Tracking Slowly Moving Clairvoyant: Optimal Dynamic Regret of Online Learning with True and Noisy Gradient

Tianbao Yang TIANBAO-YANG@UIOWA.EDU
Department of Computer Science, The University of Iowa, Iowa City, IA 52242, USA

Lijun Zhang ZLJZJU@GMAIL.COM
National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

Rong Jin JINRONG.JR@ALIBABA-INC.COM
Institute of Data Science and Technologies at Alibaba Group, Seattle, USA

Jinfeng Yi JINFENGY@US.IBM.COM
IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA

Abstract

This work focuses on dynamic regret of online convex optimization that compares the performance of online learning to a clairvoyant who knows the sequence of loss functions in advance and hence selects the minimizer of the loss function at each step. By assuming that the clairvoyant moves slowly (i.e., the minimizers change slowly), we present several improved variation-based upper bounds of the dynamic regret under the true and noisy gradient feedback, which are *optimal* in light of the presented lower bounds. The key to our analysis is to explore a regularity metric that measures the temporal changes in the clairvoyant's minimizers, to which we refer as *path variation*. Firstly, we present a general lower bound in terms of the path variation, and then show that under full information or gradient feedback we are able to achieve an optimal dynamic regret. Secondly, we present a lower bound with noisy gradient feedback and then show that we can achieve optimal dynamic regrets under a stochastic gradient feedback and two-point bandit feedback. Moreover, for a sequence of smooth loss functions that admit a small variation in the gradients, our dynamic regret under the two-point bandit feedback matches what is achieved with full information.

Proceedings of the 33^{rd} International Conference on Machine Learning, New York, NY, USA, 2016. JMLR: W&CP volume 48.

1. Introduction

Online convex optimization (OCO) can be deemed as a repeated game between an online player and an adversary, in which an online player iteratively chooses a decision and then her decisions incur (possibly different) losses by the loss functions chosen by the adversary. These loss functions are unknown to the decision maker ahead of time, and can be adversarial or even depend on the action taken by the decision maker. To formulate the problem mathematically, let $\Omega \subseteq \mathbb{R}^d$ denote a convex decision set (i.e., the feasible set of the decision vector), $\mathbf{w}_t \in \Omega$ denote the decision vector and $f_t(\cdot) : \mathbb{R}^d \to \mathbb{R}$ denote the loss function at the t-th step, respectively. The goal of the online learner is to minimize her cumulative loss $\sum_{t=1}^T f_t(\mathbf{w}_t)$. The traditional performance metric - the regret of the decision maker, is defined as the difference between the total cost she has incurred and that of the best fixed decision in hindsight, i.e.,

$$\sum_{t=1}^T f_t(\mathbf{w}_t) - \min_{\mathbf{w}\in\Omega} \sum_{t=1}^T f_t(\mathbf{w}). \tag{1}$$

Recently, there emerges a surge of interest (Besbes et al., 2013; Hall & Willett, 2013; Jadbabaie et al., 2015) in the dynamic regret that compares the performance of online learning to a sequence of optimal solutions. If we denote by $\mathbf{w}_t^* \in \Omega$ an optimal solution of $f_t(\mathbf{w})$, the dynamic regret is defined as

$$\sum_{t=1}^T (f_t(\mathbf{w}_t) - f_t(\mathbf{w}_t^*)) = \sum_{t=1}^T (f_t(\mathbf{w}_t) - \min_{\mathbf{w}\in\Omega} f_t(\mathbf{w})) \tag{2}$$

i.e., the performance of the online learner is compared to a clairvoyant who knows the sequence of loss functions in

advance, and hence selects the minimizer $\mathbf{w}_t^*$ at each step. Compared to the traditional regret in (1) (termed as static regret), the dynamic regret is more aggressive since the performance of the clairvoyant in the dynamic regret model is always better than that in the static regret model, i.e., $\sum_{t=1}^T f_t(\mathbf{w}_t^*) \leq \min_{\mathbf{w}\in\Omega} \sum_{t=1}^T f_t(\mathbf{w})$. It was pointed out that algorithms that achieve performance close to the best fixed decision may perform poorly in terms of dynamic regret (Besbes et al., 2013).

Unfortunately, it is impossible to achieve a sublinear dynamic regret for *any* sequences of loss functions (c.f. Proposition 1). In order to achieve a sublinear dynamic regret, one has to impose some regularity constraints on the sequence of loss functions. In this work, we leverage a notion of variation that measures how fast the clairvoyant moves, i.e., how fast the minimizers of the sequence of loss functions change, to which we refer as **path variation** in order to differentiate with other variation definitions. Formally the path variation is defined as

$$V_T^p \triangleq \max_{\{\mathbf{w}_t^*\in\Omega_t^*\}_{t=1}^T} \sum_{t=1}^{T-1} \|\mathbf{w}_t^* - \mathbf{w}_{t+1}^*\|_2 \tag{3}$$

where Ω_t^* denotes the set of all minimizers of $f_t(\mathbf{w})$ to account for the potential non-uniqueness. We aim to develop optimal dynamic regrets when the clairvoyant moves slowly given (noisy) gradient feedback (including bandit feedback) for non-strongly convex loss functions. The main results are summarized in Table 1 and the contributions of this paper are summarized below.

- We present a general lower bound dependent solely on V_T^p and show that under true gradient feedback for smooth functions with vanishing gradients in the feasible domain, one can achieve the optimal dynamic regret of $O(V_T^p)$ comparable to that with full information feedback.
- We present a lower bound under a noisy (sub)gradient feedback dependent on V_T^p and T, and then show that online gradient descent (OGD) with an appropriate step size can achieve the optimal dynamic regret of $O(\sqrt{V_T^p T})$ under both stochastic gradient feedback and two-point bandit feedback.
- When the loss functions are smooth, we establish an improved dynamic regret under the two-point bandit feedback, which could match the bound achieved with the full information feedback in a certain condition.

We note that a regularity metric similar to the path variation (possibly measured in different norms) has been explored in shifting regret analysis (Herbster & Warmuth, 1998) and drifting regret analysis (Cesa-Bianchi et al., 2012; Buchbinder et al., 2012). The regret against the shifting experts was studied in tracking the best expert, where the best sequence of minimizers are assumed to change for a constant number of times. In drifting regret analysis, the constraint is relaxed to that the path variation is small. In fact, a similar dynamic regret bound to $\sqrt{V_T^p T}$ has been established for online convex optimization over the simplex (Cesa-Bianchi et al., 2012), where the path variation is measured in ℓ_1 norm. The present work focuses on OCO in the Euclidean space and considers noisy gradient feedback. A more general variation is considered in (Hall & Willett, 2013), where a sequence of (or a family of) dynamic models $\phi_1, \ldots, \phi_T$ are revealed by the environment for the learner to predict the decision in the next step. Their variation is defined as $V_T^\phi = \sum_{t=1}^{T-1} \|\mathbf{w}_{t+1}^* - \phi_t(\mathbf{w}_t^*)\|$ for a sequence of comparators and their dynamic regret scales as $V_T^\phi \sqrt{T}$, which is worse than our bounds when $\phi_t(\mathbf{w}) = \mathbf{w}$.

