Proceedings of Machine Learning Research vol 195:1–28, 2023 36th Annual Conference on Learning Theory

Online Nonconvex Optimization with Limited Instantaneous Oracle Feedback

Ziwei Guan GUAN.283@BUCKEYEMAIL.OSU.EDU
2015 Neil Ave, Columbus, Ohio, USA

Yi Zhou YI.ZHOU@UTAH.EDU
50 Central Campus Dr #2110, Salt Lake City, Utah, USA

Yingbin Liang LIANG.889@OSU.EDU
2015 Neil Ave, Columbus, Ohio, USA

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

We investigate online nonconvex optimization from a local regret minimization perspective. Previous studies along this line implicitly required the access to sufficient gradient oracles at each time instance in order to design double-loop algorithms. In this work, we focus on more challenging but practical settings where only limited number of oracles are available in online nonconvex optimization, including window-smoothed single gradient oracle (Window-SGO), single function value oracle (Window-SVO) and multiple function value oracles (Window-MVO). Specifically, in the Window-SGO setting which allows only single-loop algorithm design, we derive a local regret lower bound, which indicates that single-loop algorithms are provably worse than double-loop algorithms. Further, the simple classical OGD algorithm achieves the window-unconditioned lower bound. Moreover, in the Window-SVO setting, we propose a novel single-loop online algorithm named SkipOGD, and show that it achieves a near-optimal local regret that matches the Window-SGO regret lower bound up to a factor of the dimension d due to the function value feedback. Lastly, in the Window-MVO setting, we propose a new double-loop online algorithm named LoopOGD and show that it achieves a smooth trade-off between regret minimization and sample complexity over the number of oracle calls K per time instance. In particular, with $K = 1$ and wd, LoopOGD respectively achieves our regret lower bound with Window-SGO (up to the factor d due to function value feedback) and the existing regret lower bound with multiple gradient oracle feedback.

Keywords: Online nonconvex optimization; limited instantaneous oracle; window-smoothed function; local regret.

1. Introduction

As a popular and powerful machine learning framework, online optimization models the learning process in a dynamic environment and has been extensively studied in the literature (Hazan et al., 2016; Orabona, 2019). Specifically, the learning objective function $f_t : \mathbb{R}^d \to \mathbb{R}$ in online optimization varies over time t. At each time t, the agent has access to the historical information of $f_1, \ldots, f_{t-1}$, and is required to submit a control variable x_t to the environment. Then, certain feedbacks concerning f_t (such as $f_t(x_t)$ or $\nabla f_t(x_t)$) are revealed to the agent after the decision is made. The goal of the agent is to minimize the gap between the cumulative function value at the decisions and minimized with hindsight.

Most previous studies of online optimization have focused on the **convex** setting, where the learning objectives are convex functions. Studies on online **nonconvex** optimization so far have

been very limited. Among these studies, most work such as (Krichene et al., 2015; Agarwal et al., 2019; Lesage-Landry et al., 2020; Héliou et al., 2020) studied the notion of the *global regret* with an assumption that the access to global minimum points of *nonconvex* objective functions is available, which is typically not feasible in practice. More recently, several studies (Hazan et al., 2017; Aydore et al., 2019; Hallak et al., 2021) focused on a more feasible notion of the *local regret*, defined as the cumulative gradient norm of window-smoothed objective functions, and studied the case where only gradient oracles of nonconvex objective functions are available. Those studies further designed algorithms and showed that the corresponding local regret decreases sublinearly as long as the window size of objective functions is chosen properly. The success of these algorithms crucially rely on a double-loop design, where the inner loop can obtain an accurate stationary point of the window-smoothed objective functions. This requires that at each time, sufficiently many gradient oracles are available in order to iteratively update the decision variables to a target accuracy. In many practical applications (Roughgarden and Schrijvers, 2017; Hao et al., 2017), there are hard constraints on the number of feedback oracles accessible at each time (for example, only a single oracle feedback is available), and clearly the existing algorithms are not applicable.

In this paper, we investigate the local regret of online *nonconvex* optimization with the hard constraint on the number of oracles accessible at each time, and we focus on the following three main settings. Below we discuss the main challenges in each setting and describe our contributions.

(a) **Window-smoothed single gradient oracle (Window-SGO) feedback**, where only one instantaneous gradient oracle, with respect to the window-smoothed objective functions, is available at a time, and clearly no inner loop is allowed to update the variable to a target accuracy. This is not an issue for online convex optimization, where the classical single-loop online gradient descent (OGD) achieves the optimal regret. But for online *nonconvex* optimization, the existing literature has been exclusively focused on double-loop design, and it is unclear whether single-loop algorithms are provably worse than double-loop algorithms. To address this problem, we develop a lower bound on the local regret that all such algorithms must satisfy, which indicates that it is impossible for single-loop algorithms to achieve the same regret of $\mathcal{O}\left(\frac{T}{w^2}\right)$ that double-loop algorithms can achieve for all window size w, where T denotes the time horizon. Further, the simple OGD algorithm achieves the regret of $\Omega\left(\frac{T}{w}\right)$ which is the best possible regret (that holds uniformly for all window sizes) for single-loop algorithms.

(b) **Window-smoothed single function value oracle (Window-SVO) feedback**, where only one oracle of the window-smoothed functions is available at a time, and no inner loop is allowed to update the variable to a target accuracy. In this setting, we show that directly applying the best-known one-point residual (OR) gradient estimator to the online case does not yield an optimal regret (that matches the lower bound for single-loop algorithms). Our main contribution here lies in proposing a novel algorithm (called SkipOGD), which queries the function values alternatively at perturbed and un-perturbed point over adjacent times so that the oracle at un-perturbed point can serve as a better residual to improve the gradient estimation accuracy. As a result, such a scheme necessarily leads to skipped updates of variables (at un-perturbed times). Interestingly, we show that the regret bound of SkipOGD matches the first-order lower bound in the dependency of window size w and time T with the additional scaling factor dimension d due to the function value feedback. This indicates that the benefit gained due to un-perturbed points significantly outweigh the potential loss due to skipped updates.

(c) **Window-smoothed multiple function value oracles (Window-MVO) feedback**, where multiple oracles of the window-smoothed functions are available at a time, but subject to the hard

constraint of $2K$ ($K \geq 1$) oracles per time. For such a setting, we design a double-loop algorithm named LoopOGD, where the inner loop takes K updates and each update uses a two-point gradient estimator. We show that LoopOGD achieves a full spectrum of regret performance between the settings of single and multiple oracle calls. In particular, with $K = 1$ and wd, LoopOGD respectively achieves our window-unconditioned regret lower bound of $\mathcal{O}\left(\frac{T}{w}\right)$ for Window-SGO (up to the factor d due to function value feedback) and the existing regret lower bound of $\mathcal{O}\left(\frac{T}{w^2}\right)$ for window-smoothed multiple gradient oracle (Window-MGO).

2. Related Works

Online convex optimization: Online convex optimization has been extensively studied, both in terms of static regret analysis, e.g., Zinkevich (2003); Hazan and Kale (2011); Rakhlin et al. (2011); Shamir and Zhang (2013), and non-stationary regret, e.g., Besbes et al. (2015); Jadbabaie et al. (2015); Zhang et al. (2018); Zhao et al. (2020, 2021). Interested readers can refer to (Hazan et al., 2016; Orabona, 2019) and the references therein for a more comprehensive list of studies on online convex optimization.

As first-order information is typically hard to obtain in many applications, online convex optimization with *bandit feedback* has attracted much attention in the past. Flaxman et al. (2005); Saha and Tewari (2011) studied such a setting with a one-point gradient estimator. Later, Agarwal et al. (2010) adopted a better two-point estimator, which significantly reduced gradient estimation variance. Those one-point and two-point gradient estimators were further applied to other online settings with functional constraints (Cao and Liu, 2018, 2019; Kim and Lee, 2023) and distributed agents (Yi et al., 2020; Wang et al., 2022).

Online nonconvex optimization: One line of research on online *nonconvex* optimization studied the metric of the *global regret*. Specifically, Krichene et al. (2015) designed an algorithm by generalizing the Hedge algorithm to the continuous and high-dimensional domain, which was later improved by Yang et al. (2018) with a novel weighting strategy. Another set of studies assumed that the offline global optimizers are available. Agarwal et al. (2019) provided an upper bound on the global regret for the classical algorithm follow-the-perturbed-leader (FTPL), which was then further improved by (Suggala and Netrapalli, 2020) via a refined analysis. Lesage-Landry et al. (2020) investigated a slow-changing online nonconvex optimization problem and showed that the Newton's method can have a good track of the global minimum when given access to an offline nonconvex optimization oracle. Besides, Gao et al. (2018) studied an online nonconvex problem where functions satisfy weak pseudo-convex conditions and analyzed an online normalized gradient descent algorithm. *Online nonconvex optimization with bandit feedback* has also been studied from the global regret aspect. Specifically, (Héliou et al., 2020) provided upper bounds on both the global static and dynamic regret for a dual averaging algorithm with the access of function value feedback. Those bounds were further improved in (Héliou et al., 2021). Gao et al. (2018) also extended the aforementioned study to the zeroth-order type of algorithm.

Recently, several studies focused on the case where only gradient oracles of window-smoothed nonconvex objective functions are available and provided bounds on the *local regret*. Specifically, Hazan et al. (2017) provided a lower bound of $\Omega\left(\frac{T}{w^2}\right)$ on the local regret, and proposed a double-loop scheme that achieves such a lower bound. Later, Aydore et al. (2019); Hallak et al. (2021) extended the results to dynamic environment applications and the nonsmooth settings, respectively. All of the aforementioned studies require the access to sufficient gradient oracles at each time in

order to design double-loop algorithms, which is often not satisfied. In contrast, we focus on characterizing local regret online nonconvex optimization in a more practical setting with a hard threshold on the number of oracle queries each time. Further, along this line of research on the local regret, there has not been any study on the setting with bandit feedback, which is also the focus of our paper.

Gradient estimation methods with bandit feedback: Various gradient estimation methods via function values (i.e., bandit feedbacks) have been developed in the past. Existing such approaches can be mainly divided into one-point methods (i.e., gradient estimation via one function value oracle) (Flaxman et al., 2005; Dekel et al., 2015; Gasnikov et al., 2017), and two-point methods (i.e., gradient estimation via two function values) (Agarwal et al., 2010; Nesterov and Spokoiny, 2017; Shamir, 2017). Compared to two-point methods, one-point methods yield much larger variance, and consequently zeroth-order algorithms designed based on one-point methods have much slower convergence. Recently, Zhang et al. (2022) proposed a variance-reduction technique for one-point methods by subtracting the feedback from the previous iteration. Such a method, although yields better convergence rate, still cannot match the performance of two-point gradient estimators. In this paper, we proposed a novel mechanism to incorporate one-point gradient estimators in online setting that features an alternating update scheme, so that the online algorithm matches the performance of a two-point estimator based algorithm. We believe that such a design can inspire improved methods in the future for both online optimization and offline zeroth-order optimization.

3. Problem Formulation

In this paper, we study the online nonconvex optimization problem. Specifically, consider a sequence of functions $f_1, f_2, \ldots, f_t, \ldots$, which are assumed to be continuously differentiable. At each time t, a player is required to submit a decision variable x_t to the environment based on some information such as the gradients or function values of the previous functions $f_1, \ldots, f_{t-1}$. And then the oracle information of f_t is revealed to the player. In online convex optimization, the performance of the player is typically evaluated by the cumulative function value of the decision variables over time with comparison to that of the minimal point obtained with the full information of all functions. However, for a *nonconvex* functions, finding its global minimum is typically NP-hard, and hence such a minimal point may not be reasonable to serve as a baseline comparator. Instead, the following *local regret* defined with respect to the stationary points of nonconvex functions is widely adopted as a good performance metric for online nonconvex optimization (Hazan et al., 2017).

Definition 1 *The (dynamic) local regret of online nonconvex optimization is defined as:*

$$\mathfrak{R}_w(T) := \sum_{t=1}^{T} \|\nabla F_{t,w}(x_t)\|^2, \tag{1}$$

where $F_{t,w}(x)$ is the window-smoothed function with window size $w \in \{1, \ldots, T\}$ defined by

$$F_{t,w}(x) := \frac{1}{w} \sum_{i=t-w+1}^{t} f_i(x). \tag{2}$$

Note that in eq. (2), the function $f_i(x)$ is defined to be zero for all $i \leq 0$.

