}\big(f(x_t + \delta u_t) - f(x_{t-1} + \delta u_{t-1})\big)^2 \|u_t\|^2] \\
&\leq \frac{2}{\delta^2}\mathbb{E}[\big(f(x_t + \delta u_t) - f(x_{t-1} + \delta u_t)\big)^2 \|u_t\|^2] \\
&\quad + \frac{2}{\delta^2}\mathbb{E}[\big(f(x_{t-1} + \delta u_t) - f(x_{t-1} + \delta u_{t-1})\big)^2 \|u_t\|^2].
\end{aligned} \tag{8}$$

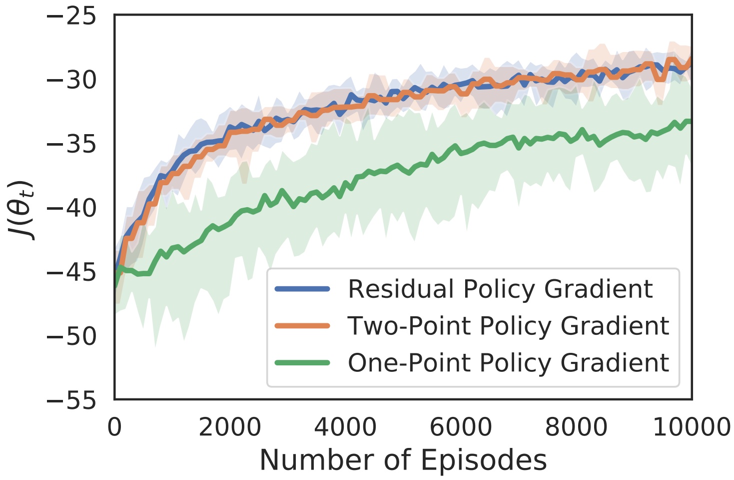

Figure 2: The convergence rate of applying the proposed residual one-point feedback (4) (blue), the two-point oracle (2) in Nesterov & Spokoiny (2017) (orange) and the one-point oracle (1) in Flaxman et al. (2005) (green) to the large-scale stochastic multi-stage resource allocation problem. The vertical axis represents the total rewards and the horizontal axis represents the number of episodes the agents take to evaluate their policy parameter iterates during the policy optimization procedure.

Since function $f \in C^{0,0}$ with Lipschitz constant L_0, we obtain that

$$\mathbb{E}[\|\tilde{g}(x_t)\|^2] \leq \frac{2L_0^2}{\delta^2}\mathbb{E}[\|x_t - x_{t-1}\|^2\|u_t\|^2] + 2L_0^2\mathbb{E}[\|u_t - u_{t-1}\|^2\|u_t\|^2]. \tag{9}$$

Since u_t is independently sampled from $x_t - x_{t-1}$, we have that $\mathbb{E}[\|x_t - x_{t-1}\|^2\|u_t\|^2] = \mathbb{E}[\|x_t - x_{t-1}\|^2]\mathbb{E}[\|u_t\|^2]$. Since u_t is subject to standard multivariate normal distribution, $\mathbb{E}[\|u_t\|^2] = d$. Furthermore, using Lemma 1 in Nesterov & Spokoiny (2017), we get that $\mathbb{E}[\|u_t - u_{t-1}\|^2\|u_t\|^2] \leq 2\mathbb{E}[(\|u_t\|^2 + \|u_{t-1}\|^2)\|u_t\|^2] = 2\mathbb{E}[(\|u_t\|^4] + 2\mathbb{E}[\|u_{t-1}\|^2\|u_t\|^2] \leq 4(d+4)^2$. Plugging these bounds into inequality (9), we have that

$$\mathbb{E}[\|\tilde{g}(x_t)\|^2] \leq \frac{2dL_0^2}{\delta^2}\mathbb{E}[\|x_t - x_{t-1}\|^2] + 8L_0^2(d+4)^2.$$

Since $x_t = x_{t-1} - \eta\tilde{g}(x_{t-1})$, we get that

$$\mathbb{E}[\|\tilde{g}(x_t)\|^2] \leq \frac{2dL_0^2\eta^2}{\delta^2}\mathbb{E}[\|\tilde{g}(x_{t-1})\|^2] + 8L_0^2(d+4)^2.$$

Next, we show the bound when we have the additional smoothness condition $f(x) \in C^{1,1}$ with constant L_1. Given the gradient estimate in (4), we have that

$$\mathbb{E}[\|\tilde{g}(x_t)\|^2] \leq \mathbb{E}[\frac{(f(x_t + \delta u_t) - f(x_{t-1} + \delta u_{t-1}))^2}{\delta^2}\|u_t\|^2]. \tag{10}$$

Next, we bound the term $(f(x_t + \delta u_t) - f(x_{t-1} + \delta u_{t-1}))^2$. Adding and subtracting $f(x_{t-1} + \delta u_t)$ inside the square, and applying the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, we can obtain

$$\begin{aligned}(f(x_t + \delta u_t) - f(x_{t-1} + \delta u_{t-1}))^2 \leq & 2(f(x_t + \delta u_t) - f(x_{t-1} + \delta u_t))^2 \\ & + 2(f(x_{t-1} + \delta u_t) - f(x_{t-1} + \delta u_{t-1}))^2. \end{aligned} \tag{11}$$

Since the function $f(x)$ is also Lipschitz continuous with constant L_0, we get that

$$(f(x_t + \delta u_t) - f(x_{t-1} + \delta u_t))^2 \leq L_0^2\|x_t - x_{t-1}\|^2 = L_0^2\eta^2\|\tilde{g}(x_{t-1})\|^2. \tag{12}$$

Next, we bound the term $(f(x_{t-1} + \delta u_t) - f(x_{t-1} + \delta u_{t-1}))^2$. Adding and subtracting $f(x_{t-1})$, $\langle\nabla f(x_{t-1}), \delta u_t\rangle$ and $\langle\nabla f(x_{t-1}), \delta u_{t-1}\rangle$ inside the square term, we have that

$$\begin{aligned}&(f(x_{t-1} + \delta u_t) - f(x_{t-1} + \delta u_{t-1}))^2 \\ \leq\ & 2\langle\nabla f(x_{t-1}), \delta(u_t - u_{t-1})\rangle^2 + 4(f(x_{t-1} + \delta u_t) - f(x_{t-1}) - \langle\nabla f(x_{t-1}), \delta u_t\rangle)^2 \\ & + 4(f(x_{t-1} + \delta u_{t-1}) - f(x_{t-1}) - \langle\nabla f(x_{t-1}), \delta u_{t-1}\rangle)^2. \end{aligned} \tag{13}$$

Since $f(x) \in C^{1,1}$ with constant L_1, we get that $|f(x_{t-1} + \delta u_t) - f(x_{t-1}) - \langle \nabla f(x_{t-1}), \delta u_t \rangle| \leq \frac{1}{2} L_1 \delta^2 \|u_t\|^2$, according to (6) in Nesterov & Spokoiny (2017). And similarly, we also have $|f(x_{t-1} + \delta u_{t-1}) - f(x_{t-1}) - \langle \nabla f(x_{t-1}), \delta u_{t-1} \rangle| \leq \frac{1}{2} L_1 \delta^2 \|u_{t-1}\|^2$. Substituting these inequalities into (13), we obtain that

$$(f(x_{t-1}+\delta u_t) - f(x_{t-1}+\delta u_{t-1}))^2 \leq 2\langle \nabla f(x_{t-1}), \delta(u_t - u_{t-1}) \rangle^2 + L_1^2 \delta^4 \|u_t\|^4 + L_1^2 \delta^4 \|u_{t-1}\|^4. \tag{14}$$

Moreover, substituting the inequalities (12) and (14) in the upper bound in (11), we get that

$$\begin{aligned} (f(x_t + \delta u_t) - f(x_{t-1} + \delta u_{t-1}))^2 \leq & 2L_0^2 \eta^2 \|\tilde{g}(x_{t-1})\|^2 + 4\langle \nabla f(x_{t-1}), \delta(u_t - u_{t-1}) \rangle^2 \\ & + 2L_1^2 \delta^4 \|u_t\|^4 + 2L_1^2 \delta^4 \|u_{t-1}\|^4 \end{aligned} \tag{15}$$

Using the bound (15) in inequality (10), and applying the bounds $\mathbb{E}[\|u_t\|^6] \leq (d+6)^3$ and $\mathbb{E}[\|u_{t-1}\|^4 \|u_t\|^2] \leq (d+6)^3$, we have that

$$\mathbb{E}[\|\tilde{g}(x_t)\|^2] \leq \frac{2dL_0^2\eta^2}{\delta^2} \mathbb{E}[\|\tilde{g}(x_{t-1})\|^2] + 4\mathbb{E}[\langle \nabla f(x_{t-1}), u_t - u_{t-1} \rangle^2 \|u_t\|^2] + 4L_1^2 (d+6)^3 \delta^2. \tag{16}$$

Since $\langle \nabla f(x_{t-1}), u_t - u_{t-1} \rangle^2 \leq 2\langle \nabla f(x_{t-1}), u_t \rangle^2 + 2\langle \nabla f(x_{t-1}), u_{t-1} \rangle^2$, we get that

$$\mathbb{E}[\langle \nabla f(x_{t-1}), u_t - u_{t-1} \rangle^2 \|u_t\|^2] \leq 2\mathbb{E}[\langle \nabla f(x_{t-1}), u_t \rangle^2 \|u_t\|^2] + 2\mathbb{E}[\langle \nabla f(x_{t-1}), u_{t-1} \rangle^2 \|u_t\|^2]. \tag{17}$$

For the term $\mathbb{E}[\langle \nabla f(x_{t-1}), u_{t-1} \rangle^2 \|u_t\|^2]$, we have that $\mathbb{E}[\langle \nabla f(x_{t-1}), u_{t-1} \rangle^2 \|u_t\|^2] \leq \mathbb{E}[\|\nabla f(x_{t-1})\|^2 \|u_{t-1}\|^2 \|u_t\|^2] \leq d^2 \mathbb{E}[\|\nabla f(x_{t-1})\|^2]$. For the term $\mathbb{E}[\langle \nabla f(x_{t-1}), u_t \rangle^2 \|u_t\|^2]$, according to Theorem 3 in Nesterov & Spokoiny (2017), we have a stronger bound $\mathbb{E}[\langle \nabla f(x_{t-1}), u_t \rangle^2 \|u_t\|^2] \leq (d+4)\mathbb{E}[\|\nabla f(x_{t-1})\|^2]$. Substituting these bounds into (17), and because $d^2 + d + 4 \leq (d+4)^2$, we have that

$$\mathbb{E}[\langle \nabla f(x_{t-1}), u_t - u_{t-1} \rangle^2 \|u_t\|^2] \leq 2(d+4)^2 \mathbb{E}[\|\nabla f(x_{t-1})\|^2]. \tag{18}$$

Substituting the bound (18) into inequality (16), we complete the proof.