There have been different notions of variation that measure the point-wise changes in the sequence of loss functions at any feasible points. For example, Besbes et al. (2013) considered the functional variation defined as

$$V_T^f = \sum_{t=1}^{T-1} \max_{\mathbf{w}\in\Omega} |f_t(\mathbf{w}) - f_{t+1}(\mathbf{w})|. \tag{4}$$

Besbes et al. considered two feedback structures, i.e., the noisy gradient and the noisy cost, and established sublinear dynamic regret for both feedback structures. For Lipschitz continuous loss functions, their results are presented in Table 1 [1]. An annoying fact is that even the sequence of Lipschitz loss functions change slowly (namely the functional variation is small), Besbes' dynamic regret is worse than $O(\sqrt{T})$, the optimal rate for static regret. In comparison, our results match that for static regret when the clairvoyant moves slowly such that the path variation is a constant. Another variation that measures point-wise difference between loss functions is the gradient variation introduced in (Chiang et al., 2012), which is defined as

$$V_T^g \triangleq \sum_{t=1}^{T} \max_{\mathbf{w}\in\Omega} \|\nabla f_t(\mathbf{w}) - \nabla f_{t-1}(\mathbf{w})\|_2^2. \tag{5}$$

The gradient variation has been explored for bounding the static regret (Chiang et al., 2012; Rakhlin & Sridharan, 2013; Yang et al., 2014). Recently, Jadbabaie et al. (2015) used the three variations and developed possibly better dynamic regret than using a single variation measure for non-strongly convex loss functions. They considered the full information feedback (i.e., the whole loss function is revealed to the learner) and a true gradient feedback for a sequence of bounded functions. Their results are also presented in Table 1. In comparison, our results could be potentially better when the clairvoyant moves slowly. Different from (Jadbabaie et al., 2015), we consider the noisy

[1] For strongly convex loss functions, better bounds were also established in (Besbes et al., 2013)

Table 1. Summary of dynamic regret bounds in this paper and comparison with the previous work. N.B. means that no better bounds are explicitly given. However, the bounds in the degenerate case may apply. N.A. means that not available. * marks the result is restricted to a family of smooth loss functions with vanishing gradients in the feasible domain. Please note that the bandit feedback in this work refers to two-point bandit feedback, and that in (Besbes et al., 2013) refers to noisy one-point bandit feedback.

		This paper	(Besbes et al., 2013)	(Jadbabaie et al., 2015)
Loss function	Feedback	path variation	functional variation	three variations
Lipschitz	Full Information	$O(V_T^p)$	$O(V_T^f)$	$O(\min(\sqrt{V_T^p V_T^g}, (V_T^g)^{1/3}T^{1/3}(V_T^f)^{1/3}))$
Lipschitz	True Gradient	N.B.	N.B.	$O(\sqrt{V_T^g V_T^p})$
Smooth	True Gradient	$O(V_T^p)^*$	N.B.	N.B.
Lipschitz	Stochastic Grad.	$O(\sqrt{V_T^p T})$	$O((V_T^f)^{1/3}T^{2/3})$	N.A.
Lipschitz	Bandit	$O(\sqrt{V_T^p T})$	$O((V_T^f)^{1/5}T^{4/5})$	N.A.
Smooth/Linear	Bandit	$O(\max\{\sqrt{V_T^p V_T^g}, V_T^p\})$	N.B.	N.A.
Lower Bounds		Yes	Yes	No

gradient feedback (including the bandit feedback) and develop both upper bounds and lower bounds.

2. Optimal Dynamic Regret with Noiseless Information

In this section, we present an optimal dynamic regret bound dependent solely on V_T^p and present algorithms with matched upper bounds.

2.1. Preliminaries and a Lower Bound

Since it is impossible to achieve a sublinear dynamic regret for any sequence of loss functions. We consider the following family of functions that admit a path variation constraint:

$$\mathcal{V}_p = \{\{f_1, \ldots, f_T\} : V_T^p \leq B_T\} \tag{6}$$

where B_T is the budget. For a (randomized) policy π that generates a sequence of solutions $\mathbf{w}_1, \ldots, \mathbf{w}_T$ for a sequence of loss functions $f_1, \ldots, f_T$ under the feedback structure ϕ, its **dynamic regret** is defined as

$$\mathcal{R}_\phi^\pi(\{f_1, \ldots, f_T\}) = \mathrm{E}^\pi\left[\sum_{t=1}^T f_t(\mathbf{w}_t)\right] - \sum_{t=1}^T f_t(\mathbf{w}_t^*)$$

The worst dynamic regret of π over $f \in \mathcal{V}_p$ is

$$\mathcal{R}_\phi^\pi(\mathcal{V}_p, T) = \sup_{f \in \mathcal{V}_p} \mathcal{R}_\phi^\pi(\{f_1, \ldots, f_T\})$$

Note that the dynamic regret remains the same for different sequences of optimal solutions $\mathbf{w}_t^*, t = 1, \ldots, T$.

Below, we establish a general lower bound of the dynamic regret for the following family of policies:

$$\mathcal{A} = \left\{\pi : \mathbf{w}_t = \begin{cases} \pi_1(U), & t = 1 \\ \pi_t(\{\phi_\tau(f_\tau)\}_{\tau=1}^{t-1}, U), & t \geq 1 \end{cases}\right\} \tag{7}$$

where $\phi_t(f_t) \in \mathbb{R}^k$ denotes any feedback of f_t, and $U \in \mathbb{U}$ denotes a random variable, $\pi_1 : \mathbb{U} \to \Omega$, $\pi_t : \mathbb{R}^{(t-1)k} \times \mathbb{U} \to \Omega$ are measurable functions.