The above definition of the local regret in eq. (1) implicitly takes the comparator to be the zero gradient norm each time, and therefore it can also be interpreted as dynamic regret with the environment changing at each time. Further, the window-smoothed function in eq. (2) is necessarily adopted in the definition of the local regret, because it has been shown in Theorem 2.7 of (Hazan et al., 2017) that there exists a sequence of functions for which without window-smoothing (i.e., setting $w = 1$ in eq. (2)) sublinear regret is impossible to achieve.

Previous studies of the local regret on online nonconvex optimization required the access to *sufficient* gradient oracles at each time instance in order to design double-loop algorithms, so that the inner loop can iteratively update the decision variables to a target accuracy. Although the overall number of oracle queries is upper-bounded, there is no instantaneous limit on the oracle access. In this work, we focus on more challenging but practical settings where there are hard constraints on the number of oracles available at each time. In particular, we will study the following three settings

- **Window-smoothed single gradient oracle (Window-SGO) feedback:** At each time t, only the gradient of $F_{t,w}$ at a single variable is provided. Hence, an inner loop that updates the variable to a target accuracy is not allowed.
- **Window-smoothed single function value oracle (Window-SVO) feedback:** At each time t, only one function value of $F_{t,w}$ at a single variable is provided. Hence, an inner loop that updates the variable to a target accuracy is not allowed.
- **Window-smoothed multiple function value oracle (Window-MVO) feedback:** At each time t, the function values of $F_{t,w}$ at $2K$ variables are provided. Here, $K \in \mathbb{N}$ is a prescribed hard constraint threshold. Without loss of generality, we consider the number of oracles to be even for the presentation convenience.

4. Online Nonconvex Optimization with Window-SGO Feedback

To our best knowledge, all existing algorithms as in (Hazan et al., 2017; Aydore et al., 2019; Hallak et al., 2021) developed for achieving a good local regret in online nonconvex optimization require querying *sufficiently many* gradient oracles of $\nabla F_{t,w}(x)$ at each time t. However, many online learning systems may provide only a *single* gradient oracle (Window-SGO) feedback of $F_{t,w}(x)$ at each time t. It is thus interesting to investigate fundamentally what is the limit on the regret (i.e., a lower bound on the regret) for such a Window-SGO system, and what is the regret that an algorithm with Window-SGO feedback can achieve.

In this section, we first provide a regret lower bound for the Window-SGO setting that all algorithms satisfying the linear-span assumption will satisfy. Then, we show that the single-loop online gradient descent (OGD) algorithm indeed matches such a lower bound up to $\mathcal{O}(1)$.

We first specify the algorithm class that we consider.

Definition 2 (Linear-span (Nesterov, 2003)) *The online learning algorithm $\mathcal{A}$ generates a sequence of $\{x_t\}_{t=1}^{\infty}$ that satisfies*

$$x_{t+1} \in \left\{ x_1 + \sum_{i=1}^{t} a_i G_i(x_i) : a_i \in \mathbb{R}, i = 1, \ldots, t \right\},$$

where $G_i(\cdot)$ is either $\nabla F_{i,w}(\cdot)$ or its stochastic estimation.

The algorithm class defined above takes only Window-SGO feedback and includes various classes of first-order online methods such as online gradient descent, online accelerated gradient descent, online gradient descent with momentum, etc.

We next make the following standard assumptions on the objective functions.

Assumption 1 *For every $t = 1, 2, \ldots$, the objective function f_t is L_0-Lipschitz continuous and L_1-smooth on $\mathbb{R}^d$. Namely, for any $x, y \in \mathbb{R}^d$, we have*

$$|f_t(x) - f_t(y)| \leq L_0\|x - y\|,$$
$$\|\nabla f_t(x) - \nabla f_t(y)\| \leq L_1\|x - y\|.$$

Furthermore, $f_t(x)$ is bounded, i.e., for all $x \in \mathbb{R}^d$, we have $|f_t(x)| \leq M$.

Assumption 2 *For every $t = 1, 2, \ldots$, the stochastic estimations $G_t(x)$ of $\nabla F_{t,w}(x)$ are such that for all $x \in \mathbb{R}^d$,*

$$\mathbb{E}\left[G_t(x)\right] = \nabla F_{t,w}(x), \qquad \mathbb{E}\left[\|G_t(x) - \nabla F_{t,w}(x)\|^2\right] \leq \sigma^2.$$

Assumption 1 poses the smoothness and boundedness conditions on the functions f_t's, which are standard and have been widely adopted for online learning (Agarwal et al., 2010; Flaxman et al., 2005; Hallak et al., 2021; Hazan et al., 2017). Assumption 2 requires the stochastic gradient to be unbiased and possess a bounded variance σ^2. Any result assuming Assumption 2 can naturally specialize to the exact gradient feedback case (i.e., $G_t(x) = \nabla F_{t,w}(x)$) by setting $\sigma = 0$ and removing the expectation concerning the randomness of G_t.

The following theorem provides a lower bound on the regret that all algorithms with Window-SGO feedback defined in Definition 2 must satisfy.

Theorem 1 *Suppose that the algorithm $\mathcal{A}$ satisfies Definition 2. Also suppose the window-size satisfies $w \leq \mathcal{O}\left(T^\alpha\right)$ with $0 \leq \alpha \leq 1$, and let the constant $\gamma \in (\frac{1}{\alpha}, \infty)$. Then, there exist $\{f_t\}_{t=1}^\infty$ and $\{G_t\}_{t=1}^\infty$ satisfying Assumptions 1 and 2, respectively, for which*

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \geq \Omega\left(\frac{T}{w^\gamma}\right) + \Omega\left(\frac{\sigma T}{w^{\frac{\gamma}{2}}}\right).$$

Discussion. To elaborate, consider the exact gradient feedback case (i.e., $\sigma = 0$). It is interesting to observe that if $1/2 \leq \alpha \leq 1$, then Theorem 1 implies that $\mathfrak{R}_w(T) \geq \Omega\left(\frac{T}{w^\gamma}\right)$ holds for any $\gamma > 2$. This means that for any such window sizes, the regret of $\mathcal{O}\left(\frac{T}{w^2}\right)$ is impossible to achieve by Window-SGO algorithms with exact gradient. For the stochastic gradient feedback case (i.e., $\sigma > 0$), such a window further expands to $1/4 \leq \alpha \leq 1$, since $\mathbb{E}\left[\mathfrak{R}_w(T)\right] \geq \Omega\left(\frac{\sigma T}{w^{\frac{\gamma}{2}}}\right)$ holds for any $\gamma > 4$. However, it has been shown that algorithms with sufficiently many gradient oracles (Hazan et al., 2017; Hallak et al., 2021) can achieve the regret of $\mathcal{O}\left(\frac{T}{w^2}\right)$ for any window size. Such a comparison indicates that *online nonconvex optimization with Window-SGO feedback is provably more challenging (i.e., has a provably higher local regret) than that with window-smoothed multiple gradient oracle (Window-MGO) feedback.*

Further, the window size w is often required to take the values of $\Theta(T)$ in practice, for example, utilizing the regret bound to achieve the state-of-the-art computational complexity in nonconvex offline optimization (Hazan et al., 2017; Hallak et al., 2021). In such a case (i.e., $w = \mathcal{O}(T)$),

Algorithm 1 Online Gradient Descent (OGD)

Input: Initial point x_1, stepsizes η
for $t = 1, \ldots, T$ **do**
  $x_{t+1} = x_t - \eta G_t(x_t)$
end for

the lower bound with Window-MGO feedback is $\Omega(\frac{1}{T})$, which is attained by the adaptive Window-MGO algorithm proposed by Hazan et al. (2017). As a comparison, Theorem 1 shows that algorithms with Window-SGO have a much worse regret lower bound of $\Omega(1)$ with exact gradient feedback and of $\Omega(\sqrt{T})$ with stochastic gradient feedback. The following corollary captures the worst-case lower bound that holds uniformly for all choices of window size, which we call as the window-unconditioned lower bound.

Corollary 1 *Suppose the window-size is chosen arbitrarily from* 1 *to* T *and the sequences of functions* $\{f_t\}_{t=1}^{\infty}$ *satisfy Assumption 1. Consider an algorithm* $\mathcal{A}$ *that satisfies Definition 2.*

(i). *For Window-SGO with exact gradient feedback, there exist objective functions* $\{f_t\}_{t=1}^{\infty}$*, for which*

$$\mathfrak{R}_w(T) \geq \Omega\left(\frac{T}{w}\right).$$

(ii). *For Window-SGO with stochastic gradient feedback, there exist objective functions* $\{f_t\}_{t=1}^{\infty}$ *and stochastic gradients* $\{G_t\}_{t=1}^{\infty}$ *satisfying Assumption 2, for which*

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \geq \Omega\left(\frac{\sigma T}{\sqrt{w}}\right).$$

Corollary 1 further illustrates the hardness of the Window-SGO setting. We next show that as a natural Window-SGO algorithm satisfying Definition 2, the online gradient descent (OGD) algorithm achieves a sublinear regret, which matches the window-unconditioned lower bound. The OGD algorithm with window size w is provided in Algorithm 1, where at each time t, the next variable x_{t+1} is updated based on a single Window-SGO feedback.

The following theorem characterizes the regret of OGD for online nonconvex optimization.

Theorem 2 *Consider Algorithm 1 with* $\eta \leq \frac{1}{2L_1}$*. Suppose Assumptions 1 and 2 hold. Then the regret satisfies:*

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \leq \mathcal{O}\left(\frac{T}{w\eta}\right) + \mathcal{O}\left(\eta\sigma^2 T\right). \tag{3}$$

The regret upper bound in Theorem 2 consists of two terms: the convergence error of x_t to the stationary points of $F_{t,w}(x)$ captured by $\mathcal{O}\left(\frac{T}{w\eta}\right)$, and the variance of the stochastic gradients captured by $\mathcal{O}\left(\eta\sigma^2 T\right)$. By choosing the stepsize η properly in Theorem 2, we obtain the best regret upper bounds for OGD in the Window-SGO setting with exact and stochastic gradient oracles, respectively.

Corollary 2 *Consider the same setting as Theorem 2.*

(i). When given access to exact gradients, i.e., $\sigma = 0$, by taking $\eta = \frac{1}{2L_1}$, the regret of Algorithm 1 satisfies

$$\mathfrak{R}_w(T) \leq \mathcal{O}\left(\frac{T}{w}\right).$$

(ii). When given access to stochastic gradients, i.e., $\sigma > 0$, by taking $\eta = \min\left\{\frac{1}{2L_1}, \frac{1}{\sigma\sqrt{w}}\right\}$, the regret of Algorithm 1 satisfies

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \leq \mathcal{O}\left(\frac{\sigma T}{\sqrt{w}}\right).$$

Clearly, the above regret upper bound matches the lower bound in Corollary 1, which indicates that OGD is optimal for online nonconvex optimization with Window-SGO feedback.

5. Online Nonconvex Optimization with Window-SVO Feedback

In this section, we study the online nonconvex optimization problem with the player having only a single function value oracle (Window-SVO) feedback at each time instance. Such a Window-SVO scenario is much more challenging than Window-SGO, because with only a single function value, the gradient of the function cannot be estimated accurately. Although there are single-point gradient estimators available in the existing research of offline optimization problems (Flaxman et al., 2005; Dekel et al., 2015; Gasnikov et al., 2017; Zhang et al., 2022), applying these estimators straightforwardly cannot yield the desired regret for online nonconvex optimization as we elaborate in Section 5.1. Hence, we are motivated to develop a novel approach for Window-SVO feedback and establish its regret performance in online nonconvex optimization.

5.1. Straightforward Method based on Offline Optimization with Window-SVO Feedback

In the existing literature of offline optimization, two main gradient estimators with Window-SVO (i.e., one-point gradient estimators) have been proposed. In this section, we explain that straightforward applications of these estimators to online optimization do not yield desirable regret bounds.

First, the following one-point gradient estimator (Flaxman et al., 2005) was commonly used in offline gradient-free algorithms and in online convex optimization (Agarwal et al., 2010):

$$g_t \coloneqq \frac{d}{\delta} f(x_t + \delta u_t) u_t, \tag{4}$$

where $\delta > 0$ is a perturbation hyperparameter and u_t is sampled from the unit surface uniformly at random. By applying such an estimator to replace the window gradient in Algorithm 1 of the single-point OGD algorithm, we can show that the resulting algorithm achieves a regret bound of $\mathcal{O}\left(\frac{dT}{w^{1/3}}\right)$ (see Appendix B for the detailed proof). Such a bound has a clear gap in terms of the order dependency on w compared to the lower bound for the Window-SGO setting. This is mainly because the gradient estimator in eq. (4) has a large bias.