C Proof of Theorem 3.3

Since we have that $f(x) \in C^{0,0}$, according to Lemma 2.2, the function $f_\delta(x)$ has $L_1(f_\delta)$-Lipschitz continuous gradient where $L_1(f_\delta) = \frac{\sqrt{d}}{\delta} L_0$. Furthermore, according to Lemma 1.2.3 in Nesterov (2013), we can get the following inequality

$$\begin{aligned} f_\delta(x_{t+1}) &\leq f_\delta(x_t) + \langle \nabla f_\delta(x_t), x_{t+1} - x_t \rangle + \frac{L_1(f_\delta)}{2} \|x_{t+1} - x_t\|^2 \\ &= f_\delta(x_t) - \eta \langle \nabla f_\delta(x_t), \tilde{g}(x_t) \rangle + \frac{L_1(f_\delta)\eta^2}{2} \|\tilde{g}(x_t)\|^2 \\ &= f_\delta(x_t) - \eta \langle \nabla f_\delta(x_t), \Delta_t \rangle - \eta \|\nabla f_\delta(x_t)\|^2 + \frac{L_1(f_\delta)\eta^2}{2} \|\tilde{g}(x_t)\|^2, \end{aligned} \tag{19}$$

where $\Delta_t = \tilde{g}(x_t) - \nabla f_\delta(x_t)$. According to Lemma 3.1, we can get that $\mathbb{E}_{u_t}[\tilde{g}(x_t)] = \nabla f_\delta(x_t)$. Therefore, taking expectation over u_t on both sides of inequality (19) and rearranging terms, we have that

$$\eta \mathbb{E}[\|\nabla f_\delta(x_t)\|^2] \leq \mathbb{E}[f_\delta(x_t)] - \mathbb{E}[f_\delta(x_{t+1})] + \frac{L_1(f_\delta)\eta^2}{2} \mathbb{E}[\|\tilde{g}(x_t)\|^2]. \tag{20}$$

Telescoping above inequalities from $t = 0$ to $T - 1$ and dividing both sides by η, we obtain that

$$\begin{aligned} \sum_{t=0}^{T-1} \mathbb{E}[\|\nabla f_\delta(x_t)\|^2] &\leq \frac{\mathbb{E}[f_\delta(x_0)] - \mathbb{E}[f_\delta(x_T)]}{\eta} + \frac{L_1(f_\delta)\eta}{2} \sum_{t=0}^{T-1} \mathbb{E}[\|\tilde{g}(x_t)\|^2] \\ &\leq \frac{\mathbb{E}[f_\delta(x_0)] - f_\delta^*}{\eta} + \frac{L_1(f_\delta)\eta}{2} \sum_{t=0}^{T-1} \mathbb{E}[\|\tilde{g}(x_t)\|^2], \end{aligned} \tag{21}$$

where f_δ^* is the lower bound of the smoothed function $f_\delta(x)$. f_δ^* must exist because we assume the orignal function $f(x)$ is lower bounded and the smoothed function has a bounded distance from $f(x)$ due to Lemma 2.2.

Recall the contraction result of the second moment $\mathbb{E}[\|\tilde{g}(x_t)\|^2]$ in Lemma 3.2 when $f(x) \in C^{0,0}$. Denote the contraction rate $\frac{2dL_0^2\eta^2}{\delta^2}$ as α and the constant perturbation term $M = 8L_0^2(d+4)^2$. Then, we get that

$$\mathbb{E}[\|\tilde{g}(x_t)\|^2] \le \alpha^t \mathbb{E}[\|\tilde{g}(x_0)\|^2] + \frac{1-\alpha^t}{1-\alpha} M. \tag{22}$$

Summing the above inequality over time, we obtain

$$\begin{aligned}\sum_{t=0}^{T-1} \|\tilde{g}(x_t)\|^2 &\le \frac{1-\alpha^T}{1-\alpha}\mathbb{E}[\|\tilde{g}(x_0)\|^2] + \sum_{t=0}^{T-1}\big(\frac{1-\alpha^t}{1-\alpha}M\big) \\ &\le \frac{1}{1-\alpha}\mathbb{E}[\|\tilde{g}(x_0)\|^2] + \frac{1}{1-\alpha}MT.\end{aligned} \tag{23}$$

Plugging the bound in (23) into inequality (21), and since $L_1(f_\delta) = \frac{\sqrt{d}}{\delta}L_0$, we have that

$$\sum_{t=0}^{T-1} \mathbb{E}[\|\nabla f_\delta(x_t)\|^2] \le \frac{\mathbb{E}[f_\delta(x_0)] - f_\delta^*}{\eta} + \frac{d^{\frac{1}{2}}L_0\eta}{\delta}\big(\frac{1}{1-\alpha}\mathbb{E}[\|\tilde{g}(x_0)\|^2] + \frac{1}{1-\alpha}8L_0^2(d+4)^2T\big). \tag{24}$$

To fullfill the requirement that $|f(x) - f_\delta(x)| \le \epsilon_f$, we set the exporation parameter $\delta = \frac{\epsilon_f}{d^{\frac{1}{2}}L_0}$. In addition, let the stepsize be $\eta = \frac{\sqrt{\epsilon_f}}{2dL_0^2T^{\frac{1}{2}}}$. We have that $\alpha = \frac{1}{2T\epsilon_f} \le \frac{1}{2}$ and $\frac{1}{1-\alpha} \le 2$, when $T \ge \frac{1}{\epsilon_f}$. Plugging the choices of η and δ into inequality (24), we obtain that

$$\sum_{t=0}^{T-1} \mathbb{E}[\|\nabla f_\delta(x_t)\|^2] \le 2L_0^2\big(\mathbb{E}[f_\delta(x_t)] - f_\delta^*\big)\frac{d}{\sqrt{\epsilon_f}}\sqrt{T} + \mathbb{E}[\|\tilde{g}(x_0)\|^2] + 8L_0^2\frac{(d+4)^2}{\sqrt{\epsilon_f}}\sqrt{T}.$$

Dividing both sides of above inequality by T, we complete the proof.

D Proof of Theorem 3.4

Following the same process in the beginning of the proof of Theorem 3.3, we can get

$$\sum_{t=0}^{T-1} \mathbb{E}[\|\nabla f_\delta(x_t)\|^2] \le \frac{\mathbb{E}[f_\delta(x_0)] - f_\delta^*}{\eta} + \frac{L_1\eta}{2}\sum_{t=0}^{T-1}\mathbb{E}[\|\tilde{g}(x_t)\|^2]. \tag{25}$$

Since $\frac{1}{2}\mathbb{E}[\|\nabla f(x_t)\|^2] \le \mathbb{E}[\|\nabla f_\delta(x_t)\|^2] + \mathbb{E}[\|\nabla f(x_t) - \nabla f_\delta(x_t)\|^2]$, and according to the bound (25) and Lemma 2.2, we have that

$$\frac{1}{2}\sum_{t=0}^{T-1} \|\mathbb{E}[\|\nabla f(x_t)\|^2\|] \le \frac{\mathbb{E}[f_\delta(x_0)] - f_\delta^*}{\eta} + \frac{L_1\eta}{2}\sum_{t=0}^{T-1}\mathbb{E}[\|\tilde{g}(x_t)\|^2] + L_1^2(d+3)^3\delta^2T. \tag{26}$$

In addition, similar to the process to derive the bound in (23), according to Lemma 3.2, when $f(x) \in C^{1,1}$, we can get that

$$\sum_{t=0}^{T-1} \|\tilde{g}(x_t)\|^2 \le \frac{1}{1-\alpha}\mathbb{E}[\|\tilde{g}(x_0)\|^2] + \frac{8}{1-\alpha}(d+4)^2\sum_{t=0}^{T-1}\|\nabla f(x_t)\|^2 + \frac{4}{1-\alpha}L_1^2(d+6)^3\delta^2T. \tag{27}$$

Plugging the bound (27) into (26), we have that

$$\begin{aligned}\frac{1}{2}\sum_{t=0}^{T-1} \|\mathbb{E}[\|\nabla f(x_t)\|^2\|] \le& \frac{\mathbb{E}[f_\delta(x_0)] - f_\delta^*}{\eta} + \frac{L_1\eta}{2}\big(\frac{1}{1-\alpha}\mathbb{E}[\|\tilde{g}(x_0)\|^2] + \frac{4}{1-\alpha}L_1^2(d+6)^3\delta^2T \\ &+ \frac{8}{1-\alpha}(d+4)^2\sum_{t=0}^{T-1}\mathbb{E}[\|\nabla f(x_t)\|^2]\big) + L_1^2(d+3)^3\delta^2T.\end{aligned} \tag{28}$$

Recalling that $\tilde{L} = \max\{32L_1, 2L_0\}$, let $\eta = \frac{1}{\tilde{L}(d+4)^2 T^{\frac{1}{3}}}$ and $\delta = \frac{1}{\sqrt{d}T^{\frac{1}{3}}}$, and we have that $\alpha = 2dL_0^2 \frac{\eta^2}{\delta^2} \leq \frac{1}{2}$. In addition, the coefficient before the term $\|\nabla f(x_t)\|^2$ in the upper bound above $\frac{L_1\eta}{2}\frac{8}{1-\alpha}(d+4)^2 \leq \frac{1}{4}$. Therefore, we obtain that

$$\begin{aligned}\frac{1}{4}\sum_{t=0}^{T-1}\|\mathbb{E}[\|\nabla f(x_t)\|^2\|] \leq & \tilde{L}(\mathbb{E}[f_\delta(x_0)] - f_\delta^*)(d+4)^2 T^{\frac{1}{3}} + \frac{1}{32(d+4)^2 T^{\frac{1}{3}}}\mathbb{E}[\|\tilde{g}(x_0)\|^2] \\ & + \frac{L_1^2}{8}\frac{(d+6)^3}{(d+4)^2 d} + L_1^2\frac{(d+3)^3}{d}T^{\frac{1}{3}}.\end{aligned} \tag{29}$$

Dividing both sides of above inequality by T, we complete the proof.

E Proof of Theorem 3.5

First, according to iteration (5), we have that

$$\|x_{t+1} - x^*\|^2 \leq \|x_t - \eta\tilde{g}(x_t) - x^*\|^2 = \|x_t - x^*\|^2 - 2\eta\langle\tilde{g}(x_t), x_t - x^*\rangle + \eta^2\|\tilde{g}(x_t)\|^2. \tag{30}$$

Taking expectation on both sides, and since $\mathbb{E}[\tilde{g}(x_t)] = \nabla f_\delta(x_t)$, we obtain that

$$\mathbb{E}[\|x_{t+1} - x^*\|^2] \leq \mathbb{E}[\|x_t - x^*\|^2] - 2\eta\langle\nabla f_\delta(x_t), x_t - x^*\rangle + \eta^2\mathbb{E}[\|\tilde{g}(x_t)\|^2]. \tag{31}$$

Due to the convexity, we have that $\langle\nabla f_\delta(x_t), x_t - x^*\rangle \geq f_\delta(x_t) - f_\delta(x^*)$. Plugging this inequality into (31), we have that

$$\mathbb{E}[\|x_{t+1} - x^*\|^2] \leq \mathbb{E}[\|x_t - x^*\|^2] - 2\eta(f_\delta(x_t) - f_\delta(x^*)) + \eta^2\mathbb{E}[\|\tilde{g}(x_t)\|^2]. \tag{32}$$

When $f(x) \in C^{0,0}$, using Lemma (2.2), we can replace $f_\delta(x)$ with $f(x)$ in above inequality and get

$$\mathbb{E}[\|x_{t+1} - x^*\|^2] \leq \mathbb{E}[\|x_t - x^*\|^2] - 2\eta(f(x_t) - f(x^*)) + \eta^2\mathbb{E}[\|\tilde{g}(x_t)\|^2] + 4L_0\sqrt{d}\delta\eta. \tag{33}$$

Rearranging the terms and telescoping from $t = 0$ to $T - 1$, we obtain that

$$\begin{aligned}\sum_{t=0}^{T-1}\mathbb{E}[f(x_t)] - Tf(x^*) &\leq \frac{1}{2\eta}(\|x_0 - x^*\|^2 - \mathbb{E}[\|x_T - x^*\|^2]) + \frac{\eta}{2}\sum_{t=0}^{T-1}\mathbb{E}[\|\tilde{g}(x_t)\|^2] + 2L_0\sqrt{d}\delta T \\ &\leq \frac{1}{2\eta}\|x_0 - x^*\|^2 + \frac{\eta}{2}\sum_{t=0}^{T-1}\mathbb{E}[\|\tilde{g}(x_t)\|^2] + 2L_0\sqrt{d}\delta T\end{aligned} \tag{34}$$

Since function $f(x) \in C^{0,0}$, we can plug the bound (23) into the above inequality and get that

$$\sum_{t=0}^{T-1}\mathbb{E}[f(x_t)] - Tf(x^*) \leq \frac{1}{2\eta}\|x_0 - x^*\|^2 + \frac{\eta}{2(1-\alpha)}\mathbb{E}[\|\tilde{g}(x_0)\|^2] + \frac{4\eta}{1-\alpha}L_0^2(d+4)^2 T + 2L_0\sqrt{d}\delta T. \tag{35}$$

Let $\eta = \frac{1}{2dL_0\sqrt{T}}$ and $\delta = \frac{1}{\sqrt{T}}$. We have that $\alpha = 2dL_0^2\frac{\eta^2}{\delta^2} = \frac{1}{2d} \leq \frac{1}{2}$. Therefore, $\frac{1}{1-\alpha} \leq 2$. Applying this bound and the choice of η and δ into above inequality, we have that

$$\begin{aligned}\sum_{t=0}^{T-1}\mathbb{E}[f(x_t)] - Tf(x^*) \leq & L_0\|x_0 - x^*\|^2 d\sqrt{T} + \frac{1}{2dL_0\sqrt{T}}\mathbb{E}[\|\tilde{g}(x_0)\|^2] \\ & + 4L_0\frac{(d+4)^2}{d}\sqrt{T} + 2L_0\sqrt{d}\sqrt{T}.\end{aligned} \tag{36}$$

Recalling that $f(\bar{x}) \leq \frac{1}{T}\sum_{t=0}^{T-1} f(x_t)$ due to convexity and dividing both sides of above inequality by T, the proof of the nonsmooth case is complete.