Proposition 1. *Let C, C_1, C_2 be positive constants independent of T and V_T^p, and let π be any policy in $\mathcal{A}$. (i) If $B_T \geq C_1 T$, then there exists a positive constant C_2 such that $\mathcal{R}_\phi^\pi(\mathcal{V}_p, T) \geq C_2 T$. (ii) For any $\gamma \in (0, 1)$, there exists a sequence of loss functions $f_1, \ldots, f_T$ and a positive constant C such that $V_T^p \leq o(T)$ and $\mathcal{R}_\phi^\pi(\{f_1, \ldots, f_t\}) \geq C(V_T^p)^\gamma$.*

Remark: The first part indicates that it is impossible to achieve a sublinear dynamic regret if there is no constraint on the sequence of loss functions. Therefore, in the sequel, we only consider $B_T \leq o(T)$. A similar result to the first part using V_T^f as the regularity measure has been established (Besbes et al., 2013). The second part is novel, which indicates that it is impossible to achieve a better dynamic regret bound of $O((V_T^p)^\alpha)$ with $\alpha < 1$. If otherwise, it then contradicts to the lower bound in the second part of Proposition 1.

Proof. Fix $T \geq 1$ and $\gamma \in (0, 1)$. To generate the sequence of loss functions, we create a sequence of random variables $\varepsilon_1, \ldots, \varepsilon_T$, where each ε_t is sampled independently from $\{\sigma, -\sigma\}$ with equal probabilities. It is obvious that $\mathrm{E}[\varepsilon_t] = 0$ and $\mathrm{E}[\varepsilon_t^2] = \sigma^2$. For each ε_t, we define a loss function $f_t(w) = \frac{1}{2}(w - \varepsilon_t)^2$. Assume $\sigma \in (0, 1)$ whose value will be specified later. Let the feasible domain be $\Omega = [-1, 1]$.

$$\mathrm{E}\left[\mathcal{R}_\phi^\pi(\{f_1, \ldots, f_t\})\right] = \mathrm{E}\left[\sum_{t=1}^T f_t(w_t) - f_t(\varepsilon_t)\right]$$

$$= \sum_{t=1}^T \mathrm{E}\left[w_t^2/2 + \varepsilon_t^2/2 - w_t\varepsilon_t\right] \geq T\sigma^2/2$$

where $\mathrm{E}[\cdot]$ denotes the expectation over the randomness in the sequence of loss functions $f_1, \ldots, f_t$ and the policy π and the last inequality is due to that w_t is independent of ε_t. We also have $V_T^p = \sum_{t=1}^{T-1} |\varepsilon_t - \varepsilon_{t+1}| \leq 2\sigma T$. To

prove the first part, we let σ be a constant $C_1/2$, then any sequences of loss functions generated as above constitute a subset $\mathcal{V}'_p \subset \mathcal{V}_p$. Then

$$\mathcal{R}^\pi_\phi(\mathcal{V}_p, T) \geq \mathcal{R}^\pi_\phi(\mathcal{V}'_p, T) \geq \mathrm{E}\left[\mathcal{R}^\pi_\phi(\{f_1, \ldots, f_t\})\right] \geq \frac{C_1^2}{8} T$$

To prove the second part, we set $\sigma = T^{-\mu}$ with $\mu = (1-\gamma)/(2-\gamma) \in (0, 1/2)$. Then there exits a positive constant C such that $\mathrm{E}\left[\mathcal{R}^\pi_\phi(\{f_1, \ldots, f_T\}) - C(V_T^p)^\gamma\right] \geq 0$, which implies that there exists a sequence of loss functions $f_1, \ldots, f_T$ such that $\mathcal{R}^\pi_\phi(\{f_1, \ldots, f_T\}) \geq C(V_T^p)^\gamma$.

We note that if $\gamma = 1$, we have $\mu = 0$ and therefore $B_T = \Omega(T)$ which reduces to the lower bound in the first part. Therefore, we restrict $\gamma \in (0, 1)$. □

An interesting question is that whether an $O(V_T^p)$ dynamic regret bound is achievable, if not what is the best we can achieve. In particular, we are interested in scenarios when the feedback $\phi_t(f_t) = \phi_t(\mathbf{w}_t, f_t)$ only gives a (noisy) gradient of $f_t(\mathbf{w})$ at $\mathbf{w}_t$. Before delving into the noisy gradient feedback, we first show that an $O(V_T^p)$ upper bound is achievable with full information of the loss functions or with full gradient feedback provided that the loss functions are smooth and have vanishing gradients. We make the following assumptions throughout the paper without explicitly mentioning it in the sequel.

Assumption 1. *For $\{f_1, \ldots, f_T\} \in \mathcal{V}_p$, there exists a $r > 0$ such that $\sup_{\mathbf{w}_t^* \in \Omega_t^*} \|\mathbf{w} - \mathbf{w}_t^*\|_2 \leq r$, for any $\mathbf{w} \in \Omega$ and $1 \leq t \leq T$.*

2.2. Online Learning with Full Information

Assume that at each step the full information of the loss function $f_t(\mathbf{w})$ is revealed after the decision $\mathbf{w}_t$ is submitted, and each loss function $f_t(\mathbf{w})$ is G-Lipschitz continuous. Then we can update $\mathbf{w}_{t+1}$ by

$$\mathbf{w}_{t+1} = \min_{\mathbf{w} \in \Omega} f_t(\mathbf{w}), t \geq 1$$

with $\mathbf{w}_1$ be any point in Ω. To analyze the dynamic regret, we denote by $\mathbf{w}_0^* = \mathbf{w}_1$.

$$\sum_{t=1}^T f_t(\mathbf{w}_t) - \sum_{t=1}^T f_t(\mathbf{w}_t^*) = \sum_{t=1}^T f_t(\mathbf{w}_{t-1}^*) - \sum_{t=1}^T f_t(\mathbf{w}_t^*)$$
$$\leq \sum_{t=1}^T G\|\mathbf{w}_{t-1}^* - \mathbf{w}_t^*\|_2 = G\|\mathbf{w}_1 - \mathbf{w}_1^*\|_2$$
$$+ G\sum_{t=1}^{T-1} \|\mathbf{w}_t^* - \mathbf{w}_{t+1}^*\|_2 \leq Gr + GV_T^p = O(\max(V_T^p, 1)).$$

It is notable that a similar upper bound of $O(\max(V_T^f, 1))$ with the full information can be achieved (Jadbabaie et al., 2015).

2.3. Online Learning with Gradient Feedback

Full information may not be available. In practice, only some partial information of the $f_t(\mathbf{w})$ regarding the decision vector $\mathbf{w}_t$ is available. In this subsection, we assume that only the gradient information $\nabla f_t(\mathbf{w}_t)$ is available after the decision $\mathbf{w}_t$ is submitted. Below, we will first present several examples showing that $O(V_T^p)$ is achievable and generalize the analysis to a broad family.