More recently, the following one-point gradient estimator was proposed in (Zhang et al., 2022) by incorporating the residual feedback (called OR), which is given below.

$$g_t \coloneqq \frac{d}{\delta}\left(f(x_t + \delta u_t) - f(x_{t-1} + \delta u_{t-1})\right) u_t,$$

which mimics the two-point gradient estimator by further subtracting the function value at the perturbation of the previous point in order to reduce the estimation bias. Applying the above estimator to online nonconvex optimization yields the following window-smoothed gradient estimator:

$$\bar{\nabla} F_{t,w}(x_t) = \frac{d}{\delta} \left(F_{t,w}(x_t + \delta u_t) - F_{t-1,w}(x_{t-1} + \delta u_{t-1}) \right) u_t, \tag{5}$$

which can be shown to have the following variance bound

$$\mathbb{E}\left[\left\| \bar{\nabla} F_{t,w}(x_t) \right\|^2 \right] \leq \frac{3d^2\eta^2 L_0^2}{\delta^2} \mathbb{E}\left[\left\| \bar{\nabla} F_{t-1,w}(x_{t-1}) \right\|^2 \right] + \frac{12M^2d^2}{w^2\delta^2} + 12d^2L_0^2. \tag{6}$$

To control the iterative upper bound of the variance in eq. (6), we need to set $\eta = \mathcal{O}(\frac{\delta}{d})$. Otherwise the bound explodes exponentially fast. Such a requirement of η slows down the convergence speed of x_t to a stationary point, since the increment in each iteration satisfies $\|x_{t+1} - x_t\| \leq \mathcal{O}(\eta)$. Indeed, by setting $\eta = \frac{1}{3L_0 d\sqrt{w}}$ and $\delta = \frac{1}{\sqrt{w}}$, we can obtain the following optimized regret bound by leveraging the convergence of the gradient descent update of x_t and the variance bound in eq. (6) (see Appendix C for the detailed proof).

$$\mathbb{E}\left[\mathfrak{R}_w(T) \right] \leq \mathcal{O}\left(\frac{dT}{\sqrt{w}} \right). \tag{7}$$

The above regret bound improves the regret bound of $\mathcal{O}\left(\frac{dT}{w^{1/3}}\right)$ for the algorithm based on the one-point estimator in eq. (4). Intuitively, this is because by iteratively applying eq. (6), the variance of $\bar{\nabla} F_{t,w}(x_t)$ can be shown to be bounded by an improved dependency of $\mathcal{O}(d^2)$.

However, there is still a performance gap in terms of the dependency on w compared to the regret lower bound of Window-SGO with exact gradient in Corollary 1, which indicates that the best upper bound can scale with $\mathcal{O}(\frac{1}{w})$. Hence, it is interesting to explore if online nonconvex optimization with Window-SVO feedback can match the regret lower bound with exact gradient in terms of the dependency on w.

5.2. A Novel Window-SVO Scheme with Improved Regret

In this subsection, we propose a novel online design that improves the method discussed in the previous subsection based on the straightforward use of the OR estimator in eq. (5). The main drawback there is that both function values in the gradient estimator in eq. (5) are evaluated at perturbed points, which causes a large variance. To address this issue, we propose the following new one-point residual gradient estimator.

$$\widehat{\nabla} F_{t,w}(x_t) := \frac{d}{\delta} \left(F_{t,w}(x_t + \delta u_t) - F_{t-1,w}(x_t) \right) u_t, \tag{8}$$

where we subtract the previous oracle at the unperturbed current point x_t. As we will show in Lemma 1 that such an estimator indeed has a much smaller variance. On the other hand, since such a gradient estimator evaluates $F_{t,w}$ and $F_{t-1,w}$ respectively at different points $x_t + \delta u_t$ and x_t, we need two iterations to generate one gradient estimator and hence one gradient descent update. This leads to an algorithm design with **skipped updates** as shown in Algorithm 2 (named SkipOGD). Specifically, we query the function value of x_t in the odd iterations and that of $x_t + \delta u_t$ in the

Algorithm 2 OGD with Skipped One-point Residual Estimation (SkipOGD)

Input: Initial point $x_1 \in \mathbb{R}^d$, stepsizes η, estimation radius δ
for $t = 1, \ldots, T-1$ **do**
  if $t = 2j + 1$ **then**
    Observe $F_{t,w}(x_t)$
    $x_{t+1} = x_t$
  else
    Drawn u_t uniformly from $\mathcal{S}^{d-1}$ independently
    Observe $F_{t,w}(x_t + \delta u_t)$
    Estimate gradient based on Equation (8)
    $x_{t+1} = x_t - \eta \widehat{\nabla} F_{t,w}(x_t)$
  end if
end for

even iterations to construct the gradient estimation in eq. (8). Consequently, such an algorithm design skips one variable update for every two iterations. It turns out for those odd time instances, although their variable updates are skipped, the sliding window functions still track the desirable variable good enough such that the overall regret performance of SkipOGD is better than the original algorithm that takes one update in every iteration using the OR estimator in eq. (5). Essentially, this implies that the variance reduction due to skipped updates outweighs the all-time update scheme with the OR estimator, leading to the match of the regret lower bound in the dependency on w as we show in Corollary 3.

The following lemma captures the advantage of the gradient variance of our estimator in eq. (8) over the OR estimator in eq. (5).

Lemma 1 *Suppose Assumption 1 holds. For every even t, we have*

$$\mathbb{E}\left[\left\|\widehat{\nabla} F_{t,w}(x_t)\right\|^2 \middle| \mathcal{F}_t\right] \leq 3d \left\|\nabla F_{t,w}(x_t)\right\|^2 + \frac{12M^2d^2}{w^2\delta^2} + L_1^2 d^2 \delta^2, \tag{9}$$

where $\mathcal{F}_t$ is the filtration defined as $\mathcal{F}_t \coloneqq \sigma\left(u_2, u_4, \ldots, u_{t-2}\right)$.

Comparing the above variance bound with that of the OR estimation in (6), the variance of our estimator no longer depends on the previous updates and therefore does not pose any restriction on the stepsize η. Thus, such a new variance bound allows a better flexibility of choosing a larger stepsize, and can hence lead to faster convergence intuitively. The next theorem formally establishes an improved regret bound of our proposed SkipOGD.

Theorem 3 *Under Assumption 1, Algorithm 2 with $\eta = \frac{1}{3L_1 d}$ satisfies the following regret bound.*

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \leq \mathcal{O}\left(\frac{dT}{w}\right) + \mathcal{O}\left(\frac{dT}{w^2\delta^2}\right) + \mathcal{O}(d\delta^2 T).$$

The above regret bound captures three different algorithmic statistics: (a) the term $\mathcal{O}(\frac{dT}{w})$ captures the convergence error of x_t to the stationary points of $F_{t-1,w}(x)$; (b) the term $\mathcal{O}(d\delta^2 T)$ captures the gradient estimation bias; and (c) the term $\mathcal{O}\left(\frac{dT}{w^2\delta^2}\right)$ captures the variance of gradient estimation.

We note that the first and third terms in the previous variance bound (9) are absorbed into the convergence error and estimation bias, respectively, as they take the same order. In particular, the proof of the above regret bound critically leverages the variance bound in Lemma 1 and incorporates it into the convergence analysis for nonconvex optimization.

Choosing proper hyperparameters in the above theorem, we further obtain the following corollary that indicates that the regret of our new algorithm SkipOGD matches the regret lower bound with exact gradient oracles in (i) of Corollary 1 on the dependency of w and T. The gap of a factor of $\mathcal{O}(d)$ is due to the bandit feedback which is somewhat unavoidable even for online convex optimization.

Corollary 3 *Let $\eta = \frac{1}{3L_1 d}$ and $\delta = \frac{1}{\sqrt{w}}$. Theorem 3 implies that Algorithm 2 has the regret bound*

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \leq \mathcal{O}\left(\frac{dT}{w}\right).$$

6. Online Nonconvex Optimization with Window-MVO Feedback

In this section, we study online nonconvex optimization under the setting where up to $2K$ ($K \geq 1$) function value oracles can be queried at each time t. In such a setting with window-smoothed multiple function value oracle (Window-MVO) queries, the regret can be improved in two aspects: (a) better gradient estimation based on two-point estimators; and (b) an inner-loop update of the variable to a better tracking of a stationary point of $F_{t,w}(x)$.

Regarding the first aspect, our design of Algorithm 3 adopts the following standard two-point gradient estimator:

$$\widetilde{\nabla} F_{t,w}(x) = \frac{d}{\delta}\left(F_{t,w}(x+\delta u) - F_{t,w}(x)\right) u, \tag{10}$$

where u is drawn uniformly from $\mathcal{S}^{d-1}$. The following lemma captures the variance bound of the two-point gradient estimator.

Lemma 2 *Suppose Assumption 1 hold. For every even t, we have*

$$\mathbb{E}\left[\left\|\widetilde{\nabla} F_{t,w}(y_t^k)\right\|^2 \middle| \mathcal{F}_t^k\right] \leq 2d\left\|\nabla F_{t,w}(y_t^k)\right\|^2 + \frac{L_1^2 d^2 \delta^2}{2}, \tag{11}$$

where $\mathcal{F}_t^k$ is the filtration defined as $\mathcal{F}_t^k \coloneqq \sigma\left(u_t^1, u_t^2, \ldots, u_t^{k-2}, u_t^{k-1}\right)$.

Compared with Lemma 1 for the one-point gradient estimator $\widehat{\nabla} F_{t,w}(x_t)$, the variance of $\widetilde{\nabla} F_{t,w}(y_t^k)$ in Lemma 2 not only has smaller constants, but also is exempted from having the $\frac{12M^2d^2}{w^2\delta^2}$ term caused by the difference between adjacent $F_{t,w}(x)$.

Regarding the second aspect, we adopt a double-loop structure in Algorithm 3 (and hence our algorithm is called LoopOGD), where the inner loop takes K updates due to the hard oracle constraint. Such an algorithm is different from the existing double-loop algorithms for online nonconvex optimization (Hazan et al., 2017; Hallak et al., 2021), where the update steps in each inner loop there do not satisfy a hard constraint on the number of oracle queries. Hence, our regret analysis is also different, because our inner loop does not guarantee to output a variable with a target accuracy to a stationary point. The following theorem provides the regret bound for Algorithm 3 based on the two-point gradient estimator.

Algorithm 3 OGD with Looped Two-Point Estimation (LoopOGD)

Input: Initial point $x_1 \in \mathbb{R}^d$, stepsizes η, constant ξ, estimation radius δ
for $t = 1, \ldots, T-1$ **do**
  $y_t^0 \leftarrow y_{t-1}^K$ ($y_1^0 \leftarrow x_1$)
  for $k = 0, \ldots, K-1$ **do**
    Drawn u_t^k from $\mathcal{S}^{d-1}$ uniformly at random
    Observe $F_{t,w}(y_t^k)$ and $F_{t,w}(y_t^k + \delta u_t^k)$
    Estimate $\widetilde{\nabla} F_{t,w}(y_t^k)$ based on eq. (10)
    $y_t^{k+1} = y_t^k - \eta \widetilde{\nabla} F_{t,w}(y_t^k)$
  end for
  Pick I_t from $0, 1, \ldots, K-1$ uniformly at random
  $x_{t+1} = y_t^{I_t}$
end for

Theorem 4 *Suppose Assumption 1 holds. Consider Algorithm 3. Let* $\eta = \frac{1}{2L_1 d}$. *We have*

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \leq \mathcal{O}\left(\frac{dT}{wK}\right) + \mathcal{O}\left(\frac{T}{w^2}\right) + \mathcal{O}\left(d\delta^2 T\right). \tag{12}$$

The above regret bound consists of three terms: (a) the term $\mathcal{O}\left(\frac{dT}{wK}\right)$ that captures the convergence error of x_t to the stationary points of $F_{t-1,w}(x)$; (b) the term $\mathcal{O}(d\delta^2 T)$ that captures the gradient estimation bias; and (c) the term $\mathcal{O}\left(\frac{T}{w^2}\right)$ that captures the error caused by the time-variation of the window function $F_{t,w}(x)$. Note that the first and second term in eq. (11) are absorbed into the above convergence error and gradient estimation bias terms, respectively, as they are of the same order.