When function $f(x) \in C^{1,1}$, it is straightforward to see that we also have the inequality (32). In addition, according to Lemma 2.2, we can replace $f_\delta(x)$ with $f(x)$ in above inequality and get

$$\mathbb{E}[\|x_{t+1} - x^*\|^2] \leq \mathbb{E}[\|x_t - x^*\|^2] - 2\eta(f(x_t) - f(x^*)) + \eta^2\mathbb{E}[\|\tilde{g}(x_t)\|^2] + 4L_1 d\delta^2\eta. \tag{37}$$

Similarly to the above analysis, we telescope the above inequality from $t = 0$ to $T - 1$, apply the bound on $\sum_{t=0}^{T-1} \mathbb{E}[\|\tilde{g}(x_t)\|^2]$ in (27) and obtain that

$$\sum_{t=0}^{T-1} \mathbb{E}[f(x_t)] - Tf(x^*) \le \frac{1}{2\eta}\|x_0 - x^*\|^2 + \frac{\eta}{2(1-\alpha)}\mathbb{E}[\|\tilde{g}(x_0)\|^2] + \frac{2\eta}{1-\alpha}L_1^2(d+6)^3\delta^2 T$$
$$+ \frac{4\eta}{1-\alpha}(d+4)^2 \sum_{t=0}^{T-1} \mathbb{E}[\|\nabla f(x_t)\|^2] + 2L_1 d\delta^2 T. \tag{38}$$

Since $f(x) \in C^{1,1}$ is convex, we have that $\|\nabla f(x_t)\|^2 \le 2L_1(f(x_t) - f(x^*))$ according to (2.1.7) in Nesterov (2013). Applying this bound into the above inequality, we get that

$$\sum_{t=0}^{T-1} \mathbb{E}[f(x_t)] - Tf(x^*) \le \frac{1}{2\eta}\|x_0 - x^*\|^2 + \frac{\eta}{2(1-\alpha)}\mathbb{E}[\|\tilde{g}(x_0)\|^2] + \frac{2\eta}{1-\alpha}L_1^2(d+6)^3\delta^2 T$$
$$+ \frac{8\eta}{1-\alpha}L_1(d+4)^2\big(\sum_{t=0}^{T-1} \mathbb{E}[f(x_t)] - Tf(x^*)\big) + 2L_1 d\delta^2 T. \tag{39}$$

Let $\eta = \frac{1}{2\tilde{L}(d+4)^2 T^{\frac{1}{3}}}$ and $\delta = \frac{\sqrt{d}}{T^{\frac{1}{3}}}$ where $\tilde{L} = \max\{L_0, 16L_1\}$. Then, we have that $\alpha = 2dL_0^2\frac{\eta^2}{\delta^2} \le \frac{1}{2(d+4)^4} \le \frac{1}{2}$. In addition, we have that $\frac{8\eta}{1-\alpha}L_1(d+4)^2 \le \frac{1}{2T^{\frac{1}{3}}} \le \frac{1}{2}$. Applying these two bounds into above inequality and rearranging terms, we have that

$$\frac{1}{2}\sum_{t=0}^{T-1} \mathbb{E}[f(x_t)] - Tf(x^*) \le \tilde{L}\|x_0 - x^*\|^2(d+4)^2 T^{\frac{1}{3}} + \frac{1}{2\tilde{L}(d+4)^2 T^{\frac{1}{3}}}\mathbb{E}[\|\tilde{g}(x_0)\|^2]$$
$$+ \frac{L_1}{8}\frac{(d+6)^3 d}{(d+4)^2} + 2L_1 d^2 T^{\frac{1}{3}}. \tag{40}$$

Recalling that $f(\bar{x}) \le \frac{1}{T}\sum_{t=0}^{T-1} f(x_t)$ due to convexity and dividing both sides of above inequality by T, the proof of the smooth case is complete.

F Proof of Lemma 4.3

The analysis is similar to the proof in Appendix B. First, consider the case when $F(x,\xi) \in C^{0,0}$ with $L_0(\xi)$. According to (6), we have that

$$\begin{aligned}
\mathbb{E}[\|\tilde{g}(x_t)\|^2] &= \mathbb{E}[\frac{1}{\delta^2}\big(F(x_t + \delta u_t, \xi_t) - F(x_{t-1} + \delta u_{t-1}, \xi_{t-1})\big)^2\|u_t\|^2] \\
&\le \frac{2}{\delta^2}\mathbb{E}[\big(F(x_t + \delta u_t, \xi_t) - F(x_{t-1} + \delta u_{t-1}, \xi_t)\big)^2\|u_t\|^2] \\
&\quad + \frac{2}{\delta^2}\mathbb{E}[\big(F(x_{t-1} + \delta u_{t-1}, \xi_t) - F(x_{t-1} + \delta u_{t-1}, \xi_{t-1})\big)^2\|u_t\|^2].
\end{aligned} \tag{41}$$

Using the bound in Assumption 4.1, we get that $\frac{2}{\delta^2}\mathbb{E}[\big(F(x_{t-1} + \delta u_{t-1}, \xi_t) - F(x_{t-1} + \delta u_{t-1}, \xi_{t-1})\big)^2\|u_t\|^2] \le \frac{8d\sigma^2}{\delta^2}$. In addition, adding and subtracting $F(x_{t-1} + \delta u_t, \xi_t)$ in $\big(F(x_t + \delta u_t, \xi_t) - F(x_{t-1} + \delta u_{t-1}, \xi_t)\big)^2$ in above inequality, we obtain that

$$\begin{aligned}
\mathbb{E}[\|\tilde{g}(x_t)\|^2] \le& \frac{4}{\delta^2}\mathbb{E}[\big(F(x_t + \delta u_t, \xi_t) - F(x_{t-1} + \delta u_t, \xi_t)\big)^2\|u_t\|^2] \\
&+ \frac{4}{\delta^2}\mathbb{E}[\big(F(x_{t-1} + \delta u_t, \xi_t) - F(x_{t-1} + \delta u_{t-1}, \xi_t)\big)^2\|u_t\|^2] + \frac{8d\sigma^2}{\delta^2}.
\end{aligned} \tag{42}$$

Using Assumption 4.2, we can bound the first two items on the right hand side of above inequality following the same procedure after inequality (9) and get that

$$\mathbb{E}[\|\tilde{g}(x_t)\|^2] \le \frac{4dL_0^2\eta^2}{\delta^2}\mathbb{E}[\|\tilde{g}(x_{t-1})\|^2] + 16L_0^2(d+4)^2 + \frac{8d\sigma^2}{\delta^2}. \tag{43}$$

The proof is complete.

G Proof of Theorem 4.4

When function $F(x) \in C^{0,0}$ with $L_0(\xi)$, using Assumption 4.2 and following the same procedure in Appendix C, we have that

$$\sum_{t=0}^{T-1} \mathbb{E}[\|\nabla f_\delta(x_t)\|^2] \leq \frac{\mathbb{E}[f_\delta(x_0)] - f_\delta^*}{\eta} + \frac{L_1(f_\delta)\eta}{2} \sum_{t=0}^{T-1} \mathbb{E}[\|\tilde{g}(x_t)\|^2], \tag{44}$$

where $L_1(f_\delta) = \frac{\sqrt{d}}{\delta} L_0$. In addition, according to Lemma 4.3, we get that

$$\sum_{t=0}^{T-1} \mathbb{E}[\|\tilde{g}(x_t)\|^2] \leq \frac{1}{1-\alpha} \mathbb{E}[\|\tilde{g}(x_0)\|^2] + \frac{16 L_0^2}{1-\alpha}(d+4)^2 T + \frac{8\sigma^2}{1-\alpha} \frac{d}{\delta^2} T, \tag{45}$$

where $\alpha = \frac{4dL_0^2\eta^2}{\delta^2}$. Plugging (45) into the bound in (44), we obtain that

$$\begin{aligned} \sum_{t=0}^{T-1} \mathbb{E}[\|\nabla f_\delta(x_t)\|^2] \leq & \frac{\mathbb{E}[f_\delta(x_0)] - f_\delta^*}{\eta} + \frac{\sqrt{d} L_0}{2(1-\alpha)} \mathbb{E}[\|\tilde{g}(x_0)\|^2] \frac{\eta}{\delta} \\ & + \frac{8 L_0^3 \sqrt{d}}{1-\alpha}(d+4)^2 \frac{\eta}{\delta} T + \frac{4\sigma^2 L_0}{1-\alpha} d^{1.5} \frac{\eta}{\delta^3} T. \end{aligned} \tag{46}$$

Similar to Appendix C, to fullfill the requirement that $|f(x) - f_\delta(x)| \leq \epsilon_f$, we set the exporation parameter $\delta = \frac{\epsilon_f}{d^{\frac{1}{2}} L_0}$. In addition, let the stepsize be $\eta = \frac{\epsilon_f^{1.5}}{2\sqrt{2} L_0^2 d^{1.5} T^{\frac{1}{2}}}$. Then, we have that $\alpha = \frac{4dL_0^2\eta^2}{\delta^2} = \frac{\epsilon_f}{2dT} \leq \frac{1}{2}$ when $T \geq \frac{1}{d\epsilon_f}$. Therefore, we have that $\frac{1}{1-\alpha} \leq 2$. Applying this bound and the choices of η and δ into the bound (46), we obtain that

$$\begin{aligned} \sum_{t=0}^{T-1} \mathbb{E}[\|\nabla f_\delta(x_t)\|^2] \leq & 2\sqrt{2} L_0^2 (\mathbb{E}[f_\delta(x_0)] - f_\delta^*) \frac{d^{1.5}\sqrt{T}}{\epsilon_f^{1.5}} + \frac{L_0 \epsilon_f^{0.5}}{2\sqrt{2dT}} \mathbb{E}[\|\tilde{g}(x_0)\|^2] \\ & + 4\sqrt{2} L_0^2 \frac{(d+4)^2}{\sqrt{d}} \sqrt{\epsilon_f T} + 2\sqrt{2}\sigma^2 L_0^2 \frac{d^{1.5}\sqrt{T}}{\epsilon_f^{1.5}}. \end{aligned} \tag{47}$$

Dividing both sides by T, the proof for the nonsmooth case is complete.