We consider two loss functions $g_1(w) = \max(w, 0)^2$ and $g_2(w) = (w - \alpha)^2$ defined in the domain $\Omega = [-1, 3]$ and divide all iterations $1, \ldots, T$ into a number m of batches with each batch size of Δ_T. Assume the adversary selects $g_1(\cdot)$ in odd batches and $g_2(\cdot)$ in even batches, and at each step the full gradient feedback is available, i.e., $\phi_t(w_t, f_t) = f'_t(w_t)$. The example is similar to that presented in (Besbes et al., 2013) except that $g_1(w)$ is not strongly convex. Below, we consider two instances of the above example with different Δ_T and α. For the updates, we adopt the OGD, i.e.,

$$w_{t+1} = \Pi_\Omega[w_t - \eta f'_t(w_t)], \quad t = 1, \ldots, T-1$$

where $\Pi_\Omega[\cdot]$ denotes the projection into the domain Ω.

Instance 1. $\Delta_T = \lceil T/2 \rceil$ and $\alpha = 1$. Then $V_T^p = 1$.

Given the value of Δ_T, there are two batches. Let $\Gamma_1, \Gamma_2 \subseteq T$ denote the iteration indices in the first and the second batch, respectively, and let $\Gamma_j[1]$ denote the first iteration of the j-th batch. We adopt a constant step size $\eta = 1/2$ with a starting point $w_0 = 0$. Then $w_t = 0, t \in \Gamma_1$. For the first iteration $t \in \Gamma_2$ we have $w_t = \Pi_\Omega[0 - \eta g'_1(0)] = 0$. And for all remaining iterations $t \in \Gamma_2$, we have $w_t = \Pi_\Omega[w_{t-1} - \eta g'_2(w_{t-1})] = \alpha$. As a result $w_t = w_t^*, t \in [T]$ except $w_{\Gamma_2[1]} = 0 \neq w^*_{\Gamma_2[1]}$, which indicates that the dynamic regret is $f_{\Gamma_2[1]}(w_{\Gamma_2[1]}) - f_{\Gamma_2[1]}(w^*_{\Gamma_2[1]}) = g_2(0) - g_2(1) = 1$.

Instance 2. $\theta = C/\sqrt{T}$, $\Delta_T = \lfloor 1 + 1/(2\theta) \rfloor$, $\alpha = 1 + (1 - 2\theta)^{\Delta_T}$, and $T > 4C^2$ (note that $\theta < 1/2$ and $1 \leq \alpha \leq 2$). Then

$$V_T^p = \sum_{j=1}^m \alpha \leq 2m = \frac{2T}{\Delta_T} = \frac{2T}{\lfloor 1 + \sqrt{T}/(2C) \rfloor} \leq 4C\sqrt{T}$$

The example is similar to the above except that the loss functions change more frequently. We consider OGD with a constant step size $\eta = 1/2$ and $w_1 = 1$. Similar as before, we use $j = 1, \ldots, m$ to denote the batch index, $\Gamma_j \subseteq T$ to denote the indices in the j-th batch, $\Gamma_j[1]$ and $\Gamma_j[2:]$ to denote the first iteration and remaining iterations in batch j, respectively. Note that $w_t^* = 0, t \in \Gamma_{2j-1}$ and $w_t^* = \alpha, t \in \Gamma_{2j}$.

For $t \in \Gamma_1[2:]$ or $t = \Gamma_2[1]$, by induction we can show that $w_t = \Pi_\Omega[w_{t-1} - \eta g'_1(w_{t-1})] = 0$. Therefore, $w_t = w_t^*, t \in \Gamma_1[2{:}]$ and $w_{\Gamma_2[1]} = 0$. For $t \in \Gamma_2[2{:}]$ or $t = \Gamma_3[1]$,

following the OGD update $w_t = \Pi_\Omega[w_{t-1} - \eta g_2'(w_{t-1})] = \Pi_\Omega[w_{t-1} - 2\eta(w_{t-1} - \alpha)] = \alpha$. Therefore, $w_t = \alpha, t \in \Gamma_2[2{:}]$ and $t = \Gamma_3[1]$. Following the same analysis, we have $w_t = w_t^*$ for $t \in \Gamma_j[2{:}]$, $w_{\Gamma_{2j-1}[1]} = \alpha$ and $w_{\Gamma_{2j}[1]} = 0$. It means that the difference between the decision vector w_t and the optimal solutions w_t^* only happens at the first iterations of all batches. As a result, the dynamic regret is

$$\sum_{j=1}^{m} \max(g_1(\alpha) - g_1(0), g_2(0) - g_2(\alpha)) = \sum_{j=1}^{m} \alpha^2 \leq 2V_T^p$$

It is notable the key ingredient to achieve an $O(V_T^p)$ dynamic regret is to use a constant step size. Next, we generalize this result to a broad family of loss functions. In particular, we assume the sequence of loss functions satisfy the following assumption.

Assumption 2. *Assume that every loss function $f_t(\cdot)$ is defined over $\mathbb{R}^d$ and is convex and smooth, i.e., for any $\mathbf{w}, \mathbf{w}' \in \mathbb{R}^d$, we have*

$$\|\nabla f_t(\mathbf{w}) - \nabla f_t(\mathbf{w}')\|_2 \leq L\|\mathbf{w} - \mathbf{w}'\|_2,$$

where $L > 0$ is the smoothness constant. In addition, we assume that there exists $\mathbf{w}_t^* \in \Omega_t^*$ such that $\nabla f_t(\mathbf{w}_t^*) = 0$.

The condition $\nabla f_t(\mathbf{w}_t^*) = 0$ is referred to as the vanishing gradient condition. The examples considered before indeed satisfy **Assumption 2**. Consider the policy of OGD:

$$\pi: \ \mathbf{w}_t = \begin{cases} \mathbf{w}_1 \in \Omega & t = 1 \\ \Pi_\Omega[\mathbf{w}_{t-1} - \eta \nabla f_{t-1}(\mathbf{w}_{t-1})] & t > 1 \end{cases} \tag{8}$$

The following theorem states the dynamic regret bound of OGD with a constant step size.

Theorem 3. *(upper bound) Suppose* **Assumption 2** *hold. By the policy π in (8) with $\eta = 1/(2L)$, for any $\{f_1, \ldots, f_T\} \in \mathcal{V}_p$ we have*

$$\sum_{t=1}^{T} f_t(\mathbf{w}_t) - f_t(\mathbf{w}_t^*) \leq 2L\left(r^2 + 2rB_T\right).$$

To prove the theorem, we first give the following lemma whose proof is included the supplement.