Corollary 4 *Let* $\eta = \frac{1}{2L_1 d}$ *and* $\delta = \frac{1}{\sqrt{dw}}$. *Theorem 4 implies that Algorithm 3 has the following regret bound:*

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \leq \mathcal{O}\left(\frac{dT}{wK}\right) + \mathcal{O}\left(\frac{T}{w^2}\right).$$

Corollary 4 has the following implications. (i). By taking $K = 1$, Algorithm 3 becomes a single-loop algorithm (i.e., the OGD algorithm with gradients replaced by their two-point estimators), which matches the regret lower bound of the Window-SGO with exact gradient feedback up to a gap of $\mathcal{O}(d)$ (due to the bandit feedback). Interestingly, our one-point single-loop Algorithm 2 even matches the regret of such a two-point algorithm up to a constant. (ii). By taking $K = wd$, Algorithm 3 matches the existing regret lower bound (up to a gap of $\mathcal{O}(d)$ due to the bandit feedback) for online nonconvex optimization with window-smoothed multiple gradient oracle (Window-MGO) feedback given in (Hazan et al., 2017). Compared with the fact that the first-order algorithms need the number $\mathcal{O}(w)$ of exact gradient oracles on average per time instance (Hazan et al., 2017; Hallak et al., 2021), Algorithm 3 requires $\mathcal{O}(d)$ times more accesses to the function value oracles, which reflects the typical relationship between gradient and function value feedback; And (iii). By setting K between 1 and wd, Algorithm 3 provides a full spectrum of the tradeoff between the regret $\mathcal{O}(\frac{dT}{wK})$ and the computational complexity $\mathcal{O}(KT)$.

7. Conclusion

In this paper, we study online nonconvex optimization with multiple types of limited instantaneous oracle feedback, including the Window-SGO, Window-SVO and Window-MVO settings, which are different from the existing studies that allows sufficiently many Window-MGOs at each time. We develop tight lower regret bounds in the Window-SGO setting and show that the classic OGD algorithm achieves the lower bound. We further develop novel gradient estimators and algorithm schemes in the more challenging Window-SVO and Window-MVO settings and show that they achieve near-optimal local regrets. We anticipate that the new one-point gradient estimator that we devise as well as our design of online algorithms with the skipping mechanism can be useful more generally in online learning with hard instantaneous constraints on oracle queries.

Acknowledgments

The work of Ziwei Guan and Yingbin Liang was supported in part by the U.S. National Science Foundation under the grants CCF-1909291 and RINGS-2148253. The work of Yi Zhou was supported in part by U.S. National Science Foundation under the grants CCF-2106216, DMS-2134223 and CAREER-2237830.

References

Alekh Agarwal, Ofer Dekel, and Lin Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Proc. Annual Conference on Learning Theory (COLT)*, pages 28–40, 2010.

Naman Agarwal, Alon Gonen, and Elad Hazan. Learning in non-convex games with an optimization oracle. In *Proc. Annual Conference on Learning Theory (COLT)*, pages 18–29. PMLR, 2019.

Yossi Arjevani, Yair Carmon, John C Duchi, Dylan J Foster, Nathan Srebro, and Blake Woodworth. Lower bounds for non-convex stochastic optimization. *Mathematical Programming*, pages 1–50, 2022.

Sergul Aydore, Tianhao Zhu, and Dean P Foster. Dynamic local regret for non-convex online forecasting. *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019.

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations research*, 63(5):1227–1244, 2015.

Xuanyu Cao and K. J. Ray Liu. On the time-varying constraints and bandit feedback of online convex optimization. In *IEEE International Conference on Communications (ICC)*, pages 1–6, 2018. doi: 10.1109/ICC.2018.8422559.

Xuanyu Cao and K. J. Ray Liu. Online convex optimization with time-varying constraints and bandit feedback. *IEEE Transactions on Automatic Control*, 64(7):2665–2680, 2019. doi: 10.1109/TAC.2018.2884653.

Yair Carmon, John C Duchi, Oliver Hinder, and Aaron Sidford. Lower bounds for finding stationary points I. *Mathematical Programming*, 184(1):71–120, 2020.

Ofer Dekel, Ronen Eldan, and Tomer Koren. Bandit smooth convex optimization: Improving the bias-variance tradeoff. *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 28, 2015.

Abraham D. Flaxman, Adam Tauman Kalai, and H. Brendan McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *In Proc.ACM-SIAM Symposium on Discrete Algorithms (SODA)*, page 385–394, 2005.

Xiand Gao, Xiaobo Li, and Shuzhong Zhang. Online learning with non-convex losses and non-stationary regret. In *Proc. International Conference on Artifical Intelligence and Statistics (AISTATS)*, pages 235–243. PMLR, 2018.

Alexander V Gasnikov, Ekaterina A Krymova, Anastasia A Lagunovskaya, Ilnura N Usmanova, and Fedor A Fedorenko. Stochastic online optimization. single-point and multi-point non-linear multi-armed bandits. convex and strongly-convex case. *Automation and remote control*, 78:224–234, 2017.

Nadav Hallak, Panayotis Mertikopoulos, and Volkan Cevher. Regret minimization in stochastic non-convex learning via a proximal-gradient approach. In *Proc. International Conference on Machine Learning (ICML)*, pages 4008–4017. PMLR, 2021.

Shuji Hao, Peilin Zhao, Yong Liu, Steven CH Hoi, and Chunyan Miao. Online multitask relative similarity learning. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, volume 25, pages 1823–1829, 2017.

Elad Hazan and Satyen Kale. Beyond the regret minimization barrier: an optimal algorithm for stochastic strongly-convex optimization. In *Proc. Annual Conference on Learning Theory (COLT)*, pages 421–436. JMLR Workshop and Conference Proceedings, 2011.

Elad Hazan, Karan Singh, and Cyril Zhang. Efficient regret minimization in non-convex games. In *Proc. International Conference on Machine Learning (ICML)*, pages 1433–1441. PMLR, 2017.

Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

Amélie Héliou, Matthieu Martin, Panayotis Mertikopoulos, and Thibaud Rahier. Online non-convex optimization with imperfect feedback. *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 33:17224–17235, 2020.

Amélie Héliou, Matthieu Martin, Panayotis Mertikopoulos, and Thibaud Rahier. Zeroth-order non-convex learning via hierarchical dual averaging. In *Proc. International Conference on Machine Learning (ICML)*, pages 4192–4202. PMLR, 2021.

Ali Jadbabaie, Alexander Rakhlin, Shahin Shahrampour, and Karthik Sridharan. Online optimization: Competing with dynamic comparators. In *Proc. International Conference on Artifical Intelligence and Statistics (AISTATS)*, pages 398–406. PMLR, 2015.

Yeongjong Kim and Dabeen Lee. Online convex optimization with stochastic constraints: Zero constraint violation and bandit feedback. *arXiv preprint arXiv:2301.11267*, 2023.

Walid Krichene, Maximilian Balandat, Claire Tomlin, and Alexandre Bayen. The hedge algorithm on a continuum. In *International Conference on Machine Learning*, pages 824–832. PMLR, 2015.

Antoine Lesage-Landry, Joshua A Taylor, and Iman Shames. Second-order online nonconvex optimization. *IEEE Transactions on Automatic Control*, 66(10):4866–4872, 2020.

Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2003.

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

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

Alexander Rakhlin, Ohad Shamir, and Karthik Sridharan. Making gradient descent optimal for strongly convex stochastic optimization. *Proc. International Conference on Machine Learning (ICML)*, 2011.

Tim Roughgarden and Okke Schrijvers. Online prediction with selfish experts. *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 30, 2017.

Ankan Saha and Ambuj Tewari. Improved regret guarantees for online smooth convex optimization with bandit feedback. In *Proc. International Conference on Artifical Intelligence and Statistics (AISTATS)*, pages 636–642. JMLR Workshop and Conference Proceedings, 2011.

Ohad Shamir. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *Journal of Machine Learning Research (JMLR)*, 18(1):1703–1713, 2017.

Ohad Shamir and Tong Zhang. Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes. In *Proc. International Conference on Machine Learning (ICML)*, pages 71–79. PMLR, 2013.

Arun Sai Suggala and Praneeth Netrapalli. Online non-convex learning: Following the perturbed leader is optimal. In *Algorithmic Learning Theory*, pages 845–861. PMLR, 2020.

Cong Wang, Shengyuan Xu, Deming Yuan, Baoyong Zhang, and Zhengqiang Zhang. Distributed online convex optimization with a bandit primal-dual mirror descent push-sum algorithm. *Neurocomputing*, 497:204–215, 2022.

Lin Yang, Lei Deng, Mohammad H Hajiesmaili, Cheng Tan, and Wing Shing Wong. An optimal algorithm for online non-convex learning. *Proc. of the ACM on Measurement and Analysis of Computing Systems*, 2(2):1–25, 2018.

Xinlei Yi, Xiuxian Li, Tao Yang, Lihua Xie, Tianyou Chai, and Karl Henrik Johansson. Distributed bandit online convex optimization with time-varying coupled inequality constraints. *IEEE Transactions on Automatic Control*, 66(10):4620–4635, 2020.

Lijun Zhang, Shiyin Lu, and Zhi-Hua Zhou. Adaptive online learning in dynamic environments. *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 31, 2018.

Yan Zhang, Yi Zhou, Kaiyi Ji, and Michael M Zavlanos. A new one-point residual-feedback oracle for black-box learning and control. *Automatica*, 136:110006, 2022.

Peng Zhao, Yu-Jie Zhang, Lijun Zhang, and Zhi-Hua Zhou. Dynamic regret of convex and smooth functions. *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 33:12510–12520, 2020.

Peng Zhao, Yu-Jie Zhang, Lijun Zhang, and Zhi-Hua Zhou. Adaptivity and non-stationarity: Problem-dependent dynamic regret for online convex optimization. *arXiv preprint arXiv:2112.14368*, 2021.

Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proc. International Conference on Machine Learning (ICML)*, pages 928–936, 2003.

Appendix A. Proofs of Section 4

A.1. Proof of Theorem 1

We develop the proof with two cases, i.e., $\sigma = 0$ and $\sigma > 0$. When $\sigma = 0$, we aim to show $\mathfrak{R}_w(T) \geq \Omega\left(\frac{T}{w^\gamma}\right)$; and when $\sigma > 0$, $\Omega\left(\frac{\sigma T}{w^{\frac{\gamma}{2}}}\right)$ is the dominating term and we want to show that $\mathbb{E}\left[\mathfrak{R}_w(T)\right] \geq \Omega\left(\frac{\sigma T}{w^{\frac{\gamma}{2}}}\right)$.

Case (a). $\sigma = 0$: We prove $\mathfrak{R}_w(T) \geq \Omega\left(\frac{T}{w^\gamma}\right)$ by contradiction. Suppose $\mathfrak{R}_w(T) \leq \mathcal{O}\left(\frac{T}{w^\gamma}\right)$. Then, there exists a constant $c > 0$ such that $\mathfrak{R}_w(T) \leq \frac{cT}{w^\gamma}$ for all T. Moreover, we let $w = CT^\alpha$ for some $0 < C < 1$. Then, let $f_t(x) = f(x)$ for all t with $f(x)$ satisfying the smoothness and boundedness conditions in Assumption 1. Hence, $F_{t,w}(x) = f(x)$ for all $t \geq w$. By the definition of the local regret, we have that

$$\frac{1}{T-w}\sum_{t=w+1}^{T}\|\nabla f(x_t)\|^2 \leq \frac{1}{T-w}\sum_{t=1}^{T}\|\nabla f(x_t)\|^2 = \frac{\mathfrak{R}_w(T)}{T-w} \leq \frac{c}{C^\gamma T^{\alpha\gamma}(1-CT^{\alpha-1})}. \tag{13}$$

Next, we observe that

$$\inf_{t=1,\ldots,T}\|\nabla f(x_t)\| \overset{(i)}{\leq} \sqrt{\frac{\sum_{t=w+1}^{T}\|\nabla f(x_t)\|^2}{T-w}} \overset{(ii)}{\leq} \frac{\sqrt{c}}{\sqrt{C^\gamma(1-C)}T^{\frac{\alpha\gamma}{2}}}, \tag{14}$$

where (i) follows because

$$\inf_{t=1,\ldots T}\|\nabla f(x_t)\|^2 \leq \inf_{t=w+1,\ldots T}\|\nabla f(x_t)\|^2 \leq \frac{\sum_{t=w+1}^{T}\|\nabla f(x_t)\|^2}{T-w},$$

and (ii) follows from eq. (13) and $1 - CT^{\alpha-1} \geq 1 - C$.