When function $F(x, \xi) \in C^{1,1}$ with $L_1(\xi)$, according to Assumption 4.2, we also have that $f_\delta(x), f(x) \in C^{1,1}$ with constant L_1. Similarly to the proof in Appendix D, we get that

$$\frac{1}{2} \sum_{t=0}^{T-1} \|\mathbb{E}[\|\nabla f(x)\|^2\|] \leq \frac{\mathbb{E}[f_\delta(x_0)] - f_\delta^*}{\eta} + \frac{L_1\eta}{2} \sum_{t=0}^{T-1} \mathbb{E}[\|\tilde{g}(x_t)\|^2] + L_1^2 (d+3)^3 \delta^2 T. \tag{48}$$

Plugging inequality (45) into the above upper bound, we obtain that

$$\begin{aligned} \frac{1}{2} \sum_{t=0}^{T-1} \|\mathbb{E}[\|\nabla f(x)\|^2\|] \leq & \frac{\mathbb{E}[f_\delta(x_0)] - f_\delta^*}{\eta} + \frac{L_1\eta}{2(1-\alpha)} \mathbb{E}[\|\tilde{g}(x_0)\|^2] + \frac{8 L_0^2 L_1}{1-\alpha}(d+4)^2 \eta T \\ & + \frac{4 L_1 \sigma^2}{1-\alpha} \frac{d\eta}{\delta^2} T + L_1^2 (d+3)^3 \delta^2 T. \end{aligned} \tag{49}$$

Let $\eta = \frac{1}{2\sqrt{2} L_0 d^{\frac{4}{3}} T^{\frac{2}{3}}}$ and $\delta = \frac{1}{d^{\frac{5}{6}} T^{\frac{1}{6}}}$. Then, $\alpha = \frac{4dL_0^2\eta^2}{\delta^2} = \frac{1}{2T} \leq \frac{1}{2}$ and $\frac{1}{1-\alpha} \leq 2$. Plugging these results into the above inequality, we get that

$$\begin{aligned} \frac{1}{2} \sum_{t=0}^{T-1} \|\mathbb{E}[\|\nabla f(x)\|^2\|] \leq & 2\sqrt{2} L_0 (\mathbb{E}[f_\delta(x_0)] - f_\delta^*) d^{\frac{4}{3}} T^{\frac{2}{3}} + \frac{L_1}{2\sqrt{2} L_0 d^{\frac{4}{3}} T^{\frac{2}{3}}} \mathbb{E}[\|\tilde{g}(x_0)\|^2] \\ & + 4\sqrt{2} L_0 L_1 \frac{(d+4)^2}{d^{\frac{4}{3}}} T^{\frac{1}{3}} + \frac{2\sqrt{2} L_1 \sigma^2}{L_0 d^{\frac{1}{3}}} T^{\frac{1}{3}} + L_1^2 \frac{(d+3)^3}{d^{\frac{5}{3}}} T^{\frac{2}{3}}. \end{aligned} \tag{50}$$

Dividing both sides by T, the proof for the smooth case is complete.

H Proof of Theorem 4.5

When the function $f(x) \in C^{0,0}$ with constant $L_0(\xi)$ is convex, we can follow the same procedure as in Appendix E and get that

$$\sum_{t=0}^{T-1} \mathbb{E}[f(x_t)] - Tf(x^*) \leq \frac{1}{2\eta}\|x_0 - x^*\|^2 + \frac{\eta}{2}\sum_{t=0}^{T-1}\mathbb{E}[\|\tilde{g}(x_t)\|^2] + 2L_0\sqrt{d}\delta T. \tag{51}$$

Plugging the bound (45) into above inequality, we have that

$$\begin{aligned}\sum_{t=0}^{T-1} \mathbb{E}[f(x_t)] - Tf(x^*) \leq \frac{1}{2\eta}\|x_0 - x^*\|^2 + \frac{\eta}{2(1-\alpha)}\mathbb{E}[\|\tilde{g}(x_0)\|^2] + \frac{8L_0^2}{1-\alpha}(d+4)^2\eta T \\ + \frac{4\sigma^2}{1-\alpha}\frac{d\eta}{\delta^2}T + 2L_0\sqrt{d}\delta T.\end{aligned} \tag{52}$$

Let $\eta = \frac{1}{2\sqrt{2}L_0\sqrt{d}T^{\frac{3}{4}}}$ and $\delta = \frac{1}{T^{\frac{1}{4}}}$. Then, we have that $\alpha = \frac{4dL_0^2\eta^2}{\delta^2} = \frac{1}{2T} \leq \frac{1}{2}$. Plugging these results into the above inequality, we get that

$$\begin{aligned}\sum_{t=0}^{T-1} \mathbb{E}[f(x_t)] - Tf(x^*) \leq & \sqrt{2}L_0\|x_0 - x^*\|^2\sqrt{d}T^{\frac{3}{4}} + \frac{1}{2\sqrt{2}L_0\sqrt{d}T^{\frac{3}{4}}}\mathbb{E}[\|\tilde{g}(x_0)\|^2] \\ & + 4\sqrt{2}L_0\frac{(d+4)^2}{\sqrt{d}}T^{\frac{1}{4}} + \frac{2\sqrt{2}\sigma^2}{L_0}\sqrt{d}T^{\frac{3}{4}} + 2L_0\sqrt{d}T^{\frac{3}{4}}.\end{aligned} \tag{53}$$

Dividing both sides by T, the proof for the nonsmooth case is complete.

When the function $f(x) \in C^{1,1}$ with constant $L_1(\xi)$, we can also get the inequality (37) in Appendix E. Telescoping this inequality from $t = 0$ to $T - 1$ and rearranging terms, we obtain

$$\sum_{t=0}^{T-1} \mathbb{E}[f(x_t)] - Tf(x^*) \leq \frac{1}{2\eta}\|x_0 - x^*\|^2 + \frac{\eta}{2}\sum_{t=0}^{T-1}\mathbb{E}[\|\tilde{g}(x_t)\|^2] + 2L_1 d\delta^2 T. \tag{54}$$

Plugging the bound (45) into above inequality, we have that

$$\begin{aligned}\sum_{t=0}^{T-1} \mathbb{E}[f(x_t)] - Tf(x^*) \leq \frac{1}{2\eta}\|x_0 - x^*\|^2 + \frac{\eta}{2(1-\alpha)}\mathbb{E}[\|\tilde{g}(x_0)\|^2] + \frac{8L_0^2}{1-\alpha}(d+4)^2\eta T \\ + \frac{4\sigma^2}{1-\alpha}\frac{d\eta}{\delta^2}T + 2L_1 d\delta^2 T.\end{aligned} \tag{55}$$

Let $\eta = \frac{1}{2\sqrt{2}L_0 d^{\frac{2}{3}}T^{\frac{2}{3}}}$ and $\delta = \frac{1}{d^{\frac{1}{6}}T^{\frac{1}{6}}}$. Then, we have that $\alpha = \frac{4dL_0^2\eta^2}{\delta^2} = \frac{1}{2T} \leq \frac{1}{2}$. Plugging these parameters into above inequality, we get that

$$\begin{aligned}\sum_{t=0}^{T-1} \mathbb{E}[f(x_t)] - Tf(x^*) \leq & \sqrt{2}L_0\|x_0 - x^*\|^2 d^{\frac{2}{3}}T^{\frac{2}{3}} + \frac{1}{2\sqrt{2}L_0 d^{\frac{2}{3}}T^{\frac{2}{3}}}\mathbb{E}[\|\tilde{g}(x_0)\|^2] \\ & + 4\sqrt{2}L_0\frac{(d+4)^2}{d^{\frac{2}{3}}}T^{\frac{1}{3}} + \frac{2\sqrt{2}\sigma^2}{L_0}d^{\frac{2}{3}}T^{\frac{2}{3}} + 2L_1 d^{\frac{2}{3}}T^{\frac{2}{3}}.\end{aligned} \tag{56}$$

Dividing both sides by T, the proof for the smooth case is complete.

I Analysis of SGD with Mini-batch Residual Feedback

In this section, we analyze the query complexity of SGD with the mini-batch residual feedback. First, we make some additional assumptions.

Assumption I.1. *When function $F(x, \xi) \in C^{1,1}$, we assume that*

$$\mathbb{E}_\xi[\|\nabla F(x,\xi) - \mathbb{E}[\nabla F(x,\xi)]\|^2] \leq \sigma_g^2.$$

Before presenting the main results, we first establish some important lemmas. The following lemma provides a characterization for the estimation variance of the estimator $\widetilde{g}_b(x_t)$.

Lemma I.2. *When function $F(x,\xi) \in C^{0,0}$ with constant $L_0(\xi)$, given Assumptions 4.1 and 4.2, we have that*

$$\mathbb{E}\|\widetilde{g}_b(x_t)\|^2 \leq \frac{4(d+2)L_0^2}{\delta^2}\mathbb{E}\|x_t - x_{t-1}\|^2 + 16L_0^2(d+4)^2 + \frac{8(d+2)\sigma^2}{\delta^2 b}. \tag{57}$$

Furthermore, when function $F(x,\xi) \in C^{1,1}$ with constant $L_1(\xi)$, given Assumptions 4.1, 4.2 and I.1, we have that

$$\begin{aligned}\mathbb{E}\|\widetilde{g}_b(x_t)\|^2 \leq & 12L_1^2\delta^2(d+6)^3 + \frac{6(d+2)L_0^2\eta^2}{\delta^2}\mathbb{E}\|\widetilde{g}_b(x_{t-1})\|^2 \\ & + 24(d+4)\mathbb{E}(\|\nabla f(x_t)\|^2 + \|\nabla f(x_{t-1})\|^2) + \frac{48(d+4)\sigma_g^2}{b} + \frac{8(d+2)\sigma^2}{\delta^2 b}.\end{aligned}$$

Proof. When function $F(x,\xi) \in C^{0,0}$, based on the definition of $\widetilde{g}(x_t)$, we have

$$\begin{aligned}\|\widetilde{g}_b(x_t)\|^2 = & \frac{1}{\delta^2 b^2}|F(x_t+\delta u_t,\xi_{1:b}) - F(x_{t-1}+\delta u_{t-1},\xi_{1:b}) \\ & \quad + F(x_{t-1}+\delta u_{t-1};\xi_{1:b}) - F(x_{t-1}+\delta u_{t-1},\xi'_{1:b})|^2\|u_t\|^2 \\ \leq & \frac{2}{\delta^2 b^2}\big(|F(x_t+\delta u_t,\xi_{1:b}) - F(x_{t-1}+\delta u_{t-1},\xi_{1:b})|^2 + \\ & \quad |F(x_{t-1}+\delta u_{t-1},\xi_{1:b}) - F(x_{t-1}+\delta u_{t-1},\xi'_{1:b})|^2\big)\|u_t\|^2 \\ \leq & \frac{4L_0^2}{\delta^2}\|x_t - x_{t-1}\|^2\|u_t\|^2 + 4L_0^2\|u_t - u_{t-1}\|^2\|u_t\|^2 \\ & + \frac{2}{\delta^2 b^2}|F(x_{t-1}+\delta u_{t-1},\xi_{1:b}) - F(x_{t-1}+\delta u_{t-1},\xi'_{1:b})|^2\|u_t\|^2.\end{aligned}$$