Lemma 4. *Let $\mathbf{w}_t = \Pi_\Omega[\mathbf{w}_{t-1} - \eta \mathbf{g}_{t-1}], t > 1$. Then*

$$\mathbf{g}_t^\top(\mathbf{w}_t - \mathbf{w}_t^*) \leq \frac{\eta}{2}\|\mathbf{g}_t\|_2^2 + \frac{r\|\mathbf{w}_t^* - \mathbf{w}_{t+1}^*\|_2}{\eta} + \frac{\|\mathbf{w}_t - \mathbf{w}_t^*\|_2^2 - \|\mathbf{w}_{t+1} - \mathbf{w}_t^*\|_2^2 - \|\mathbf{w}_t^* - \mathbf{w}_{t+1}^*\|_2^2}{2\eta}$$

Proof of Theorem 3. Following Lemma 4 and the convexity of $f_t(\mathbf{w})$, we have

$$f_t(\mathbf{w}_t) - f_t(\mathbf{w}_t^*) \leq \frac{\eta}{2}\|\nabla f_t(\mathbf{w}_t)\|_2^2 + \frac{r\|\mathbf{w}_t^* - \mathbf{w}_{t+1}^*\|_2}{\eta} + \frac{\|\mathbf{w}_t - \mathbf{w}_t^*\|_2^2 - \|\mathbf{w}_{t+1} - \mathbf{w}_{t+1}^*\|_2^2 - \|\mathbf{w}_{t+1}^* - \mathbf{w}_t^*\|_2^2}{2\eta} \tag{9}$$

By the smoothness of $f(\mathbf{w})$, for any $\mathbf{w} \in \mathbb{R}^d$

$$f_t(\mathbf{w}) - f_t(\mathbf{w}_t) \leq \langle \nabla f_t(\mathbf{w}_t), \mathbf{w} - \mathbf{w}_t \rangle + \frac{L}{2}\|\mathbf{w} - \mathbf{w}_t\|_2^2$$

Let $\mathbf{w} = \mathbf{w}_t' = \mathbf{w}_t - \frac{1}{L}\nabla f_t(\mathbf{w}_t)$ in the above inequality, we have $f_t(\mathbf{w}_t') - f_t(\mathbf{w}_t) \leq -\frac{\|\nabla f_t(\mathbf{w}_t)\|_2^2}{2L}$. By convexity of $f_t(\mathbf{w})$,

$$f_t(\mathbf{w}_t') \geq f_t(\mathbf{w}_t^*) + \nabla f_t(\mathbf{w}_t^*)^\top(\mathbf{w}_t' - \mathbf{w}_t^*) = f_t(\mathbf{w}_t^*)$$

which follows the vanishing gradient condition. Then

$$f_t(\mathbf{w}_t^*) - f_t(\mathbf{w}_t) \leq f_t(\mathbf{w}_t') - f_t(\mathbf{w}_t) \leq -\frac{\|\nabla f_t(\mathbf{w}_t)\|_2^2}{2L}$$

Combing the inequality above with (9), we have

$$f_t(\mathbf{w}_t) - f_t(\mathbf{w}_t^*) \leq \eta L(f_t(\mathbf{w}_t) - f_t(\mathbf{w}_t^*)) + \frac{\|\mathbf{w}_t - \mathbf{w}_t^*\|_2^2 - \|\mathbf{w}_{t+1} - \mathbf{w}_{t+1}^*\|_2^2}{2\eta} + \frac{r\|\mathbf{w}_t^* - \mathbf{w}_{t+1}^*\|_2}{\eta}$$

By summing over $t = 1, \ldots, T$, we have

$$\sum_{t=1}^{T}(f_t(\mathbf{w}_t) - f_t(\mathbf{w}_t^*)) \leq \frac{1}{1 - \eta L}\left(\frac{r^2}{2\eta} + \frac{r}{\eta}V_T^p\right)$$

We complete the proof by choosing $\eta = 1/[2L]$. □

Remark: Theorem 3 exhibits that OGD can achieve an $O(\max(V_T^p, 1))$ dynamic regret for a sequence of loss functions in $\mathcal{V}_p$ that satisfy **Assumption 2** with only the gradient feedback, which is comparable to that achieved in the full information feedback. The instance 1 and 2 in Section 2 has $V_T^p = O(1)$ and $V_T^p \approx 4C\sqrt{T}$, respectively. Therefore, using OGD with $\eta = C$ we can obtain an $O(1)$ and $O(\sqrt{T})$ dynamic regret.

Finally, it is worth mentioning that the OGD with restarting proposed in Besbes et al.' work achieves an $O(T^{1/3})$ dynamic regret for instance 1 and an $O(T^{5/6})$ dynamic regret for instance 2 due to that the functional variation for the first instance is bounded by a constant and for the second instance is bounded by $O(\sqrt{T})$.

3. Optimal Dynamic Regret with Noisy Gradient

In this section, we focus on noisy gradient feedback, i.e., $\phi_t(\mathbf{w}_t, f_t)$ is only a noisy (sub)gradient of $f_t(\mathbf{w})$ at $\mathbf{w}_t$.

3.1. A Lower Bound with Noisy Gradient Feedback

Before presenting the upper bounds of the dynamic regret with noisy gradient feedback, we will first present a lower bound. For establishing the lower bound, we consider the following class of policies

$$\pi: \ \mathbf{w}_t = \begin{cases} \pi_1(U), & t = 1 \\ \pi_t(\{\phi_\tau(\mathbf{w}_\tau, f_\tau)\}_{\tau=1}^{t-1}, U), & t \geq 1 \end{cases} \tag{10}$$

where U, π_1, π_t are defined similarly as before, and $\phi_t(\mathbf{w}_t, f_t)$ is a noisy subgradient of f_t at $\mathbf{w}_t$. In particular, we assume the noisy gradient is given by $\phi_t(\mathbf{w}_t, f_t) \in \partial f_t(\mathbf{w}_t) + \epsilon_t$ with ϵ_t satisfying the following condition.

Assumption 5. *$\epsilon_t \in \mathbb{R}^d, t \geq 1$ are iid random vectors with zero mean and covariance matrix Σ with bounded entries such that $tr(\Sigma) \leq \lambda^2$. Let $P(\cdot)$ denote the cumulative function of ϵ_t. There exists a constant $\tilde{C}$ such that for any $a \in \mathbb{R}^d$, $\int \log\left(\frac{dP(\mathbf{y})}{dP(\mathbf{y}+\mathbf{a})}\right) dP(\mathbf{y}) \leq \tilde{C}\|\mathbf{a}\|_2^2$.*

When the noise vectors are independent Gaussian random vectors with zero mean and covariance matrices with entries uniformly bounded by σ^2, the above assumption is satisfied with $\tilde{C} = 1/(2\sigma^2)$ and $\lambda^2 = d\sigma^2$ (Besbes et al., 2013).