Thus, for all $T \geq \frac{c^{\frac{1}{\alpha\gamma}}}{C^{\frac{1}{\alpha}}(1-C)^{\frac{1}{\alpha\gamma}}\epsilon^{\frac{2}{\alpha\gamma}}}$ where $\epsilon > 0$ is a given target accuracy, we have

$$\inf_{t=1,\ldots,T}\|\nabla f(x_t)\| \leq \epsilon. \tag{15}$$

Since $F_{t,w}(x) = f(x)$ for all t and w, $f(x)$ can be viewed as the objective function of an offline nonconvex optimization problem. For such a problem, we take the standard *computational complexity* to achieve an ϵ-accurate stationary point as

$$\mathsf{T}_\epsilon(\mathcal{A}, f) \coloneqq \inf\{t \in \mathbb{N} : \|\nabla f(x_t)\| \leq \epsilon\},$$

where $x_t = \mathcal{A}(x_1, \{\nabla f(x_i)\}_{i=1}^{t-1})$. Then the condition on T to obtain eq. (15) implies that

$$\mathsf{T}_\epsilon(\mathcal{A}, f) \leq \mathcal{O}\left(\frac{1}{\epsilon^{\frac{2}{\alpha\gamma}}}\right) = o\left(\frac{1}{\epsilon^2}\right). \tag{16}$$

The above bound contradicts the lower bound on $\mathsf{T}_\epsilon(\mathcal{A}, f)$ given in Theorems 1 and 2 of (Carmon et al., 2020) with $p = 1$ for offline nonconvex optimization, which requires $\mathsf{T}_\epsilon(\mathcal{A}, f) \geq \Omega(\frac{1}{\epsilon^2})$. This completes the proof.

Case (*b*). $\sigma > 0$: We also prove by contradiction. Suppose there exists an algorithm satisfying Definition 2 and $\mathbb{E}\left[\mathfrak{R}_w(T)\right] \leq \mathcal{O}(\frac{\sigma T}{w^{\frac{\gamma}{2}}})$. Then, there exists a constant $c > 0$ such that $\mathbb{E}\left[\mathfrak{R}_w(T)\right] \leq \frac{c\sigma T}{w^{\frac{\gamma}{2}}}$ for all T. Moreover, let $w = CT^{\alpha}$ for some $0 < C < 1$ and set $f_t(x) = f(x)$ for some $f(x)$ satisfying the smoothness and boundedness conditions in Assumption 1. Then, we have $F_{t,w}(x) = f(x)$ for all $t \geq w$. Let $\{G_t(x)\}_{t=1}^{\infty}$ be a sequence of independent random functions over t and satisfy Assumption 2. We further define the filtration $\mathcal{F}_t = \sigma\left(G_1, \ldots, G_{t-1}\right)$. Then we obtain that

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\sum_{t=w+1}^{T}\left\|\nabla f(x_t)\right\|^2}{T-w}\right] \\
&= \mathbb{E}\left[\frac{\sum_{t=w+1}^{T}\mathbb{E}\left[\left\|\nabla f(x_t)\right\|^2\middle|\mathcal{F}_t\right]}{T-w}\right] \leq \mathbb{E}\left[\frac{\sum_{t=w+1}^{T}\mathbb{E}\left[\left\|\nabla G_t(x_t)\right\|^2\middle|\mathcal{F}_t\right]}{T-w}\right] \\
&= \mathbb{E}\left[\frac{\sum_{t=w+1}^{T}\left\|\nabla G_t(x_t)\right\|^2}{T-w}\right] \leq \mathbb{E}\left[\frac{\sum_{t=1}^{T}\left\|\nabla G_t(x_t)\right\|^2}{T-w}\right] = \frac{\mathbb{E}\left[\mathfrak{R}_w(T)\right]}{T-w} \leq \frac{c\sigma}{C^{\frac{\gamma}{2}}(1-C)T^{\frac{\alpha\gamma}{2}}}.
\end{aligned} \tag{17}
$$

Here, eq. (17) implies that

$$
\inf_{t=1,\ldots,T}\mathbb{E}\left[\left\|\nabla f(x_t)\right\|\right] \leq \sqrt{\inf_{t=w+1,\ldots,T}\mathbb{E}\left[\left\|\nabla f(x_t)\right\|^2\right]} \leq \frac{\sqrt{c\sigma}}{C^{\frac{\gamma}{4}}(1-C)^{\frac{1}{2}}T^{\frac{\alpha\gamma}{4}}}. \tag{18}
$$

Equation (18) implies that $\inf_{t=1,\ldots,T}\mathbb{E}\left[\left\|\nabla f(x_t)\right\|\right] \leq \epsilon$ for all $\epsilon > 0$ when $T \geq \frac{(c\sigma)^{\frac{2}{\alpha\gamma}}}{C^{\frac{1}{\alpha}}(1-C)^{\frac{2}{\alpha\gamma}}\epsilon^{\frac{4}{\alpha\gamma}}}$.

Since $F_{t,w}(x) = f(x)$ for all t and w, $f(x)$ can be viewed as the objective function of an offline nonconvex optimization problem solved via a stochastic gradient algorithm. For such a problem, we take the standard *computational complexity* to achieve an ϵ-accurate stationary point as $\mathfrak{m}_\epsilon := \inf\left\{t \in \mathbb{N} | \mathbb{E}\left[\left\|\nabla f(x_t)\right\|\right] \leq \epsilon\right\}$. Then we have that

$$
\mathfrak{m}_\epsilon \leq \frac{(c\sigma)^{\frac{2}{\alpha\gamma}}}{C^{\frac{1}{\alpha}}(1-C)^{\frac{2}{\alpha\gamma}}\epsilon^{\frac{4}{\alpha\gamma}}}.
$$

If $\gamma > \frac{1}{\alpha}$, the above inequality implies that $\mathfrak{m}_\epsilon \leq o\left(\frac{\sigma^2}{\epsilon^4}\right)$, which contradicts the lower bound provided in (Arjevani et al., 2022, Theorem 3) for first-order stochastic algorithms for offline nonconvex optimization, which requires $\mathfrak{m}_\epsilon \geq \Omega\left(\frac{\sigma^2}{\epsilon^4}\right)$.

A.2. Proof of Theorem 2

By the gradient L_1-Lipschitz continuity, we have that

$$
\begin{aligned}
F_{t,w}(x_{t+1}) &\leq F_{t,w}(x_t) + \left\langle \nabla F_{t,w}(x_t), x_{t+1} - x_t\right\rangle + \frac{L_1\|x_{t+1} - x_t\|^2}{2} \\
&= F_{t,w}(x_t) - \eta\left\langle \nabla F_{t,w}(x_t), G_t(x_t)\right\rangle + \frac{L_1\eta^2\|G_t\|^2}{2} \\
&= F_{t,w}(x_t) - \eta\left\langle \nabla F_{t,w}(x_t), G_t(x_t)\right\rangle + \frac{L_1\eta^2\|\nabla F_{t,w}(x_t) + G_t - \nabla F_{t,w}(x_t)\|^2}{2}
\end{aligned}
$$

$$\leq F_{t,w}(x_t) - \eta \langle \nabla F_{t,w}(x_t), G_t(x_t) \rangle + L_1 \eta^2 \|\nabla F_{t,w}(x_t)\|^2 + L_1 \eta^2 \|G_t - \nabla F_{t,w}(x_t)\|^2. \tag{19}$$

Taking the conditioned expectation (on x_t) on both sides of eq. (19), we obtain that

$$\begin{aligned}
\mathbb{E}\left[F_{t,w}(x_{t+1})|x_t\right] &\leq F_{t,w}(x_t) - \eta\left(1 - L_1\eta\right)\|\nabla F_{t,w}(x_t)\|^2 + L_1\eta^2 \mathbb{E}\left[\|G_t - \nabla F_{t,w}(x_t)\|^2 \big| x_t\right] \\
&\overset{(i)}{\leq} F_{t,w}(x_t) - \eta\left(1 - L_1\eta\right)\|\nabla F_{t,w}(x_t)\|^2 + L_1\eta^2\sigma^2 \\
&\overset{(ii)}{\leq} F_{t,w}(x_t) - \frac{\eta}{2}\|\nabla F_{t,w}(x_t)\|^2 + L_1\eta^2\sigma^2,
\end{aligned} \tag{20}$$

where (i) follows from Assumption 2 and (ii) follows from the choice $\eta \leq \frac{1}{2L_1}$.

Taking expectation on both sides of eq. (20), rearranging the terms, and telescoping from $t = 1$ to T, we obtain that

$$\begin{aligned}
&\mathbb{E}\left[\mathfrak{R}_w(T)\right] \\
&= \sum_{t=1}^{T} \mathbb{E}\left[\|\nabla F_{t,w}(x_t)\|^2\right] \\
&\leq \frac{2}{\eta}\sum_{t=1}^{T}\left(\mathbb{E}\left[F_{t,w}(x_t) - F_{t,w}(x_{t+1})\right] + L_1\eta^2\sigma^2\right) \\
&= \frac{2\left(F_{1,w} - F_{T+1,w}(x_{T+1})\right)}{\eta} + \frac{2}{\eta}\sum_{t=1}^{T}\left(F_{t+1,w}(x_{t+1}) - F_{t,w}(x_{t+1})\right) + 2\eta L_1\sigma^2 T \\
&\overset{(i)}{\leq} \frac{4M}{\eta} + \frac{4MT}{w\eta} + 2\eta L_1\sigma^2 T,
\end{aligned}$$

where (i) follows from the boundedness condition in Assumption 1 and the fact that

$$\begin{aligned}
&F_{t+1,w}(x_{t+1}) - F_{t,w}(x_{t+1}) \\
&= \left|\frac{1}{w}\sum_{i=t-w+2}^{t+1} f_i(x_{t+1}) - \frac{1}{w}\sum_{i=t-w+1}^{t} f_i(x_{t+1})\right| \\
&= \frac{1}{w}|f_{t+1}(x_{t+1}) - f_{t-w+1}(x_{t+1})| \leq \frac{2M}{w}.
\end{aligned}$$

Appendix B. Analysis of OGD with One-point Gradient Estimator in eq. (4)

In this section, we apply the one-point gradient estimator in eq. (4) in a straightforward way to OGD, and the resulting algorithm (see Algorithm 4) can be used for the Window-SVO setting.

Theorem 5 *Suppose Assumption 1 holds. Then, Algorithm 4 satisfies the following regret.*

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \leq \mathcal{O}\left(\frac{T}{\eta w}\right) + \mathcal{O}\left(\frac{d^2\eta T}{\delta^2}\right) + \mathcal{O}(\delta^2 T).$$

Corollary 5 *By taking $\eta = \frac{1}{dw^{2/3}}$ and $\delta = \frac{1}{w^{1/6}}$, Theorem 5 implies that the regret bound for Algorithm 5 satisfies*

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \leq \mathcal{O}\left(\frac{dT}{w^{\frac{1}{3}}}\right).$$

Algorithm 4 OGD with one-point estimator eq. (4)

Input: Initial point $x_1 \in \mathbb{R}^d$, stepsizes η, estimation radius δ
for $t = 1, \ldots, T-1$ **do**
Drawn u_t uniformly from $\mathcal{S}^{d-1}$ independently
Observe $F_{t,w}(x_t + \delta u_t)$
Construct residual feedback estimator as

$$\breve{\nabla} F_{t,w}(x_t) = \frac{d}{\delta} F_{t,w}(x_t + \delta u_t).$$

$x_{t+1} = x_t - \eta \breve{\nabla} F_{t,w}(x_t)$
end for

B.1. Proof of Theorem 5

Let $F_{t,w,\delta}(x) \coloneqq \mathbb{E}\left[F_{t,w}(x + \delta u)\right]$ with u being a random vector drawn uniformly from the unit sphere $\mathcal{S}^{d-1}$. Then $F_{t,w,\delta}(x)$ is also L_1-gradient Lipschitz because

$$\begin{aligned}\|\nabla F_{t,w,\delta}(x) - \nabla F_{t,w,\delta}(y)\| &= \|\nabla \mathbb{E}[F_{t,w}(x + \delta u)] - \nabla \mathbb{E}[F_{t,w}(y + \delta u)]\| \\ &\leq \mathbb{E}\left[\|\nabla F_{t,w}(x + \delta u) - \nabla F_{t,w}(x + \delta u)\|\right] \\ &\overset{(i)}{\leq} L_1 \|x - y\|, \end{aligned} \tag{21}$$

where (i) follows from the L_1-Lipschitz continuity condition of Assumption 1.