Taking expectation over the above inequality yields

$$\begin{aligned}\mathbb{E}\|\widetilde{g}_b(x_t)\|^2 \leq & \frac{4L_0^2}{\delta^2}\mathbb{E}\big(\|x_t - x_{t-1}\|^2\|u_t\|^2\big) + 4L_0^2\mathbb{E}\big(\|u_t - u_{t-1}\|^2\|u_t\|^2\big) \\ & + \frac{2}{\delta^2 b^2}\mathbb{E}\big(|F(x_{t-1}+\delta u_{t-1},\xi_{1:b}) - F(x_{t-1}+\delta u_{t-1},\xi'_{1:b})|^2\|u_t\|^2\big) \\ \leq & \frac{4L_0^2}{\delta^2}\mathbb{E}\big(\|x_t - x_{t-1}\|^2\mathbb{E}_{u_t}\|u_t\|^2\big) + 8L_0^2\mathbb{E}\big(\|u_t\|^4 + \|u_{t-1}\|^2\|u_t\|^2\big) \\ & + \frac{2}{\delta^2}\mathbb{E}\big(|F(x_{t-1}+\delta u_{t-1},\xi_{1:b}) - F(x_{t-1}+\delta u_{t-1},\xi'_{1:b})|^2\|u_t\|^2\big) \\ \overset{(i)}{\leq} & \frac{4(d+2)L_0^2}{\delta^2}\mathbb{E}\|x_t - x_{t-1}\|^2 + 8L_0^2\big((d+4)^2 + (d+2)^2\big) \\ & + \frac{2}{\delta^2 b^2}\mathbb{E}\big(|F(x_{t-1}+\delta u_{t-1},\xi_{1:b}) - F(x_{t-1}+\delta u_{t-1},\xi'_{1:b})|^2\|u_t\|^2\big) \\ \leq & \frac{4(d+2)L_0^2}{\delta^2}\mathbb{E}\|x_t - x_{t-1}\|^2 + \frac{4}{\delta^2 b^2}\underbrace{\mathbb{E}|F(x_{t-1}+\delta u_{t-1},\xi_{1:b}) - bf(x_{t-1}+\delta u_{t-1})|^2\|u_t\|^2}_{(P)} \\ & + \frac{4}{\delta^2 b^2}\underbrace{\mathbb{E}|F(x_{t-1}+\delta u_{t-1},\xi'_{1:b}) - bf(x_{t-1}+\delta u_{t-1})|^2\|u_t\|^2}_{(Q)} + 16L_0^2(d+4)^2,\end{aligned} \tag{58}$$

where (i) follows from Lemma 1 in Nesterov & Spokoiny (2017) that $\mathbb{E}\|u\|^p \leq (d+p)^{p/2}$ for a d-dimensional standard Gaussian random vector. Our next step is to upper-bound (P) and (Q) in the above inequality. For (P), conditioning on x_{t-1} and u_{t-1} and noting that $\xi_{1:b}$ is independent of u_t, we have

$$
\begin{aligned}
(P) &= \mathbb{E}_{\xi_{1:b}}\Big|\sum_{\xi\in\xi_{1:b}}\big(F(x_{t-1}+\delta u_{t-1},\xi)-f(x_{t-1}+\delta u_{t-1})\big)\Big|^2\mathbb{E}_{u_t}\|u_t\|^2\\
&\le (d+2)\mathbb{E}_{\xi_{1:b}}\Big|\sum_{\xi\in\xi_{1:b}}\big(F(x_{t-1}+\delta u_{t-1},\xi)-f(x_{t-1}+\delta u_{t-1})\big)\Big|^2\\
&= b(d+2)\mathbb{E}_{\xi}|F(x_{t-1}+\delta u_{t-1},\xi)-f(x_{t-1}+\delta u_{t-1})|^2+(d+2)\sum_{i\neq j,\xi_i,\xi_j\in\xi_{1:b}}\\
&\quad\langle\mathbb{E}_{\xi_i}F(x_{t-1}+\delta u_{t-1},\xi_i)-f(x_{t-1}+\delta u_{t-1}),\mathbb{E}_{\xi_j}F(x_{t-1}+\delta u_{t-1},\xi_j)-f(x_{t-1}+\delta u_{t-1})\rangle\\
&= b(d+2)\mathbb{E}_{\xi}\|F(x_{t-1}+\delta u_{t-1},\xi)-f(x_{t-1}+\delta u_{t-1})\|^2\le b(d+2)\sigma^2.
\end{aligned}\tag{59}
$$

Unconditioning on x_{t-1} and u_{t-1} in the above equality yields $(P)\le b(d+2)\sigma^2$. For Q, we have

$$
\begin{aligned}
(Q) &= \mathbb{E}\big(|F(x_{t-1}+\delta u_{t-1},\xi'_{1:b})-bf(x_{t-1}+\delta u_{t-1})|^2\mathbb{E}_{u_t}\|u_t\|^2\,\big|x_{t-1},\xi'_{1:b},u_{t-1}\big)\\
&\le (d+2)\mathbb{E}|F(x_{t-1}+\delta u_{t-1},\xi'_{1:b})-bf(x_{t-1}+\delta u_{t-1})|^2,
\end{aligned}
$$

which, using an approach similar to the steps in (59), yields

$$
(Q)\le b(d+2)\sigma^2. \tag{60}
$$

Combining (58), (59) and (60) yields the proof when $F(x,\xi)\in C^{0,0}$.

When function $F(x,\xi)\in C^{1,1}$, based on the definition of $\widetilde{g}_b(x_t)$, we have

$$
\begin{aligned}
\|\widetilde{g}_b(x_t)\|^2 &= \frac{1}{\delta^2b^2}|F(x_t+\delta u_t,\xi_{1:b})-F(x_{t-1}+\delta u_{t-1},\xi_{1:b})\\
&\qquad\qquad+F(x_{t-1}+\delta u_{t-1},\xi_{1:b})-F(x_{t-1}+\delta u_{t-1},\xi'_{1:b})|^2\|u_t\|^2\\
&\le \frac{2}{\delta^2b^2}|F(x_t+\delta u_t,\xi_{1:b})-F(x_{t-1}+\delta u_{t-1},\xi_{1:b})|^2\|u_t\|^2\\
&\qquad\qquad+\frac{2}{\delta^2b^2}|F(x_{t-1}+\delta u_{t-1},\xi_{1:b})-F(x_{t-1}+\delta u_{t-1},\xi'_{1:b})|^2\|u_t\|^2,
\end{aligned}
$$

which, taking expectation and using an approach similar to (58), yields

$$
\mathbb{E}\|\widetilde{g}(x_t)\|^2\le\frac{2}{\delta^2b^2}\underbrace{\mathbb{E}|F(x_t+\delta u_t,\xi_{1:b})-F(x_{t-1}+\delta u_{t-1},\xi_{1:b})|^2\|u_t\|^2}_{(P)}+\frac{8(d+2)\sigma^2}{\delta^2b}. \tag{61}
$$

Our next step is to upper-bound (P) in the above inequality. We first divide (P) into the following three parts.

$$
\begin{aligned}
(P)\le 3\mathbb{E}\Big[&\underbrace{|F(x_t+\delta u_t,\xi_{1:b})-F(x_t,\xi_{1:b})-\langle\delta u_t,\nabla F(x_t,\xi_{1:b})\rangle+\langle\delta u_t,\nabla F(x_t,\xi_{1:b})\rangle|^2\|u_t\|^2}_{(P_1)}\\
&+\underbrace{|F(x_t,\xi_{1:b})-F(x_{t-1},\xi_{1:b})|^2\|u_t\|^2}_{(P_2)}\\
&+\underbrace{|F(x_{t-1},\xi_{1:b})-F(x_{t-1}+\delta u_{t-1},\xi_{1:b})+\langle\delta u_{t-1},\nabla F(x_{t-1},\xi_{1:b})\rangle-\langle\delta u_{t-1},\nabla F(x_{t-1},\xi_{1:b})\rangle|^2\|u_t\|^2}_{(P_3)}\Big].
\end{aligned}\tag{62}
$$

Using the assumption that $F(x;\xi)\in C^{0,0}\cap C^{1,1}$, we have

$$
\begin{aligned}
&P_1\le b^2L_1^2\delta^4\|u_t\|^6+2\delta^2|\langle u_t,\nabla F(x_t,\xi_{1:b})\rangle|^2\|u_t\|^2,\\
&P_2\le b^2L_0^2\|x_t-x_{t-1}\|^2\|u_t\|^2,\\
&P_3\le b^2L_1^2\delta^4\|u_{t-1}\|^4\|u_t\|^2+2\delta^2|\langle u_{t-1},\nabla F(x_{t-1},\xi_{1:b})\rangle|^2\|u_t\|^2.
\end{aligned}
$$

Plugging the above inequalities into (62), we have

$$
\begin{aligned}
(P)\le & 3b^2L_1^2\delta^4\mathbb{E}\|u_t\|^6+6\delta^2\mathbb{E}|\langle u_t,\nabla F(x_t,\xi_{1:b})\rangle|^2\|u_t\|^2+3b^2L_0^2\mathbb{E}\|x_t-x_{t-1}\|^2\|u_t\|^2\\
&+3b^2L_1^2\delta^4\mathbb{E}\|u_{t-1}\|^4\|u_t\|^2+6\delta^2\mathbb{E}|\langle u_{t-1},\nabla F(x_{t-1},\xi_{1:b})\rangle|^2\|u_t\|^2,
\end{aligned}\tag{63}
$$

Based on the results in Nesterov & Spokoiny (2017), we have $\mathbb{E}_u[\|u\|^p] \leq (d+p)^{p/2}$, $\mathbb{E}[\langle u_t, \nabla F(x_t,\xi_{1:b})\rangle^2 \|u_t\|^2] \leq (d+4)\|\nabla F(x_t,\xi_{1:b})\|^2$, $\mathbb{E}[\langle u_{t-1}, \nabla F(x_{t-1},\xi_{1:b})\rangle^2] \leq \|\nabla F(x_t,\xi_{1:b})\|^2$, which, in conjunction with (63), yields

$$\begin{aligned}
(P) \leq\, & 6b^2L_1^2\delta^4(d+6)^3 + 3b^2(d+2)L_0^2\mathbb{E}\|x_t - x_{t-1}\|^2 \\
& + 6(d+4)\delta^2\mathbb{E}\|\nabla F(x_t,\xi_{1:b})\|^2 + 6(d+2)\delta^2\mathbb{E}\|\nabla F(x_{t-1},\xi_{1:b})\|^2 \\
\overset{(i)}{\leq}\, & 6b^2L_1^2\delta^4(d+6)^3 + 3b^2(d+2)L_0^2\mathbb{E}\|x_t - x_{t-1}\|^2 + 12b^2(d+4)\delta^2\mathbb{E}\|\nabla f(x_t)\|^2 \\
& + 12b(d+4)\delta^2\sigma_g^2 + 12b^2(d+2)\delta^2\mathbb{E}\|\nabla f(x_{t-1})\|^2 + 12b(d+2)\delta^2\sigma_g^2 \\
\leq\, & 6b^2L_1^2\delta^4(d+6)^3 + 3b^2(d+2)L_0^2\mathbb{E}\|x_t - x_{t-1}\|^2 \\
& + 12b^2(d+4)\delta^2\mathbb{E}\|\nabla f(x_t)\|^2 + 12b^2(d+2)\delta^2\mathbb{E}\|\nabla f(x_{t-1})\|^2 + 24b(d+4)\delta^2\sigma_g^2. \quad (64)
\end{aligned}$$

Combining (64) and (61) yields

$$\begin{aligned}
\mathbb{E}\|\widetilde{g}_b(x_t)\|^2 \leq & 12L_1^2\delta^2(d+6)^3 + \frac{6(d+2)L_0^2\eta^2}{\delta^2}\mathbb{E}\|\widetilde{g}_b(x_{t-1})\|^2 \\
& + 24(d+4)\mathbb{E}(\|\nabla f(x_t)\|^2 + \|\nabla f(x_{t-1})\|^2) + \frac{48(d+4)\sigma_g^2}{b} + \frac{8(d+2)\sigma^2}{\delta^2 b}, \quad (65)
\end{aligned}$$

which finishes the proof. ☐

First, we analyze the convergence when the problem is non-smooth. Based on Lemma I.2, we provide an upper bound on $\mathbb{E}\|x_{t+1} - x_t\|^2$.