Theorem 6. *(lower bound) For any $1 \leq B_T \leq T$ and $\kappa \in (1/2, 1)$, there exist $\mathcal{V}_p' \subset \mathcal{V}_p$ and $C(\kappa) > 0$ independent of T and B_T such that for any policy π in (10) under the noisy gradient feedback that satisfies* **Assumption 5***, we have*

$$\mathcal{R}_\phi^\pi(\mathcal{V}_p', T) \geq C(\kappa) B_T^\kappa T^{1-\kappa}.$$

Remark: Note that from the proof presented below when $\kappa \to 1/2$, $C(\kappa) \to 0$. However, the above lower bound can be used to argue that it is impossible to achieve a better dynamic regret than $O(B_T^{1/2}T^{1/2})$ with the noisy gradient feedback for any sequence of loss functions. We prove this by contradiction. In particular, assume there exists an algorithm under the noisy gradient feedback achieves better bound than $O(B_T^{1/2}T^{1/2})$ for any sequence of loss functions. We can consider two lower orders $O(B_T^\alpha T^{1-\alpha})$ with $1 > \alpha > 1/2$ and $O(B_T^\alpha T^\beta)$ with $\alpha \leq 1/2, \beta \leq 1/2$ and $\alpha + \beta < 1$. First, we assume $O(B_T^\alpha T^{1-\alpha})$ is achievable. By Theorem 6 we know that there exists $\alpha < \kappa < 1$ (e.g., $\kappa = \frac{\alpha+1}{2}$) and $\mathcal{V}_p'$ such that $\mathcal{R}_\phi^\pi(\mathcal{V}_p', T) \geq \Omega(B_T^\kappa T^{1-\kappa}) \geq \Omega(B_T^\alpha T^{1-\alpha})$, which yields a contradiction. To show that the second lower bound is unachievable, we can construct a B_T such that $B_T^\kappa T^{1-\kappa} \geq \Omega(B_T^\alpha T^\beta)$, i.e., $B_T \geq \Omega(T^{\frac{\beta+\kappa-1}{\kappa-\alpha}})$, where $\beta + \kappa - 1 < \kappa - \alpha$.

Proof. We construct two functions over the domain $\Omega = [-1/2, +1/2]$. They are

$$f(x) = \begin{cases} \frac{1}{1+\gamma}\delta^{1+\gamma} - \delta^\gamma x & x \in [-1/2, 0] \\ \frac{1}{1+\gamma}|x-\delta|^{1+\gamma} & x \in [0, 2\delta] \\ -\frac{1+2\gamma}{1+\gamma}\delta^{1+\gamma} + \delta^\gamma x & x \in [2\delta, 1/2] \end{cases}, \quad (11)$$

$$g(x) = \begin{cases} -\frac{1+2\gamma}{1+\gamma}\delta^{1+\gamma} - \delta^\gamma x & x \in [-1/2, -2\delta] \\ \frac{1}{1+\gamma}|x+\delta|^{1+\gamma} & x \in [-2\delta, 0] \\ \frac{1}{1+\gamma}\delta^{1+\gamma} + \delta^\gamma x & x \in [0, 1/2] \end{cases} \quad (12)$$

where $0 < \delta < 1/2$ and $\gamma > 0$ will be determined later. It is easy to verify that both $f(\cdot)$ and $g(\cdot)$ are convex but not strongly convex. It is also easy to see that the optimal solutions for $f(\cdot)$ and $g(\cdot)$ are $x_f^* = \delta$ and $x_g^* = -\delta$, respectively. Hence $|x_f^* - x_g^*| = 2\delta$. For a given budget B_T, we will construct a subset of $\mathcal{V}_p$ by only considering the sequence of these two loss functions. For some $\Delta_T \in \{1, \ldots, T\}$ that, we divide the entire sequence T into $m = \lceil T/\Delta_T \rceil$ batches, denoted by $\mathcal{T}_1, \ldots, \mathcal{T}_m$, each with size of Δ_T (except perhaps $\mathcal{T}_m$), i.e. $\mathcal{T}_j = \{(j-1)\Delta_T + 1, \ldots, \min(j\Delta_t, T)\}, j = 1, \ldots, m$. To generate the sequence of loss functions $f_1, \ldots, f_T$, at the beginning of each batch $\mathcal{T}_j$, we randomly choose between the two functions $f(\cdot)$ and $g(\cdot)$ and the same loss function will be used throughout the batch. We denote by $\mathcal{V}_p' = \{\{f_t(\cdot), t = 1, \ldots, T\}\}$ the set of a sequence of randomly sampled loss functions, and by $X_1, \ldots, X_T$ the sequence of solutions generated by any policy in (10). Let $\delta = B_T\Delta_T/2T$. For any $f \in \mathcal{V}_p'$, we have

$$V_T^p = \sum_{j=2}^m |x_j^* - x_{j-1}^*| \leq (\lceil T/\Delta_T \rceil - 1)\, 2\delta \leq \frac{T2\delta}{\Delta_T} = B_T$$

Therefore, $\mathcal{V}_p' \subset \mathcal{V}_p$. We denote by $\mathbb{P}_f^\pi$ the probability measure under policy π when f is the sequence of the loss functions, and by E_f^π the associated expectation operator. Set $\Delta_T = \max\left\{(\frac{4^\gamma}{4\tilde{C}})^{1/(2\gamma+1)}(\frac{T}{B_T})^{2\gamma/(2\gamma+1)}, 1\right\}$. Then

$$\begin{aligned} \tilde{C}\mathrm{E}_f^\pi\left[\sum_{t\in\mathcal{T}_j}(\nabla f(X_t) - \nabla g(X_t))^2\right] &\leq \tilde{C}\sum_{t\in\mathcal{T}_j} 4\delta^{2\gamma} \\ &\leq 4\tilde{C}\Delta_T\delta^{2\gamma} \leq 4\tilde{C}\frac{B_T^{2\gamma}\Delta^{2\gamma+1}}{2^{2\gamma}T^{2\gamma}} \leq \max(1, \frac{4\tilde{C}B_T^{2\gamma}}{T^{2\gamma}4^\gamma}) \\ &\leq \max(1, \frac{4\tilde{C}}{4^\gamma}) \leq \max(1, 4\tilde{C}) \triangleq \beta \end{aligned}$$

where we use the condition that $B_T \leq T$. Using Lemma A-1 and A-2 from (Besbes et al., 2013), we have

$$\max\left\{\mathbb{P}_f^\pi\{X_t > 0\}, \mathbb{P}_g^\pi\{X_t \leq 0\}\right\} \geq \frac{1}{4e^\beta}, \forall t$$