By the gradient L_1-Lipschitz condition of $F_{t,w,\delta}(x)$, we obtain that

$$\begin{aligned} F_{t,w,\delta}(x_{t+1}) &\leq F_{t,w,\delta}(x_t) + \langle \nabla F_{t,w,\delta}(x_t), x_{t+1} - x_t \rangle + \frac{L_1}{2}\|x_{t+1} - x_t\|^2 \\ &= F_{t,w,\delta}(x_t) - \eta \left\langle \nabla F_{t,w,\delta}(x_t), \breve{\nabla} F_{t,w}(x_t) \right\rangle + \frac{L_1 \eta^2}{2} \|\breve{\nabla} F_{t,w}(x_t)\|^2. \end{aligned} \tag{22}$$

Taking expectation on both sides of eq. (22) conditioned on x_t, and then taking expectation with respect to x_t, we obtain that

$$\mathbb{E}\left[F_{t,w,\delta}(x_{t+1})\right] \leq \mathbb{E}\left[F_{t,w,\delta}(x_t)\right] - \eta \mathbb{E}\left[\|\nabla F_{t,w,\delta}(x_t)\|^2\right] + \frac{L_1 \eta^2}{2} \mathbb{E}\left[\left\|\breve{\nabla} F_{t,w}(x_t)\right\|^2\right]. \tag{23}$$

Rearranging eq. (23), we obtain that

$$\mathbb{E}\left[\|\nabla F_{t,w,\delta}(x_t)\|^2\right] \overset{(i)}{\leq} \frac{\mathbb{E}[F_{t,w,\delta}(x_t) - F_{t,w,\delta}(x_{t+1})]}{\eta} + \frac{L_1 d^2 M^2 \eta}{\delta^2}, \tag{24}$$

where (i) follows from the fact that

$$\mathbb{E}\left[\left\|\breve{\nabla} F_{t,w}(x_t)\right\|^2\right] = \frac{d^2}{\delta^2} \mathbb{E}\left[\left(F_{t,w}(x_t + \delta u_t)\right)^2\right] \leq \frac{d^2 M^2}{\delta^2}.$$

On the other hand, recall that

$$\|\nabla F_{t,w}(x_t)\|^2 = \|\nabla F_{t,w,\delta}(x_t) - \nabla F_{t,w,\delta}(x_t) + \nabla F_{t,w}(x_t)\|^2$$

$$\leq 2\left\|\nabla F_{t,w,\delta}(x_t)\right\|^2 + 2\left\|\nabla F_{t,w}(x_t) - \nabla F_{t,w,\delta}(x_t)\right\|^2$$
$$\leq 2\left\|\nabla F_{t,w,\delta}(x_t)\right\|^2 + 2L_1^2\delta^2. \tag{25}$$

Combining eqs. (24) and (25), we obtain that

$$\begin{aligned}
\mathbb{E}\left[\mathfrak{R}_w(T)\right] &= \sum_{t=1}^{T} \mathbb{E}\left[\left\|\nabla F_{t,w}(x_t)\right\|^2\right] \\
&\leq \frac{2\sum_{t=1}^{T}\mathbb{E}[F_{t,w,\delta}(x_t) - F_{t,w,\delta}(x_{t+1})]}{\eta} + \frac{2L_1 d^2 M^2 \eta T}{\delta^2} + 2L_1^2\delta^2 T \\
&= \frac{2\mathbb{E}[F_{1,w,\delta}(x_1) - F_{T+1,w,\delta}(x_{T+1})]}{\eta} + \frac{2\sum_{t=1}^{T}\mathbb{E}[F_{t+1,w,\delta}(x_{t+1}) - F_{t,w,\delta}(x_{t+1})]}{\eta} \\
&\quad + \frac{2L_1 d^2 M^2 \eta}{\delta^2} + 2L_1^2\delta^2 T \\
&\overset{(i)}{\leq} \frac{4M}{\eta} + \frac{4MT}{\eta w} + \frac{2L_1 d^2 M^2 \eta}{\delta^2} + 2L_1^2\delta^2 T,
\end{aligned} \tag{26}$$

where (i) follows from the facts that $|F_{t,w,\delta}(x)| \leq M$ for all t, x and

$$|F_{t+1,w,\delta}(x_{t+1}) - F_{t,w,\delta}(x_{t+1})| = \frac{1}{w}|f_{t+1,\delta}(x_{t+1}) - f_{t+1-w,\delta}(x_{t+1})| \leq \frac{2M}{w}.$$

Appendix C. Analysis of OGD with One-point Gradient Estimator OR in eq. (5)

In this section, we apply the one-point gradient estimator OR in eq. (5) in a straightforward way to OGD, and the resulting algorithm (see Algorithm 5) can be used for the Window-SVO setting.

Algorithm 5 OGD with gradient estimator OR in eq. (5)

Input: Initial point $x_1 \in \mathbb{R}^d$, stepsizes η, estimation radius δ
for $t = 1, \ldots, T-1$ **do**
  Drawn u_t uniformly from $\mathcal{S}^{d-1}$ independently
  Observe $F_{t,w}(x_t + \delta u_t)$
  Construct residual feedback estimator as

$$\bar{\nabla} F_{t,w}(x_t) = \frac{d}{\delta}\left(F_{t,w}(x_t + \delta u_t) - F_{t-1,w}(x_{t-1} + \delta u_{t-1})\right) u_t,$$

  where we take $F_{t,w}(x_t + \delta u_t) = 0$ for $t \leq 0$.
  $x_{t+1} = x_t - \eta \bar{\nabla} F_{t,w}(x_t)$
end for

We first show that the gradient estimator OR satisfies the following iterative upper bound as given in eq. (6).

Lemma 3 *Suppose Assumption 1 holds, we have*

$$\mathbb{E}\left[\left\|\bar{\nabla} F_{t,w}(x_t)\right\|^2\right] \leq \frac{3d^2\eta^2 L_0^2}{\delta^2}\left\|\bar{\nabla} F_{t-1,w}(x_{t-1})\right\|^2 + \frac{12M^2 d^2}{w^2\delta^2} + 12d^2 L_0^2. \tag{27}$$

Proof By the definition of $\bar{\nabla} F_{t,w}(x_t)$, we have

$$\begin{aligned}
&\left\|\bar{\nabla} F_{t,w}(x_t)\right\|^2 \\
&= \frac{d^2}{\delta^2}\left(F_{t,w}(x_t+\delta u_t) - F_{t-1,w}(x_{t-1}+\delta u_{t-1})\right)^2 \\
&= \frac{d^2}{\delta^2}\left(F_{t,w}(x_t+\delta u_t) - F_{t-1,w}(x_t+\delta u_t) + F_{t-1,w}(x_t+\delta u_t) - F_{t-1,w}(x_{t-1}+\delta u_t)\right. \\
&\qquad\qquad \left. + F_{t-1,w}(x_{t-1}+\delta u_t) - F_{t-1,w}(x_{t-1}+\delta u_{t-1})\right)^2 \\
&\le \frac{3d^2}{\delta^2}\left(F_{t,w}(x_t+\delta u_t) - F_{t-1,w}(x_t+\delta u_t)\right)^2 + \frac{3d^2}{\delta^2}\left(F_{t-1,w}(x_t+\delta u_t) - F_{t-1,w}(x_{t-1}+\delta u_t)\right)^2 \\
&\qquad + \frac{3d^2}{\delta^2}\left(F_{t-1,w}(x_{t-1}+\delta u_t) - F_{t-1,w}(x_{t-1}+\delta u_{t-1})\right)^2 \\
&\overset{(i)}{\le} \frac{3d^2\eta^2 L_0^2}{\delta^2}\left\|\bar{\nabla} F_{t-1,w}(x_{t-1})\right\|^2 + \frac{12M^2d^2}{w^2\delta^2} + 12d^2L_0^2, \qquad (28)
\end{aligned}$$

where (i) follows from the update rule of x_t, Assumption 1 and the fact that

$$|F_{t,w}(z) - F_{t-1,w}(z)| = \frac{1}{w}|f_t(z) - f_{t-w}(z)| \le \frac{2M}{w}.$$

■

Based on Lemma 3 and the gradient Lipschitz continuity, we prove the following regret bound of Algorithm 5.

Theorem 6 *Suppose Assumption 1 holds and $\eta \le \frac{\delta}{3dL_0}$. Then Algorithm 5 satisfies the following regret bound.*

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \le \mathcal{O}\left(\frac{T}{\eta w}\right) + \mathcal{O}\left(\delta^2 T\right) + \mathcal{O}\left(\frac{d^2\eta T}{w^2\delta^2}\right) + \mathcal{O}\left(d^2\eta T\right) + \mathcal{O}\left(\frac{d^2\eta}{\delta^2}\right).$$

By properly choosing η and δ, Theorem 6 implies the regret bound of $\mathcal{O}\left(\frac{dT}{\sqrt{w}}\right)$ as given by the following corollary.

Corollary 6 *Taking $\eta = \frac{1}{3L_0 d\sqrt{w}}$, and $\delta = \frac{1}{\sqrt{w}}$, Theorem 6 implies that the regret bound for Algorithm 5 satisfies*

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \le \mathcal{O}\left(\frac{dT}{\sqrt{w}}\right).$$

C.1. Proof of Theorem 6

Let $F_{t,w,\delta}(x) := \mathbb{E}\left[F_{t,w}(x+\delta u)\right]$ with u being a random vector drawn uniformly from the unit sphere $\mathcal{S}^{d-1}$. By the gradient L_1-Lipschitz continuity of $F_{t,w,\delta}(x)$ (see eq. (21)), we obtain that

$$\begin{aligned}
F_{t,w,\delta}(x_{t+1}) &\le F_{t,w,\delta}(x_t) + \left\langle \nabla F_{t,w,\delta}(x_t), x_{t+1} - x_t\right\rangle + \frac{L_1}{2}\|x_{t+1} - x_t\|^2 \\
&= F_{t,w,\delta}(x_t) - \eta\left\langle \nabla F_{t,w,\delta}(x_t), \bar{\nabla} F_{t,w}(x_t)\right\rangle + \frac{L_1\eta^2}{2}\|\bar{\nabla} F_{t,w}(x_t)\|^2. \qquad (29)
\end{aligned}$$

Taking expectation on both sides of Equation (29) conditioned on x_t, and then taking expectation with respect to x_t, we have

$$\mathbb{E}\left[F_{t,w,\delta}(x_{t+1})\right] \leq \mathbb{E}\left[F_{t,w,\delta}(x_t)\right] - \eta\mathbb{E}\left[\left\|\nabla F_{t,w,\delta}(x_t)\right\|^2\right] + \frac{L_1\eta^2}{2}\mathbb{E}\left[\left\|\bar{\nabla} F_{t,w}(x_t)\right\|^2\right]. \tag{30}$$

Rearranging eq. (30), we obtain that

$$\mathbb{E}\left[\left\|\nabla F_{t,w,\delta}(x_t)\right\|^2\right] \leq \frac{\mathbb{E}[F_{t,w,\delta}(x_t) - F_{t,w,\delta}(x_{t+1})]}{\eta} + \frac{L_1\eta}{2}\mathbb{E}\left[\left\|\bar{\nabla} F_{t,w}(x_t)\right\|^2\right]. \tag{31}$$

On the other hand, recall that

$$\begin{aligned}\left\|\nabla F_{t,w}(x_t)\right\|^2 &= \left\|\nabla F_{t,w,\delta}(x_t) - \nabla F_{t,w,\delta}(x_t) + \nabla F_{t,w}(x_t)\right\|^2 \\ &\leq 2\left\|\nabla F_{t,w,\delta}(x_t)\right\|^2 + 2\left\|\nabla F_{t,w}(x_t) - \nabla F_{t,w,\delta}(x_t)\right\|^2 \\ &\leq 2\left\|\nabla F_{t,w,\delta}(x_t)\right\|^2 + 2L_1\delta^2.\end{aligned} \tag{32}$$