Lemma I.3. *Suppose Assumptions 4.1 and 4.2 are satisfied. Then, we have*

$$\mathbb{E}\|x_{t+1} - x_t\|^2 \leq \beta_1^t\Big(\mathbb{E}\|x_1 - x_0\|^2 - \frac{\beta_2}{1-\beta_1}\Big) + \frac{\beta_2}{1-\beta_1}, \quad (66)$$

where $\beta_1 = \frac{4\eta^2(d+2)L_0^2}{\delta^2}$ *and* $\beta_2 = 16\eta^2L_0^2(d+4)^2 + \frac{8\eta^2(d+2)\sigma^2}{\delta^2 b}$.

Proof. Based on the update that $x_{t+1} - x_t = -\eta\widetilde{g}_b(x_t)$ and Lemma I.2, we have

$$\begin{aligned}
\mathbb{E}\|x_{t+1} - x_t\|^2 = & \eta^2\|\widetilde{g}_b(x_t)\|^2 \leq \eta^2\Big(\frac{4(d+2)L_0^2}{\delta^2}\mathbb{E}\|x_t - x_{t-1}\|^2 + 16L_0^2(d+4)^2 + \frac{8(d+2)\sigma^2}{\delta^2 b}\Big) \\
= & \frac{4\eta^2(d+2)L_0^2}{\delta^2}\mathbb{E}\|x_t - x_{t-1}\|^2 + 16\eta^2L_0^2(d+4)^2 + \frac{8\eta^2(d+2)\sigma^2}{\delta^2 b}. \\
= & \beta_1\mathbb{E}\|x_t - x_{t-1}\|^2 + \beta_2.
\end{aligned}$$

Then, telescoping the above inequality yields the proof. ☐

Nonsmooth Nonconvex Geometry

Based on the above lemmas, we next provide the convergence and complexity analysis for our proposed algorithm for the case where $F(x;\xi)$ is nonconvex and belongs to $C^{0,0}$.

Theorem I.4. *Suppose Assumptions 4.1 and 4.2 are satisfied. Choose* $\eta = \frac{\epsilon_f^{1/2}}{2(d+2)^{3/2}T^{1/2}L_0^2}$, $\delta = \frac{\epsilon_f}{(d+2)^{1/2}L_0}$ *and* $b = \frac{\sigma^2}{\epsilon_f^2} \geq 1$ *for certain* $\epsilon_f < 1$. *Then, we have* $\mathbb{E}\|\nabla f_\delta(x_\zeta)\|^2 \leq \mathcal{O}\left(\frac{d^{3/2}}{\epsilon_f^{1/2}\sqrt{T}}\right)$ *with the approximation error* $|f_\delta(x_\zeta) - f(x_\zeta)| < \theta$, *where* ζ *is uniformly sampled from* $\{0,1,\ldots,T-1\}$. *Then, to achieve an* ϵ*-accurate stationary point of* f_δ, *the corresponding total function query complexity is given by*

$$Tb = \mathcal{O}\left(\frac{\sigma^2 d^3}{\epsilon_f^3\epsilon^2}\right). \quad (67)$$

Proof. Recall that $f_\delta(x) = \mathbb{E}_u f(x+\delta u)$ is a smoothed approximation of $f(x)$, where $u \in \mathbb{R}^d$ is a standard Gaussian random vector. Based on Lemma 2 in Nesterov & Spokoiny (2017), we have $f_\delta \in C^{1,1}$ with gradient-Lipschitz constant L_δ satisfying $L_\delta \le \frac{d^{1/2}}{\delta} L_0$, and thus

$$\begin{aligned} f_\delta(x_{t+1}) \le & f_\delta(x_t) + \langle \nabla f_\delta(x_t), x_{t+1} - x_t \rangle + \frac{L_\delta}{2} \|x_{t+1} - x_t\|^2 \\ \le & f_\delta(x_t) + \langle \nabla f_\delta(x_t), x_{t+1} - x_t \rangle + \frac{d^{1/2} L_0}{2\delta} \|x_{t+1} - x_t\|^2 \\ = & f_\delta(x_t) - \eta \langle \nabla f_\delta(x_t), \widetilde{g}(x_t) \rangle + \frac{d^{1/2} L_0}{2\delta} \|x_{t+1} - x_t\|^2. \end{aligned}$$

Taking expectation over the above inequality and using $\mathbb{E}(\widetilde{g}(x_t)|x_t) = \nabla f_\delta(x_t)$, we have

$$\mathbb{E} f_\delta(x_{t+1}) \le \mathbb{E} f_\delta(x_t) - \eta \mathbb{E} \|\nabla f_\delta(x_t)\|^2 + \frac{d^{1/2} L_0}{2\delta} \mathbb{E} \|x_{t+1} - x_t\|^2,$$

which, in conjunction with Lemma I.3, yields

$$\mathbb{E} f_\delta(x_{t+1}) \le \mathbb{E} f_\delta(x_t) - \eta \mathbb{E} \|\nabla f_\delta(x_t)\|^2 + \frac{d^{1/2} L_0}{2\delta} \beta_1^t \Big(\mathbb{E} \|x_1 - x_0\|^2 - \frac{\beta_2}{1-\beta_1} \Big) + \frac{d^{1/2} L_0}{2\delta} \frac{\beta_2}{1-\beta_1}. \tag{68}$$

Telescoping the above inequality over t from 0 to $T-1$ yields

$$\begin{aligned} & \sum_{t=0}^{T-1} \eta \mathbb{E} \|\nabla f_\delta(x_t)\|^2 \\ \le & \, f_\delta(x_0) - \inf_x f_\delta(x) + \frac{d^{1/2} L_0 T}{2\delta} \frac{\beta_2}{1-\beta_1} + \frac{d^{1/2} L_0}{2\delta} \Big(\mathbb{E} \|x_1 - x_0\|^2 - \frac{\beta_2}{1-\beta_1} \Big) \sum_{t=0}^{T-1} \beta_1^t \\ = & \, f_\delta(x_0) - \inf_x f_\delta(x) + \frac{d^{1/2} L_0 T}{2\delta} \frac{\beta_2}{1-\beta_1} + \frac{d^{1/2} L_0}{2\delta} \Big(\mathbb{E} \|x_1 - x_0\|^2 - \frac{\beta_2}{1-\beta_1} \Big) \frac{1-\beta_1^T}{1-\beta_1} \\ \le & \, f(x_0) - \inf_x f(x) + 2\delta L_0 d^{1/2} + \frac{d^{1/2} L_0 T}{2\delta} \frac{\beta_2}{1-\beta_1} + \frac{d^{1/2} L_0}{2\delta} \Big(\mathbb{E} \|x_1 - x_0\|^2 - \frac{\beta_2}{1-\beta_1} \Big) \frac{1-\beta_1^T}{1-\beta_1}, \end{aligned}$$

where the last inequality follows from Equation (3.11) in Ghadimi & Lan (2013) and Equation (18) in Nesterov & Spokoiny (2017). Choose $\eta = \frac{\epsilon_f^{1/2}}{2(d+2)^{3/2} T^{1/2} L_0^2}$ and $\delta = \frac{\epsilon_f}{(d+2)^{1/2} L_0}$ with certain $\epsilon_f < 1$, and set $T > \frac{1}{2\epsilon_f d}$. Then, we have $\beta_1 = \frac{1}{\epsilon_f (d+2) T} < \frac{1}{2}$, and thus the above inequality yields

$$\sum_{t=0}^{T-1} \eta \mathbb{E} \|\nabla f_\delta(x_t)\|^2 \le f(x_0) - \inf_x f(x) + 2\delta L_0 d^{1/2} + \frac{\beta_2 d^{1/2} L_0 T}{\delta} + \frac{d^{1/2} L_0 \eta^2}{\delta} \mathbb{E} \|\widetilde{g}_b(x_0)\|^2. \tag{69}$$

Choosing ζ from $0, 1, ..., T-1$ uniformly at random, and rearranging the above inequality, we have

$$\begin{aligned} \mathbb{E} \|\nabla f_\delta(x_\zeta)\|^2 \le & \frac{f(x_0) - \inf_x f(x)}{\eta T} + \frac{2\delta L_0 d^{1/2}}{\eta T} + \frac{d^{1/2} L_0 \eta}{\delta T} \mathbb{E} \|\widetilde{g}(x_0)\|^2 \\ & + \frac{16\eta (d+4)^2 d^{1/2} L_0^3}{\delta} + \frac{8\eta (d+2) d^{1/2} L_0 \sigma^2}{\delta^3 b} \\ \le & \mathcal{O}\left(\frac{1}{\eta T} + \frac{\delta L_0 d^{1/2}}{\eta T} + \frac{d^{1/2} L_0 \eta}{\delta T} + \frac{\eta d^{5/2} L_0^3}{\delta} + \frac{\eta d^{3/2} L_0 \sigma^2}{\delta^3 b} \right), \end{aligned} \tag{70}$$

which, in conjunction with $\eta = \frac{\epsilon_f^{1/2}}{2(d+2)^{3/2}T^{1/2}L_0^2}$, $\delta = \frac{\epsilon_f}{(d+2)^{1/2}L_0}$ and $b = \frac{\sigma^2}{\epsilon_f^2}$, yields

$$\begin{aligned}
\mathbb{E}\|\nabla f_\delta(x_\zeta)\|^2 \leq & \mathcal{O}\left(\frac{1}{\eta T} + \frac{\delta L_0 d^{1/2}}{\eta T} + \frac{d^{1/2}L_0\eta}{\delta T} + \frac{\eta d^{5/2}L_0^3}{\delta} + \frac{\eta d^{3/2}L_0\sigma^2}{\delta^3 b}\right) \\
\leq & \mathcal{O}\left(\frac{d^{3/2}L_0^2}{\epsilon_f^{1/2}\sqrt{T}} + \frac{\epsilon_f^{1/2}d^{3/2}L_0^2}{\sqrt{T}} + \frac{1}{\epsilon_f^{1/2}d^{1/2}T^{3/2}} + \frac{d^{3/2}\sigma^2}{\epsilon_f^{5/2}\sqrt{T}b}\right) \\
\leq & \mathcal{O}\left(\left(1 + \frac{\sigma^2}{\epsilon_f^2 b}\right)\frac{d^{3/2}}{\epsilon_f^{1/2}\sqrt{T}}\right) \leq \mathcal{O}\left(\frac{d^{3/2}}{\epsilon_f^{1/2}\sqrt{T}}\right).
\end{aligned}$$

Based on $\delta = \frac{\epsilon_f}{(d+2)^{1/2}L_0}$ and Equation (18) in Nesterov & Spokoiny (2017), we have $|f_\delta(x) - f(x)| < \epsilon_f$. Then, to achieve an ϵ-accurate stationary point of the smoothed function f_δ with approximation error $|f_\delta(x) - f(x)| < \epsilon_f, \epsilon_f < 1$, we need $T \leq \mathcal{O}(\epsilon_f^{-1}d^3\epsilon^{-2})$, and thus the corresponding total function query complexity is given by

$$Tb = \mathcal{O}\left(\frac{\sigma^2 d^3}{\epsilon_f^3\epsilon^2}\right). \tag{71}$$

Then, the proof is complete. □

Nonsmooth Convex Geometry

In this part, we provide the convergence and complexity analysis for our proposed algorithm for the case where $F(x;\xi)$ is convex and belongs to $C^{0,0}$.