Using the above result, we have

$$\begin{aligned} \mathcal{R}_\phi^\pi(\mathcal{V}_p', T) &\geq \mathrm{E}\left[\sum_{j=1}^m\sum_{t\in\mathcal{T}_j} f_t(X_t) - f_t(x_j^*)\right] \\ &= \frac{1}{2}\sum_{j=1}^m \mathrm{E}_f^\pi\left[\sum_{t\in\mathcal{T}_j} f(X_t) - f(x_f^*)\right] \\ &+ \frac{1}{2}\sum_{j=1}^m \mathrm{E}_g^\pi\left[\sum_{t\in\mathcal{T}_j} g(X_t) - g(x_g^*)\right] \\ &\geq \frac{1}{2}\sum_{j=1}^m\sum_{t\in\mathcal{T}_j}\mathbb{P}\{X_t \leq 0\}\left(f(0) - f(x_f^*)\right) \\ &+ \frac{1}{2}\sum_{j=1}^m\sum_{t\in\mathcal{T}_j}\mathbb{P}\{X_t > 0\}\left(g(0) - g(x_g^*)\right) \end{aligned}$$

$$= \frac{\delta^{1+\gamma}}{2(1+\gamma)} \sum_{j=1}^{m} \sum_{t \in \mathcal{T}_j} (\mathbb{P}\{X_t \leq 0\} + \mathbb{P}\{X_t > 0\})$$
$$\geq \frac{\delta^{1+\gamma} T}{8(1+\gamma)e^{\beta}}.$$

In the above derivations, the first expectation is taking over all randomness in π and $f_1, \ldots, f_T$, the second inequality holds because

$$\mathrm{E}[f(X_t)] = \mathrm{E}[f(X_t)|X_t > 0] \Pr(X_t > 0) + \mathrm{E}[f(X_t)|X_t \leq 0] \Pr(X_t \leq 0) \geq f(0) \Pr(X_t \leq 0)$$
$$\mathrm{E}[g(X_t)] = \mathrm{E}[g(X_t)|X_t > 0] \Pr(X_t > 0) + \mathrm{E}[g(X_t)|X_t \leq 0] \Pr(X_t \leq 0) \geq g(0) \Pr(X_t > 0)$$

where the inequalities are due to $f(x) \geq 0, g(x) \geq 0$ and $f(x) \geq f(0)$ when $x \leq 0$ and $g(x) \geq g(0)$ when $x \geq 0$. To proceed, we plug in the value of δ into the lower bound of $\mathcal{R}_{\phi}^{\pi}(\mathcal{V}_p', T)$

$$\mathcal{R}_{\phi}^{\pi}(\mathcal{V}_p', T) \geq \frac{T}{8e^{\beta}(1+\gamma)} \frac{B_T^{1+\gamma} \Delta_T^{1+\gamma}}{2^{1+\gamma} T^{1+\gamma}}$$
$$\geq \frac{4^{\gamma(1+\gamma)/(1+2\gamma)}}{8e^{\beta}(1+\gamma)2^{1+\gamma}(4\tilde{C})^{(1+\gamma)/(1+2\gamma)}} \frac{B_T^{1+\gamma-(1+\gamma)\frac{2\gamma}{1+2\gamma}}}{T^{\gamma-(1+\gamma)\frac{2\gamma}{1+2\gamma}}}$$
$$= \frac{4^{\gamma(1+\gamma)/(1+2\gamma)}}{8e^{\beta}(1+\gamma)2^{1+\gamma}(4\tilde{C})^{(1+\gamma)/(1+2\gamma)}} B_T^{\frac{1+\gamma}{1+2\gamma}} T^{1-\frac{1+\gamma}{1+2\gamma}}$$

Let $\gamma = (1-\kappa)/(2\kappa-1)$, i.e., $\kappa = (1+\gamma)/(1+2\gamma)$. Then

$$\mathcal{R}_{\phi}^{\pi}(\mathcal{V}_p', T) \geq \frac{(4^{\gamma})^{\kappa}}{(1+\gamma)2^{\gamma}} \frac{1}{16e^{\beta}(4\tilde{C})^{\kappa}} B_T^{\kappa} T^{1-\kappa}.$$

□

In the next two subsections, we consider two types of noisy gradient feedback, namely a stochastic subgradient feedback that is an unbiased estimation of the true subgradient and a bandit feedback that gives an unbiased estimation of the subgradient of a smoothed function instead of the original function. We show that under the two noisy gradient feedback, we are able to achieve an optimal dynamic regret of $O(\sqrt{V_T^p T})$. Furthermore, for smooth loss functions under the two-point bandit feedback, we establish an even better upper bound by leveraging the gradient variation in the form of $O(\max(\sqrt{V_T^p V_T^g}, V_T^p))$, which when the gradient variation is small matches the lower bound presented in Proposition 1.

3.2. Online Learning with Bounded Stochastic Gradient Feedback

We adopt the policy defined by OGD using the noisy gradient feedback, i.e.,

$$\pi: \mathbf{w}_t = \begin{cases} \mathbf{w}_1 \in \Omega & t = 1 \\ \Pi_{\Omega}[\mathbf{w}_{t-1} - \eta \phi_{t-1}(\mathbf{w}_{t-1}, f_{t-1})] & t > 1 \end{cases} \tag{13}$$

where $\phi_t(\mathbf{w}_t, f_t) \in \partial f_t(\mathbf{w}_t) + \epsilon_t$ is a noisy subgradient with ϵ_t satisfying **Assumption 5**. The upper bound of the dynamic regret of OGD with an appropriate step size is presented below.