Combining eqs. (31) and (32), we obtain that

$$\begin{aligned}\mathbb{E}\left[\mathfrak{R}_w(T)\right] &= \sum_{t=1}^{T}\mathbb{E}\left[\left\|\nabla F_{t,w}(x_t)\right\|^2\right] \\ &\leq \frac{2\sum_{t=1}^{T}\mathbb{E}[F_{t,w,\delta}(x_t) - F_{t,w,\delta}(x_{t+1})]}{\eta} + L_1\eta\sum_{t=1}^{T}\mathbb{E}\left[\left\|\bar{\nabla} F_{t,w}(x_t)\right\|^2\right] + 2L_1\delta^2 T \\ &= \frac{2\mathbb{E}[F_{1,w,\delta}(x_1) - F_{T+1,w,\delta}(x_{T+1})]}{\eta} + \frac{2\sum_{t=1}^{T}\mathbb{E}[F_{t+1,w,\delta}(x_{t+1}) - F_{t,w,\delta}(x_{t+1})]}{\eta} \\ &\quad + L_1\eta\sum_{t=1}^{T}\mathbb{E}\left[\left\|\bar{\nabla} F_{t,w}(x_t)\right\|^2\right] + 2L_1\delta^2 T \\ &\overset{(i)}{\leq} \frac{2M}{\eta} + \frac{2MT}{\eta w} + L_1\eta\sum_{t=1}^{T}\mathbb{E}\left[\left\|\bar{\nabla} F_{t,w}(x_t)\right\|^2\right] + 2L_1\delta^2 T,\end{aligned} \tag{33}$$

where (i) follows from the facts that $|F_{t,w,\delta}(x)| \leq M$ for all t, x and

$$|F_{t+1,w,\delta}(x_{t+1}) - F_{t,w,\delta}(x_{t+1})| = \frac{1}{w}|f_{t+1,\delta}(x_{t+1}) - f_{t+1-w,\delta}(x_{t+1})| \leq \frac{2M}{w}.$$

Based on eq. (33), we conclude that the upper bound of $\mathfrak{R}_w(T)$ can be obtained through the upper bounds of $\sum_{t=1}^{T}\left\|\bar{\nabla} F_{t,w}(x_t)\right\|^2$. Thus, we proceed with bounding this term. Iteratively applying Lemma 3, we have that

$$\begin{aligned}\mathbb{E}\left[\left\|\bar{\nabla} F_{t,w}(x_t)\right\|^2\right] &\leq \left(\frac{12M^2d^2}{w^2\delta^2} + 12d^2L_0^2\right)\sum_{i=0}^{t}\left(\frac{3d^2\eta^2L_0^2}{\delta^2}\right)^i + \left(\frac{3d^2\eta^2L_0^2}{\delta^2}\right)^{t-1}\left\|\bar{\nabla} F_{0,w}(x_0)\right\|^2 \\ &\overset{(i)}{\leq} \frac{24M^2d^2}{w^2\delta^2} + 24d^2L_0^2 + \frac{d^2M^2}{\delta^2}\left(\frac{1}{2}\right)^{t-1},\end{aligned} \tag{34}$$

where (i) follows from the fact that $\frac{3d^2\eta^2L_0^2}{\delta^2} \leq \frac{1}{2}$ when $\eta \leq \frac{\delta}{3dL_0}$.

Substituting eq. (34) into eq. (33), we obtain that

$$\mathbb{E}\left[\mathfrak{R}_w(T)\right] \leq \frac{2M}{\eta} + \frac{2MT}{\eta w} + 2L_1\delta^2 T + \frac{24M^2d^2L_1\eta T}{w^2\delta^2} + 24d^2L_0^2L_1\eta T + \frac{2d^2M^2L_1\eta}{\delta^2}.$$

Appendix D. Proof in Section 5

D.1. Proof of Lemma 1

By the definition of $\widehat{\nabla} F_{t,w}(x_t)$, we have

$$\begin{aligned}
\widehat{\nabla} F_{t,w}(x_t) &= \frac{d}{\delta}(F_{t,w}(x_t+\delta u_t) - F_{t-1,w}(x_{t-1}))u_t \\
&= \frac{d}{\delta}(F_{t,w}(x_t+\delta u_t) - F_{t,w}(x_t) + F_{t,w}(x_t) - F_{t-1,w}(x_{t-1}))u_t \\
&\overset{(i)}{=} \frac{d}{\delta}\left(\int_0^\delta \langle \nabla F_{t,w}(x_t+vu_t), u_t\rangle dv\right) u_t + \frac{d}{\delta}\left(F_{t,w}(x_t) - F_{t-1,w}(x_{t-1})\right) \\
&= \frac{d}{\delta}\left(\int_0^\delta \langle \nabla F_{t,w}(x_t) + \nabla F_{t,w}(x_t+vu_t) - \nabla F_{t,w}(x_t), u_t\rangle dv\right) u_t \\
&\qquad + \frac{d}{\delta}\left(F_{t,w}(x_t) - F_{t-1,w}(x_{t-1})\right) u_t \\
&= \frac{d}{\delta}\left(\int_0^\delta \langle \nabla F_{t,w}(x_t) - \nabla F_{t,w}(x_t+vu_t), u_t\rangle dv\right) u_t \\
&\qquad + \frac{d}{\delta}\left(\int_0^\delta \langle \nabla F_{t,w}(x_t), u_t\rangle dv\right) u_t + \frac{d}{\delta}\left(F_{t,w}(x_t) - F_{t-1,w}(x_{t-1})\right) u_t \\
&\overset{(ii)}{\leq} \frac{du_t}{\delta}\int_0^\delta L_1 v dv + d\langle \nabla F_{t,w}(x_t), u_t\rangle u_t + \frac{2Mdu_t}{w\delta} \\
&= \frac{L_1 d\delta u_t}{2} + d\langle \nabla F_{t,w}(x_t), u_t\rangle u_t + \frac{2Mdu_t}{w\delta},
\end{aligned} \tag{35}$$

where (i) follows from the law of Calculus, and (ii) follows from the L_1-gradient Lipschitz continuity of $F_{t,w}(x_t)$, and because $x_t = x_{t-1}$ when t is even, and

$$|F_{t,w}(x) - F_{t-1,w}(x)| = \frac{1}{w}|f_t(x) - f_{t-w}(x)| \leq \frac{2M}{w}.$$

Using the fact that $\|a+b+c\|^2 \leq 3\|a\|^2 + 3\|b\|^2 + 3\|c\|^2$ and applying eq. (18), we obtain that

$$\left\|\widehat{\nabla} F_{t,w}(x_t)\right\|^2 \leq 3d^2 \left|\langle \nabla F_{t,w}(x_t), u_t\rangle\right|^2 + \frac{12M^2d^2}{w^2\delta^2} + L_1^2 d^2 \delta^2. \tag{36}$$

In order to bound $\left\|\widehat{\nabla} F_{t,w}(x_t)\right\|^2$, we need to bound the term $|\langle \nabla F_{t,w}(x_t), u_t\rangle|^2$. Let $l = \frac{\nabla F_{t,w}(x_t)}{\|\nabla F_{t,w}(x_t)\|}$, we have that $|\langle \nabla F_{t,w}(x_t), u_t\rangle|^2 = \|\nabla F_{t,w}(x_t)\|^2 |\langle l, u_t\rangle|^2$. Moreover, let $z = |\langle l, u_t\rangle|$. Equation (36) can be written as

$$\left\|\widehat{\nabla} F_{t,w}(x_t)\right\|^2 \leq 3d^2 \|\nabla F_{t,w}(x_t)\|^2 z^2 + \frac{12M^2d^2}{w^2\delta^2} + L_1^2 d^2 \delta^2. \tag{37}$$

Since u_t is uniformly sampled from the unit sphere, the probability density function of z follows

$$f(z) = \begin{cases} \frac{2\Gamma(d/2)}{\sqrt{\pi}\Gamma((d-1)/2)}(1-z^2)^{(d-3)/2}, & 0 \leq z \leq 1 \\ 0, & o.w. \end{cases},$$

where $\Gamma(\cdot)$ is the gamma-function.

We then have

$$\mathbb{E}[z^2] = \int_{z=0}^{1} z^2 f(z) dz = \frac{1}{d}. \tag{38}$$

Taking expectation on both sides of eq. (37) conditioned on $\mathcal{F}_t$ and applying eq. (38), we obtain that

$$\mathbb{E}\left[\left\|\widehat{\nabla} F_{t,w}(x_t)\right\|^2 \middle| \mathcal{F}_t\right] \leq 3d \left\|\nabla F_{t,w}(x_t)\right\|^2 + \frac{12M^2d^2}{w^2\delta^2} + L_1^2 d^2 \delta^2.$$

D.2. Proof of Theorem 3

Let $F_{t,w,\delta}(x) \coloneqq \mathbb{E}\left[F_{t,w}(x + \delta u)\right]$ with u being a random vector drawn uniformly from the unit sphere $\mathcal{S}^{d-1}$. For any $t = 2j$ with some $j \in \mathbb{N}_+$, by the gradient Lipschitz continuity of $F_{t,w,\delta}(x)$ (see eq. (21)), we have that

$$\begin{aligned} F_{t,w,\delta}(x_{t+1}) &\leq F_{t,w,\delta}(x_t) + \langle \nabla F_{t,w,\delta}(x_t), x_{t+1} - x_t \rangle + \frac{L_1}{2}\|x_{t+1} - x_t\|^2 \\ &= F_{t,w,\delta}(x_t) - \eta \left\langle \nabla F_{t,w,\delta}(x_t), \widehat{\nabla} F_{t,w}(x_t) \right\rangle + \frac{L_1\eta^2}{2}\|\widehat{\nabla} F_{t,w}(x_t)\|^2. \end{aligned} \tag{39}$$

Defining the filtration $\mathcal{F}_t \coloneqq \sigma\left(u_1, u_2, \ldots, u_{2\left\lfloor \frac{t}{2} \right\rfloor - 2}\right)$, and taking expectation on both sides of eq. (39), we obtain that

$$\begin{aligned} &\mathbb{E}\left[F_{t,w,\delta}(x_{t+1}) | \mathcal{F}_t\right] \\ &\quad \leq F_{t,w,\delta}(x_t) - \eta \left\langle \nabla F_{t,w,\delta}(x_t), \mathbb{E}\left[\widehat{\nabla} F_{t,w}(x_t) \middle| \mathcal{F}_t\right] \right\rangle + \frac{L_1\eta^2}{2} \mathbb{E}\left[\|\widehat{\nabla} F_{t,w}(x_t)\|^2 \middle| \mathcal{F}_t\right] \\ &\quad \overset{(i)}{=} F_{t,w,\delta}(x_t) - \eta \|\nabla F_{t,w,\delta}(x_t)\|^2 + \frac{L_1\eta^2}{2} \mathbb{E}\left[\|\widehat{\nabla} F_{t,w}(x_t)\|^2 \middle| \mathcal{F}_t\right] \\ &\quad \overset{(ii)}{\leq} F_{t,w,\delta}(x_t) - \eta\left(1 - \frac{3L_1 d\eta}{2}\right)\|\nabla F_{t,w,\delta}(x_t)\|^2 + \frac{6L_1\eta^2M^2d^2}{w^2\delta^2} + \frac{L_1^3\eta^2d^2\delta^2}{2}, \end{aligned} \tag{40}$$

where (i) follows from the fact that $\mathbb{E}\left[\widehat{\nabla} F_{t,w}(x_t) \middle| \mathcal{F}_t\right] = \nabla F_{t,w,\delta}(x_t)$, and (ii) follows from Lemma 1.

Using the fact that $\eta = \frac{1}{3L_1 d}$ and rearranging eq. (40), we have

$$\|\nabla F_{t,w,\delta}(x_t)\|^2 \leq \frac{2(F_{t,w,\delta}(x_t) - \mathbb{E}\left[F_{t,w,\delta}(x_{t+1}) | \mathcal{F}_t\right])}{\eta} + \frac{12L_1\eta M^2 d^2}{w^2\delta^2} + L_1^3\eta d^2\delta^2. \tag{41}$$

Recall the definition of regret

$$\begin{aligned} \mathfrak{R}_w(T) &= \sum_{t=1}^{T} \|\nabla F_{t,w}(x_t)\|^2 \\ &= \sum_{t=1}^{T} \|\nabla F_{t,w}(x_t) - \nabla F_{t,w,\delta}(x_t) + \nabla F_{t,w,\delta}(x_t)\|^2 \end{aligned}$$

$$\overset{(i)}{\leq} 2\sum_{t=1}^{T}\|\nabla F_{t,w}(x_t)-\nabla F_{t,w,\delta}(x_t)\|^2+2\sum_{t=1}^{T}\|\nabla F_{t,w,\delta}(x_t)\|^2$$
$$\overset{(ii)}{\leq} 2\sum_{t=1}^{T}\|\nabla F_{t,w,\delta}(x_t)\|^2+2L_1^2\delta^2T, \tag{42}$$

where (ii) follows from the fact that $\|\nabla F_{t,w}(x_t)-\nabla F_{t,w,\delta}(x_t)\|\leq L_1\delta$ for all $t=1,\ldots,T$, and (i) follows from the fact that

$$\begin{aligned}\|\nabla F_{t,w}(x_t)-\nabla F_{t,w,\delta}(x_t)\| &= \|\nabla F_{t,w}(x_t)-\mathbb{E}\left[\nabla F_{t,w}(x_t+\delta u)\right]\| \\ &\leq \mathbb{E}[\|\nabla F_{t,w}(x_t)-\nabla F_{t,w}(x_t+\delta u)\|] \\ &\leq L_1\delta.\end{aligned}$$