Theorem I.5. *Suppose Assumptions 4.1 and 4.2 are satisfied and $\mathbb{E}\|\widetilde{g}_b(x_0)\|^2 \leq Md^2T$ for certain constant $M > 0$. Choose $\eta = \frac{1}{(d+2)\sqrt{T}L_0}$, $\delta = \frac{(d+2)^{1/2}}{\sqrt{T}}$, $b = \frac{\sigma^2 T}{d^2}$, and $T > d^2$. Then, we have $\mathbb{E}\big(f(x_\zeta) - \inf_x f(x)\big) \leq \mathcal{O}\big(\frac{d}{\sqrt{T}}\big)$. Then, to achieve an ϵ-accurate solution of $f(x)$, the corresponding total function query complexity is given by*

$$Tb = \mathcal{O}\Big(\frac{\sigma^2 d^2}{\epsilon^4}\Big). \tag{72}$$

Proof. Let x^* be a minimizer of the function f, i.e. $x^* = \arg\min_x f(x)$. Then, we have

$$\begin{aligned}
\|x_{t+1} - x^*\|^2 &= \|x_t - \eta\widetilde{g}_b(x_t) - x^*\|^2 \\
&= \|x_t - x^*\|^2 - 2\eta\left\langle \widetilde{g}_b(x_t), x_t - x^*\right\rangle + \mathbb{E}\|x_{t+1} - x_t\|^2.
\end{aligned}$$

Telescoping the above inequality over t from 0 to $T-1$ yields that

$$\|x_T - x^*\|^2 = \|x_0 - x^*\|^2 - 2\eta\sum_{t=0}^{T-1}\left\langle \widetilde{g}_b(x_t), x_t - x^*\right\rangle + \sum_{t=0}^{T-1}\mathbb{E}\|x_{t+1} - x_t\|^2.$$

Taking expectation in the above equality using the fact that $\mathbb{E}[\widetilde{g}_b(x_t)|x_t] = \nabla f_\delta(x_t)$, we further obtain that

$$\begin{aligned}
\mathbb{E}\|x_T - x^*\|^2 &= \|x_0 - x^*\|^2 - 2\eta\sum_{t=0}^{T-1}\mathbb{E}\left\langle \nabla f_\delta(x_t), x_t - x^*\right\rangle + \sum_{t=0}^{T-1}\mathbb{E}\|x_{t+1} - x_t\|^2 \\
&\overset{(i)}{\leq} \|x_0 - x^*\|^2 - 2\eta\sum_{t=0}^{T-1}\mathbb{E}\big(f_\delta(x_t) - f_\delta(x^*)\big) + \sum_{t=0}^{T-1}\mathbb{E}\|x_{t+1} - x_t\|^2 \\
&\overset{(ii)}{\leq} \|x_0 - x^*\|^2 - 2\eta\sum_{t=0}^{T-1}\mathbb{E}\big(f(x_t) - f(x^*)\big) + 4\eta\delta L_0\sqrt{d}T + \sum_{t=0}^{T-1}\mathbb{E}\|x_{t+1} - x_t\|^2,
\end{aligned} \tag{73}$$

where (i) follows from the convexity of f_δ and (ii) uses the fact that $|f_\delta(x)-f(x)|\le \delta L_0\sqrt{d}$. Then, rearranging the above inequality yields

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\big(f(x_t)-f(x^*)\big)&\le \frac{\|x_0-x^*\|^2}{\eta T}+4\delta L_0\sqrt{d}+\frac{1}{\eta T}\sum_{t=0}^{T-1}\mathbb{E}\|x_{t+1}-x_t\|^2\\ &\overset{(i)}{\le}\frac{\|x_0-x^*\|^2}{\eta T}+4\delta L_0\sqrt{d}+\frac{1}{\eta T}\sum_{t=0}^{T-1}\Big(\beta_1^t\Big(\mathbb{E}\|x_1-x_0\|^2-\frac{\beta_2}{1-\beta_1}\Big)+\frac{\beta_2}{1-\beta_1}\Big)\\ &\le\frac{\|x_0-x^*\|^2}{\eta T}+4\delta L_0\sqrt{d}+\frac{1}{\eta T}\frac{1-\beta_1^T}{1-\beta_1}\Big(\mathbb{E}\|x_1-x_0\|^2-\frac{\beta_2}{1-\beta_1}\Big)+\frac{\beta_2}{\eta(1-\beta_1)},\end{aligned}$$

where (i) follows from Lemma I.3 with $\beta_1=\frac{4\eta^2(d+2)L_0^2}{\delta^2}$ and $\beta_2=16\eta^2L_0^2(d+4)^2+\frac{8\eta^2(d+2)\sigma^2}{\delta^2 b}$. Recalling $\eta=\frac{1}{(d+2)\sqrt{T}L_0}$, $\delta=\frac{(d+2)^{1/2}}{\sqrt{T}}$, $b=\frac{\sigma^2 T}{d^2}$, and $T>d^2$, we have $\beta_1<1/2$, and the above inequality yields

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\big(f(x_t)-f(x^*)\big)\le&\frac{\|x_0-x^*\|^2}{\eta T}+4\delta L_0\sqrt{d}+\frac{2\eta}{T}\mathbb{E}\|\widetilde{g}_b(x_0)\|^2+32\eta L_0^2(d+4)^2\\ &+\frac{16\eta(d+2)\sigma^2}{\delta^2 b}\le\mathcal{O}\left(\frac{d}{\sqrt{T}}+\frac{\mathbb{E}\|\widetilde{g}_b(x_0)\|^2}{dT^{3/2}}\right),\end{aligned}$$

which, combined with $\mathbb{E}\|\widetilde{g}_b(x_0)\|^2\le Md^2T$ for constant M and choosing ζ from $0,..,T-1$ uniformly at random, yields

$$\mathbb{E}\big(f(x_\zeta)-f(x^*)\big)\le\mathcal{O}\Big(\frac{d}{\sqrt{T}}\Big).\tag{74}$$

To achieve an ϵ-accurate solution, i.e., $\mathbb{E}\big(f(x_\zeta)-f(x^*)\big)<\epsilon$, we need $T=\mathcal{O}(d^2\epsilon^{-2})$, and hence the corresponding function query complexity is given by

$$Tb\le\mathcal{O}\Big(\sigma^2d^2\epsilon^{-4}\Big),\tag{75}$$

which finishes the proof. ☐

I.1 Analysis in Smooth Setting

In this section, we provide the convergence and complexity analysis for the proposed gradient estimator when function $F(x,\xi)\in C^{1,1}$

Smooth Nonconvex Geometry

In this part, we provide the convergence and complexity analysis for the proposed gradient estimator for the case where $F(x;\xi)$ is nonconvex and belongs to $C^{0,0}\cap C^{1,1}$.

Theorem I.6. *Suppose Assumptions 4.1, 4.2 and I.1 are satisfied and* $\mathbb{E}\|\widetilde{g}_b(x_0)\|^2\le MTd^{8/3}$ *for certain constant* $M>0$. *Choose* $\eta=\frac{1}{4(d+2)^{4/3}\sqrt{T}\max(L_0,L_1)}<\frac{1}{8L_1}$, $\delta=\frac{1}{(d+2)^{5/6}T^{1/4}}$ *and* $b=\max\big(\sigma^2,\frac{\sigma_g^2}{\sqrt{T}d^{5/3}}\big)\sqrt{T}$. *Then, we have* $\mathbb{E}\|\nabla f_\delta(x_\zeta)\|^2\le\mathcal{O}\Big(\frac{d^{4/3}}{\sqrt{T}}\Big)$. *Then, to achieve an* ϵ-*accurate stationary point of* f, *the total function query complexity is given by*

$$Tb=\mathcal{O}\big(\sigma^2d^4\epsilon^{-3}+\sigma_g^2d\epsilon^{-2}\big).\tag{76}$$

Proof. Based on Equation (12) in Nesterov & Spokoiny (2017), the smoothed function $f_\delta\in C^{1,1}$ with gradient-Lipschitz constant less than L_1. Then, we have

$$\begin{aligned}f_\delta(x_{t+1})\le&f_\delta(x_t)+\langle\nabla f_\delta(x_t),x_{t+1}-x_t\rangle+\frac{L_1}{2}\|x_{t+1}-x_t\|^2\\ =&f_\delta(x_t)-\eta\langle\nabla f_\delta(x_t),\widetilde{g}_b(x_t)\rangle+\frac{L_1}{2}\|x_{t+1}-x_t\|^2.\end{aligned}$$

Let $\alpha = \frac{6(d+2)L_0^2\eta^2}{\delta^2}$, $\beta = 12L_1^2\delta^2(d+6)^3 + \frac{48(d+4)\sigma_g^2}{b} + \frac{8(d+2)\sigma^2}{\delta^2 b}$ and $p_{t-1} = 24(d+4)\mathbb{E}(\|\nabla f(x_t)\|^2 + \|\nabla f(x_{t-1})\|^2)$. Then, telescoping the bound in Lemma I.2 in the smooth case yields

$$\begin{aligned}\mathbb{E}\|\widetilde{g}_b(x_t)\|^2 &\leq \alpha^t\mathbb{E}\|\widetilde{g}_b(x_0)\|^2 + \sum_{j=0}^{t-1}\alpha^{t-1-j}p_j + \beta\sum_{j=0}^{t-1}\alpha^j \\ &\leq \alpha^t\mathbb{E}\|\widetilde{g}_b(x_0)\|^2 + \sum_{j=0}^{t-1}\alpha^{t-1-j}p_j + \frac{\beta(1-\alpha^t)}{1-\alpha}. \end{aligned} \tag{77}$$

Taking expectation over the above inequality and using $\mathbb{E}(\widetilde{g}_b(x_t)|x_t) = \nabla f_\delta(x_t)$, we have

$$\mathbb{E}f_\delta(x_{t+1}) \leq \mathbb{E}f_\delta(x_t) - \eta\mathbb{E}\|\nabla f_\delta(x_t)\|^2 + \frac{L_1\eta^2}{2}\mathbb{E}\|\widetilde{g}_b(x_t)\|^2.$$

Telescoping the above inequality over t from 0 to $T-1$ yields

$$\begin{aligned}&\mathbb{E}f_\delta(x_k) \leq f_\delta(x_0) - \eta\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f_\delta(x_t)\|^2 + \frac{L_1\eta^2}{2}\sum_{t=0}^{T-1}\mathbb{E}\|\widetilde{g}_b(x_t)\|^2 \\ &\overset{(i)}{\leq} f_\delta(x_0) - \eta\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f_\delta(x_t)\|^2 + \frac{L_1\eta^2}{2}\mathbb{E}\|\widetilde{g}_b(x_0)\|^2 \\ &\qquad + \frac{L_1\eta^2}{2}\sum_{t=1}^{T-1}\left(\alpha^t\mathbb{E}\|\widetilde{g}_b(x_0)\|^2 + \sum_{j=0}^{t-1}\alpha^{t-1-j}p_j + \frac{\beta(1-\alpha^t)}{1-\alpha}\right) \\ &= f_\delta(x_0) - \eta\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f_\delta(x_t)\|^2 + \frac{1-\alpha^T}{1-\alpha}\frac{L_1\eta^2}{2}\mathbb{E}\|\widetilde{g}_b(x_0)\|^2 \\ &\qquad + \frac{L_1\eta^2}{2}\sum_{j=0}^{T-2}\sum_{t=0}^{T-2-j}\alpha^t p_j + \frac{L_1\eta^2}{2}\sum_{t=1}^{T-1}\frac{\beta(1-\alpha^t)}{1-\alpha}\end{aligned}$$

Since $\sum_{j=0}^{T-2}\sum_{t=0}^{T-2-j}\alpha^t p_j \leq \sum_{j=0}^{T-2}\sum_{t=0}^{T-2}\alpha^t p_j$, we have that

$$\begin{aligned}\mathbb{E}f_\delta(x_k) &\leq f_\delta(x_0) - \eta\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f_\delta(x_t)\|^2 + \frac{1-\alpha^T}{1-\alpha}\frac{L_1\eta^2}{2}\mathbb{E}\|\widetilde{g}_b(x_0)\|^2 \\ &\qquad + \frac{L_1\eta^2}{2}\sum_{j=0}^{T-2}\sum_{t=0}^{T-2}\alpha^t p_j + \frac{L_1\eta^2}{2}\sum_{t=1}^{T-1}\frac{\beta(1-\alpha^t)}{1-\alpha} \\ &\leq f_\delta(x_0) - \eta\sum_{t=0}^{T-1}\mathbb{E}\|\nabla f_\delta(x_t)\|^2 + \frac{1-\alpha^T}{1-\alpha}\frac{L_1\eta^2}{2}\mathbb{E}\|\widetilde{g}_b(x_0)\|^2 \\ &\qquad + \frac{L_1\eta^2}{2}\frac{1-\alpha^{T-1}}{1-\alpha}\sum_{t=0}^{T-2}p_t + \frac{L_1\eta^2}{2}\sum_{t=1}^{T-1}\frac{\beta(1-\alpha^t)}{1-\alpha}\end{aligned}$$

where (i) follows from (77). Choose $\eta = \frac{1}{4(d+2)^{4/3}\sqrt{T}\max(L_0,L_1)}$ and $\delta = \frac{1}{(d+2)^{5/6}T^{1/4}}$. Then, we have $\alpha \leq \frac{3}{8} < \frac{1}{2}$, and the above inequality yields

$$\mathbb{E}f_\delta(x_k) \leq f_\delta(x_0) - \eta\sum_{t=0}^{k-1}\mathbb{E}\|\nabla f_\delta(x_t)\|^2 + L_1\eta^2\mathbb{E}\|\widetilde{g}_b(x_0)\|^2 + L_1\eta^2\sum_{t=0}^{T-2}p_t + L_1\eta^2 T\beta.$$

Rearranging the above inequality and using $|f_\delta(x) - f| \leq \frac{\delta^2}{2} L_1 d$ and $\|\nabla f_\delta(x) - \nabla f(x)\| \leq \frac{\delta}{2} L_1 (d+3)^{3/2}$ proved in Nesterov & Spokoiny (2017), we have

$$\begin{aligned}
\mathbb{E} f(x_k) \leq & f(x_0) + \delta^2 L_1 d - \frac{\eta}{2} \sum_{t=0}^{T-1} \mathbb{E}\|\nabla f(x_t)\|^2 + \frac{\eta T}{4} \delta^2 L_1^2 (d+3)^3 \\
& + L_1 \eta^2 \mathbb{E}\|\widetilde{g}_b(x_0)\|^2 + L_1 \eta^2 \sum_{t=0}^{T-2} p_t + L_1 \eta^2 T \beta \\
\leq & f(x_0) + \delta^2 L_1 d - \frac{\eta}{2} \sum_{t=0}^{T-1} \mathbb{E}\|\nabla f(x_t)\|^2 + \frac{\eta T}{4} \delta^2 L_1^2 (d+3)^3 \\
& + L_1 \eta^2 \mathbb{E}\|\widetilde{g}_b(x_0)\|^2 + 2 L_1 \eta^2 \sum_{t=0}^{T-1} \|\nabla f(x_t)\|^2 + L_1 \eta^2 T \beta.
\end{aligned}$$

Choosing ζ from $0, ..., T-1$ uniformly at random, we obtain from the above inequality that

$$\begin{aligned}
\big(\frac{1}{2} - 2L_1 \eta\big) \mathbb{E}\|\nabla f(x_\zeta)\|^2 \leq & \frac{f(x_0) - \inf_x f(x)}{\eta T} + \frac{\delta^2 L_1 d}{\eta T} \\
& + \frac{L_1^2}{4} \delta^2 (d+3)^3 + \frac{L_1 \eta \mathbb{E}\|\widetilde{g}_b(x_0)\|^2}{T} + L_1 \eta \beta,
\end{aligned}$$

which, in conjunction with $\eta = \frac{1}{4(d+2)^{4/3}\sqrt{T}\max(L_0, L_1)} < \frac{1}{8L_1}$, $\delta = \frac{1}{(d+2)^{5/6} T^{1/4}}$, $\beta = 12 L_1^2 \delta^2 (d+6)^3 + \frac{48(d+4)\sigma_g^2}{b} + \frac{8(d+2)\sigma^2}{\delta^2 b}$, $b = \max\big(\sigma^2, \frac{\sigma_g^2}{\sqrt{T} d^{5/3}}\big)\sqrt{T}$ and $\mathbb{E}\|\widetilde{g}_b(x_0)\|^2 \leq M T d^{8/3}$, yields

$$\mathbb{E}\|\nabla f(x_\zeta)\|^2 \leq \mathcal{O}\Big(\frac{d^{4/3}}{\sqrt{T}} + \frac{d^{2/3}}{T} + \frac{d^{4/3}}{\sqrt{T}} + \frac{d^{4/3}}{\sqrt{T}} + \frac{1}{T} + \frac{\sigma_g^2}{d^{1/3} b \sqrt{T}} + \frac{d^{4/3}\sigma^2}{b}\Big) \leq \mathcal{O}\Big(\frac{d^{4/3}}{\sqrt{T}}\Big).$$

Then, to achieve an ϵ-accurate stationary point of function f, i.e., $\mathbb{E}\|\nabla f(x_\zeta)\|^2 < \epsilon$, we need $T = \mathcal{O}(d^{8/3}\epsilon^{-2})$the total number of function query is given by $Tb \leq \mathcal{O}\big(\sigma^2 d^4 \epsilon^{-3} + \sigma_g^2 d \epsilon^{-2}\big)$. $\square$

Smooth Convex Geometry

In this part, we provide the convergence and complexity analysis for the proposed gradient estimator for the case where $F(x;\xi)$ is convex and belongs to $C^{0,0} \cap C^{1,1}$.

Theorem I.7. *Suppose Assumptions 4.1, 4.2 and I.1 are satisfied and $\mathbb{E}\|\widetilde{g}_b(x_0)\|^2 \leq MTd^2$ for certain constant $M > 0$. Choose $\eta = \frac{1}{192(d+2)\sqrt{T}\max(L_0, L_1)}$ and $\delta^2 = \frac{1}{\sqrt{T}}$ and $b = \max\big(\frac{\sigma_g^2}{\sqrt{T}}, \sigma^2\big)\sqrt{T}/d$. Then, we have $\mathbb{E}\|\nabla f_\delta(x_\zeta)\|^2 \leq \mathcal{O}\Big(\frac{d}{\sqrt{T}} + \frac{d^2}{T}\Big)$. Then, to achieve an ϵ-accurate stationary point of f, the total function query complexity is given by*

$$Tb = \mathcal{O}\big(\sigma^2 d^2 \epsilon^{-3} + \sigma_g^2 d \epsilon^{-2}\big). \tag{78}$$

Proof. Using an approach similar to (73), we have

$$\mathbb{E}\|x_T - x^*\|^2 \leq \|x_0 - x^*\|^2 - 2\eta \sum_{t=0}^{T-1} \mathbb{E}\big(f(x_t) - f(x^*)\big) + 2\eta\delta^2 L_1 d T + \sum_{t=0}^{T-1} \eta^2 \mathbb{E}\|\widetilde{g}_b(x_t)\|^2,$$

where the last inequality follows from Equation (19) in Nesterov & Spokoiny (2017). Let $\alpha = \frac{6(d+2)L_0^2\eta^2}{\delta^2}$, $\beta = 12 L_1^2 \delta^2 (d+6)^3 + \frac{48(d+4)\sigma_g^2}{b} + \frac{8(d+2)\sigma^2}{\delta^2 b}$ and $p_{t-1} = 24(d+4)\mathbb{E}(\|\nabla f(x_t)\|^2 +$

$\|\nabla f(x_{t-1})\|^2)$. Then, combining the above inequality with (77) yields

$$\begin{aligned}
\mathbb{E}\|x_T - x^*\|^2 \leq& \|x_0 - x^*\|^2 - 2\eta \sum_{t=0}^{T-1} \mathbb{E}\big(f(x_t) - f(x^*)\big) + 2\eta\delta^2 L_1 dT + \eta^2 \mathbb{E}\|\widetilde{g}_b(x_0)\|^2 \\
&+ \sum_{t=1}^{T-1} \eta^2 \Big(\alpha^t \mathbb{E}\|\widetilde{g}_b(x_0)\|^2 + \sum_{j=0}^{t-1} \alpha^{t-1-j} p_j + \frac{\beta(1-\alpha^t)}{1-\alpha}\Big) \\
\leq& \|x_0 - x^*\|^2 - 2\eta \sum_{t=0}^{T-1} \mathbb{E}\big(f(x_t) - f(x^*)\big) + 2\eta\delta^2 L_1 dT + \eta^2 \frac{1-\alpha^T}{1-\alpha} \mathbb{E}\|\widetilde{g}_b(x_0)\|^2 \\
&+ \eta^2 \sum_{j=0}^{T-2} \sum_{t=0}^{T-2} \alpha^t p_j + \eta^2 \sum_{t=1}^{T-1} \frac{\beta(1-\alpha^t)}{1-\alpha} \\
\leq& \|x_0 - x^*\|^2 - 2\eta \sum_{t=0}^{T-1} \mathbb{E}\big(f(x_t) - f(x^*)\big) + 2\eta\delta^2 L_1 dT + \eta^2 \frac{1-\alpha^T}{1-\alpha} \mathbb{E}\|\widetilde{g}_b(x_0)\|^2 \\
&+ 24(d+4)\eta^2 \frac{1-\alpha^{T-1}}{1-\alpha} \sum_{t=0}^{T-2} \mathbb{E}(\|\nabla f(x_{t+1})\|^2 + \|\nabla f(x_t)\|^2) + \eta^2 \sum_{t=1}^{T-1} \frac{\beta(1-\alpha^t)}{1-\alpha}. \quad (79)
\end{aligned}$$

Recalling $\eta = \frac{1}{192(d+2)\sqrt{T}\max(L_0, L_1)}$ and $\delta^2 = \frac{1}{\sqrt{T}}$, we have $\alpha < \frac{1}{2}$, and thus the above inequality yields

$$\begin{aligned}
\mathbb{E}\|x_T - x^*\|^2 \leq& \|x_0 - x^*\|^2 - 2\eta \sum_{t=0}^{T-1} \mathbb{E}\big(f(x_t) - f(x^*)\big) + 2\eta\delta^2 L_1 dT + 2\eta^2 \mathbb{E}\|\widetilde{g}_b(x_0)\|^2 \\
&+ 48(d+4)\eta^2 \sum_{t=0}^{T-2} \mathbb{E}(\|\nabla f(x_{t+1})\|^2 + \|\nabla f(x_t)\|^2) + 2T\eta^2\beta. \quad (80)
\end{aligned}$$

Since the convexity implies that $\frac{1}{2L_1}\|\nabla f(x)\|^2 \leq f(x) - f(x^*)$ for any x, rearranging the above inequality yields

$$\begin{aligned}
&(2 - 192(d+4)L_1\eta) \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}(f(x_t) - f(x^*)) \\
&\leq \frac{\|x_0 - x^*\|^2}{\eta T} + 2\delta^2 L_1 d + \frac{2\eta \mathbb{E}\|\widetilde{g}_b(x_0)\|^2}{T} + 24\eta L_1^2 \delta^2 (d+6)^3 + \frac{96\eta(d+4)\sigma_g^2}{b} + \frac{16\eta(d+2)\sigma^2}{\delta^2 b}, \quad (81)
\end{aligned}$$

which, in conjunction with $\mathbb{E}\|\widetilde{g}_b(x_0)\|^2 \leq Md^2T$ for certain constant $M > 0$, $b = \max\big(\frac{\sigma_g^2}{\sqrt{T}}, \sigma^2\big)\sqrt{T}/d$ and recalling that ζ is chosen from $0, ..., T-1$ uniformly at random, yields

$$\mathbb{E}(f(x_\zeta) - f(x^*)) \leq \mathcal{O}\Big(\frac{d}{\sqrt{T}} + \frac{d^2}{T}\Big). \quad (82)$$

Then, to achieve an ϵ-accurate solution, i.e., $\mathbb{E}(f(x_\zeta) - f(x^*)) \leq \epsilon$, we need $T = \mathcal{O}(d^2\epsilon^{-2})$, and thus the corresponding query complexity is given by

$$Tb \leq \mathcal{O}(\sigma^2 d^2 \epsilon^{-3} + \sigma_g^2 d \epsilon^{-2}) \quad (83)$$

□