Let $\bar{T}\coloneqq 2\left\lfloor\frac{T}{2}\right\rfloor+1$. Taking expectation on both sides of eq. (42), we obtain that

$$\begin{aligned}&\mathbb{E}[\mathfrak{R}_w(T)] \\ &\leq 2\mathbb{E}\left[\sum_{t=1}^{T}\|\nabla F_{t,w,\delta}(x_t)\|^2\right]+2L_1^2\delta^2T \\ &\leq 2\mathbb{E}\left[\sum_{j=1}^{\lfloor\frac{T}{2}\rfloor}\|\nabla F_{2j,w,\delta}(x_{2j})\|^2\right]+2\mathbb{E}\left[\sum_{j=1}^{\lfloor\frac{T}{2}\rfloor+1}\|\nabla F_{2j-1,w,\delta}(x_{2j-1})\|^2\right]+2L_1^2\delta^2T \\ &\overset{(i)}{\leq} 4\mathbb{E}\left[\sum_{j=1}^{\lfloor\frac{T}{2}\rfloor}\|\nabla F_{2j,w,\delta}(x_{2j})\|^2\right]+\frac{4L_0^2T}{w^2}+2L_1^2\delta^2T \\ &\overset{(ii)}{\leq} \sum_{j=1}^{\lfloor\frac{T}{2}\rfloor}\frac{8\mathbb{E}\left[F_{2j,w,\delta}(x_{2j})-F_{2j,w,\delta}(x_{2j+1})\right])}{\eta}+\frac{24L_1\eta M^2Td^2}{w^2\delta^2}+2L_1^3\eta d^2\delta^2T+\frac{4L_0^2T}{w^2}+2L_1^2\delta^2T \\ &= \frac{4\mathbb{E}\left[F_{1,w,\delta}(x_1)-F_{\bar{T},w,\delta}\left(x_{\bar{T}}\right)\right]}{\eta}+\frac{4\mathbb{E}\left[\sum_{j=1}^{\lfloor\frac{T}{2}\rfloor}F_{2j+2,w,\delta}(x_{2j+2})-F_{2j,w,\delta}(x_{2j+1})\right]}{\eta} \\ &\qquad+\frac{24L_1\eta M^2Td^2}{w^2\delta^2}+2L_1^3\eta d^2\delta^2T+\frac{4L_0^2T}{w^2}+2L_1^2\delta^2T \\ &\overset{(iii)}{\leq} \frac{8M}{\eta}+\frac{8MT}{w\eta}+\frac{24L_1\eta M^2d^2T}{w^2\delta^2}+2L_1^3\eta d^2\delta^2T+\frac{4L_0^2T}{w^2}+2L_1^2\delta^2T \\ &= 24L_1dM+\frac{24L_1dMT}{w}+\frac{8M^2dT}{w^2\delta^2}+\frac{2L_1^2d\delta^2T}{3}+\frac{L_0^2T}{w^2}+2L_1^2\delta^2T,\end{aligned} \tag{43}$$

where (i) follows from $x_t=x_{t-1}$ when t is odd and $\|\nabla F_{t,w}(x)-\nabla F_{t-1,w}(x)\|\leq\frac{2L_0}{w}$, (ii) follows by telescoping eq. (41), and (iii) follows from Assumption 1 and the facts that $x_{2j+1}=x_{2j+2}$ and

$$\begin{aligned}&|F_{2j+2,w}(x_{2j+2})-F_{2j+1,w}(x_{2j+1})| \\ &\qquad=\frac{1}{w}|f_{2j+2}(x_{2j+2})+f_{2j+2}(x_{2j+2})-f_{2j+2-2}(x_{2j+2})-f_{2j+1-w}(x_{2j+2})| \\ &\qquad\leq\frac{4M}{w}.\end{aligned}$$

Appendix E. Proofs of Section 6

E.1. Proof of Lemma 2

Following the steps similar to those deriving eq. (35), we obtain that

$$\widetilde{\nabla} F_{t,w}(y_t^k) \le d\langle \nabla F_{t,w}(y_t^k), u_t^k\rangle u_t^k + \frac{L_1 d\delta u_t^k}{2}. \tag{44}$$

Using the fact that $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$ and eq. (44), we obtain that

$$\left\|\widetilde{\nabla} F_{t,w}(y_t^k)\right\|^2 \le 2d^2 \left|\langle \nabla F_{t,w}(y_t^k), u_t^k\rangle\right|^2 + \frac{L_1^2 d^2\delta^2}{2}.$$

Following the proof similar to that of Lemma 1 in Appendix D.1, we obtain that

$$\mathbb{E}\left[\left\|\widetilde{\nabla} F_{t,w}(y_t^k)\right\|^2 \middle| \mathcal{F}_t^k\right] \le 2d\left\|\nabla F_{t,w}(y_t^k)\right\|^2 + \frac{L_1^2 d^2\delta^2}{2}.$$

E.2. Proof of Theorem 4

Let $F_{t,w,\delta}(x) := \mathbb{E}\left[F_{t,w}(x+\delta u)\right]$ with u being a random vector drawn uniformly from the unit sphere $\mathcal{S}^{d-1}$. By the gradient Lipschitz condition of $F_{t,w,\delta}(x)$ (see eq. (21)), we have

$$\begin{aligned} F_{t,w,\delta}(y_t^{k+1}) &\le F_{t,w,\delta}(y_t^k) + \left\langle \nabla F_{t,w,\delta}(y_t^k), y_t^{k+1} - y_t^k\right\rangle + \frac{L_1}{2}\left\|y_t^{k+1} - y_t^k\right\|^2 \\ &= F_{t,w,\delta}(y_t^k) - \eta\left\langle \nabla F_{t,w,\delta}(y_t^k), \widetilde{\nabla} F_{t,w}(y_t^k)\right\rangle + \frac{L_1\eta^2}{2}\left\|\widetilde{\nabla} F_{t,w}(y_t^k)\right\|^2. \end{aligned} \tag{45}$$

Let $\mathcal{F}_{t,k} = \sigma_f(u_1^1, \ldots, u_1^K, \ldots, u_t^1, \ldots, u_t^{k-1})$. Taking expectation on both sides conditioned on $\mathcal{F}_{t,k}$, we have

$$\begin{aligned} &\mathbb{E}\left[\|\nabla F_{t,w,\delta}(y_t^{k+1})\|^2 \middle| \mathcal{F}_{t,k}\right] \\ &\le F_{t,w,\delta}(y_t^k) - \eta\left\langle \nabla F_{t,w,\delta}(y_t^k), \mathbb{E}\left[\widetilde{\nabla} F_{t,w}(y_t^k)\middle|\mathcal{F}_{t,k}\right]\right\rangle + \frac{L_1\eta^2}{2}\mathbb{E}\left[\left\|\widetilde{\nabla} F_{t,w}(y_t^k)\right\|^2\middle|\mathcal{F}_{t,k}\right] \\ &\overset{(i)}{=} F_{t,w,\delta}(y_t^k) - \eta\left\|\nabla F_{t,w,\delta}(y_t^k)\right\|^2 + \frac{L_1\eta^2}{2}\mathbb{E}\left[\left\|\widetilde{\nabla} F_{t,w}(y_t^k)\right\|^2\middle|\mathcal{F}_{t,k}\right] \\ &\overset{(ii)}{\le} F_{t,w,\delta}(y_t^k) - \eta(1 - L_1\eta d)\left\|\nabla F_{t,w,\delta}(y_t^k)\right\|^2 + \frac{L_1^3\eta^2 d^2\delta^2}{4}, \end{aligned} \tag{46}$$

where (i) follows from $\mathbb{E}\left[F_{t,w}(y_t^k)\middle|\mathcal{F}_{t,k}\right] = F_{t,w,\delta}(y_t^k)$ and (ii) follows from Lemma 2.

Using the fact that $\eta = \frac{1}{2L_1 d}$ and rearranging eq. (46), we obtain

$$\left\|\nabla F_{t,w,\delta}(y_t^k)\right\|^2 \le \frac{2\left(F_{t,w,\delta}(y_t^k) - \mathbb{E}\left[F_{t,w,\delta}(y_t^{k+1})\middle|\mathcal{F}_{t,k}\right]\right)}{\eta} + \frac{L_1^3\eta d^2\delta^2}{2}. \tag{47}$$

Taking expectation on both sides of eq. (47), we have

$$\mathbb{E}\left[\left\|\nabla F_{t,w,\delta}(y_t^k)\right\|^2\right] \le \frac{2\mathbb{E}\left[F_{t,w,\delta}(y_t^k) - F_{t,w,\delta}(y_t^{k+1})\right]}{\eta} + \frac{L_1^3\eta d^2\delta^2}{2}. \tag{48}$$

Recall the definition of the generation of I_t and that $x_{t+1} = y_t^{I_t}$. Then, we have

$$\begin{aligned}
&\mathbb{E}\left[\|\nabla F_{t,w}(x_{t+1})\|^2\right] \\
&= \frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\left[\|\nabla F_{t,w}(y_t^k)\|^2\right] \\
&\leq \frac{2}{K}\sum_{k=0}^{K-1}\mathbb{E}\left[\|\nabla F_{t,w,\delta}(y_t^k)\|^2\right] + \frac{2}{K}\sum_{k=0}^{K}\mathbb{E}\left[\|\nabla F_{t,w,\delta}(y_t^k) - \nabla F_{t,w}(y_t^k)\|^2\right] \\
&\overset{(i)}{\leq} \frac{2\mathbb{E}\left[F_{t,w,\delta}(y_t^0) - F_{t,w,\delta}(y_t^K)\right]}{\eta K} + L_1^3\eta d^2\delta^2 + 2L_1^2\delta^2, \qquad (49)
\end{aligned}$$

where (i) follows from telescoping eq. (48) and because $\|\nabla F_{t,w,\delta}(z) - \nabla F_{t,w}(z)\| \leq L_1^2\delta^2$ holds for arbitrary z.

Using eq. (49) and the fact

$$\|\nabla F_{t,w}(x)\|^2 \leq 2\|\nabla F_{t+1,w}(x)\|^2 + \frac{8L_0^2}{w^2},$$

we obtain

$$\begin{aligned}
&\mathbb{E}\left[\|\nabla F_{t+1,w}(x_{t+1})\|^2\right] \\
&\qquad \leq \frac{4\mathbb{E}\left[F_{t,w,\delta}(y_t^0) - F_{t,w,\delta}(y_t^K)\right]}{\eta K} + \frac{8L_0^2}{w^2} + 2L_1^3\eta d^2\delta^2 + 4L_1^2\delta^2. \qquad (50)
\end{aligned}$$

We then have

$$\begin{aligned}
\mathbb{E}\left[\mathfrak{R}_w(T)\right] &= \sum_{t=1}^{T}\mathbb{E}\left[\|\nabla F_{t,w}(x_t)\|^2\right] \\
&\overset{(i)}{\leq} \frac{4\sum_{t=1}^{T}\mathbb{E}\left[F_{t,w,\delta}(y_t^0) - F_{t,w,\delta}(y_t^K)\right]}{\eta K} + \frac{8L_0^2T}{w^2} + 2L_1^3\eta d^2\delta^2T + 4L_1^2\delta^2T \\
&\leq \frac{4(F_{1,w,\delta}(y_1^0) - F_{T+1,w,\delta}(y_{T+1}^K)) + 4\sum_{t=1}^{T}\mathbb{E}\left[F_{t+1,w,\delta}(y_{t+1}^0) - F_{t,w,\delta}(y_t^K)\right]}{\eta K} \\
&\qquad + \frac{8L_0^2T}{w^2} + 2L_1^3\eta d^2\delta^2T + 4L_1^2\delta^2T \\
&\overset{(ii)}{\leq} \frac{8M}{\eta K} + \frac{8MT}{\eta wK} + \frac{8L_0^2T}{w^2} + 2L_1^3\eta d^2\delta^2T + 4L_1^2\delta^2T \\
&= 16ML_1d + \frac{16L_1dMT}{wK} + \frac{8L_0^2T}{w^2} + L_1^2d\delta^2T + 4L_1^2\delta^2T, \qquad (51)
\end{aligned}$$

where (i) follows from telescoping eq. (50) and (ii) follows from the fact that $y_t^K = y_{t+1}^0$ and, for all x,

$$\|F_{t,w,\delta}(x) - F_{t+1,w,\delta}(x)\| \leq \frac{2M}{w}.$$