Theorem 7. *(upper bound) Suppose* **Assumption 5** *hold. Assume $\|\partial f_t(\mathbf{w})\|_2 \leq G$, for any $\mathbf{w} \in \Omega$ and $1 \leq t \leq T$. By the policy π in (13) with $\eta = \sqrt{\frac{r^2+2rB_T}{T(G^2+\lambda^2)}}$, we have*

$$\mathcal{R}_{\phi}^{\pi}(\mathcal{V}_p, T) \leq \sqrt{(r^2 + 2rB_T)(G^2+\lambda^2)T}.$$

Proof. Let $\mathrm{E}_t[\cdot]$ denote the expectation over the randomness in ϕ_t given the randomness before t. We abuse the notation $\mathbf{w}_{T+1}^* = \mathbf{w}_T^*$. Note that $\mathrm{E}_t\left[\|\phi_t(\mathbf{w}_t, f_t)\|_2^2\right] \leq (G^2+\lambda^2)$. Following Lemma 4 and the convexity of $f_t(\mathbf{w})$, we have

$$\mathrm{E}_t\left[f_t(\mathbf{w}_t) - f_t(\mathbf{w}_t^*)\right] \leq \mathrm{E}_t\left[\langle \phi_t(\mathbf{w}_t, f_t), \mathbf{w}_t - \mathbf{w}_t^* \rangle\right]$$
$$\leq \frac{\|\mathbf{w}_t - \mathbf{w}_t^*\|_2^2}{2\eta} - \frac{\|\mathbf{w}_{t+1} - \mathbf{w}_{t+1}^*\|_2^2}{2\eta} - \frac{\|\mathbf{w}_t^* - \mathbf{w}_{t+1}^*\|_2^2}{2\eta} + \frac{\eta}{2}(G^2+\lambda^2) + \frac{r}{\eta}\|\mathbf{w}_t^* - \mathbf{w}_{t+1}^*\|_2$$

Hence, by summing the above inequalities over $t = 1, \ldots, T$ we have

$$\mathrm{E}\left[\sum_{t=1}^{T} f_t(\mathbf{w}_t) - f_t(\mathbf{w}_t^*)\right] \leq \frac{1}{2\eta}\left(r^2 + 2rV_T^p\right) + \frac{\eta}{2} G_{\lambda}^2 T$$

where $G_{\lambda}^2 = G^2 + \lambda^2$. Since the above inequality holds for any $\mathbf{w}_t^* \in \Omega_t^*$, we thus conclude

$$\mathrm{E}\left[\sum_{t=1}^{T} f_t(\mathbf{w}_t) - f_t(\mathbf{w}_t^*)\right] \leq \frac{1}{2\eta}(r^2 + 2rB_T) + \frac{\eta}{2} G_{\lambda}^2 T$$

We complete the proof by choosing $\eta = \sqrt{\frac{r^2+2rB_T}{T(G^2+\lambda^2)}}$. □

Remark: From Theorem 7, we can see that OGD can achieve an $O(\sqrt{\max(V_T^p, 1)T})$ dynamic regret with a step size $\eta = C\sqrt{\max(V_T^p, 1)/T}$. Compared to OGD with restarting proposed in (Besbes et al., 2013), our result could be better when $\sqrt{V_T^p T} \leq O((V_T^f)^{1/3} T^{2/3})$, i.e., $V_T^p \leq O((V_T^f)^{2/3} T^{1/3})$.

3.3. Online Learning with Bandit Feedback

In this subsection, we analyze the dynamic regret with bandit feedback by building on previous work. Bandit feedback has been analyzed before for the static regret. In particular, using one-point bandit feedback Flaxman et al. (2005) showed an $O(T^{3/4})$ static regret bound, while Agarwal et al. (2010) established an optimal static regret bound of $O(\sqrt{T})$ using two-point bandit feedback. Recently, Chiang et al. (2013) derived a variational static regret bound in the two-point bandit setting that depends on $\sqrt{V_T^g}$ where V_T^g is the gradient variation defined in Section 1. In order

to have optimal dynamic regret bounds, we also consider two-point bandit setting and show that the previous algorithms in (Agarwal et al., 2010; Chiang et al., 2013) by adjusting the step size can achieve an $O(\sqrt{V_T^p T})$ dynamic regret for general Lipschitz continuous loss functions and an $O(\max(\sqrt{V_T^g V_T^p}, V_T^p))$ dynamic regret for smooth loss functions. Below, we present more details. The omitted proof can be found in the supplement.

Similar to previous work, we assume that $f_t(\mathbf{w})$ is G-Lipschitz continuous and $R_1\mathbb{B} \subseteq \Omega \subseteq R_2\mathbb{B}$ where $\mathbb{B} = \{\mathbf{w} \in \mathbb{R}^d : \|\mathbf{w}\|_2 \leq 1\}$ is the unit ball centered at 0. Let $\mathbf{u}_t \in \mathbb{R}^d$ be a random unit vector, $\mathbf{e}_i \in \mathbb{R}^d$ be the i-th canonical vector, $\mathbf{w}_1 = 0$. For Lipschitz continuous loss functions, the update is given by

$$\mathbf{w}_{t+1} = \Pi_{(1-\xi)\Omega}[\mathbf{w}_t - \hat{\mathbf{g}}_t] \tag{14}$$

where $\xi \in (0, 1)$ and $\hat{\mathbf{g}}_t$ is computed from two-point bandit feedback

$$\hat{\mathbf{g}}_t = \frac{d}{2\delta}[f_t(\mathbf{w}_t + \delta\mathbf{u}_t) - f_t(\mathbf{w}_t - \delta\mathbf{u}_t)]\mathbf{u}_t$$

with $\delta = \xi R_1$. For any $\mathbf{w}_t \in (1-\xi)\Omega$ and any unit vector $\mathbf{u}$, $\mathbf{w}_t + \delta\mathbf{u} \in \Omega$ (Flaxman et al., 2005). It can be shown that $\hat{\mathbf{g}}_t$ is an unbiased stochastic gradient of the function $\hat{f}_t(\mathbf{w}) = \mathrm{E}_{\mathbf{u}}[f_t(\mathbf{w} + \delta\mathbf{u})]$. Importantly, $\|\hat{\mathbf{g}}_t\|_2 \leq Gd$. The following theorem states the dynamic regret bound for the policy in (14).

Theorem 8. *Assume $f_t(\mathbf{w})$ is G-Lipschitz continuous. By the policy in (14) with $\xi = \frac{1}{T}$, $\delta = \xi R_1$, and $\eta = \sqrt{\frac{r^2+2rB_T}{TG^2d^2}}$, we have*

$$\mathrm{E}\left[\sum_{t=1}^{T}\frac{1}{2}(f_t(\widehat{\mathbf{w}}_t^1) + f(\widehat{\mathbf{w}}_t^2))\right] - \sum_{t=1}^{T} f_t(\mathbf{w}_t^*) \leq \sqrt{(r^2 + 2rB_T)G^2d^2T} + G(3R_1 + R_2).$$

where $\widehat{\mathbf{w}}_t^1 = \mathbf{w}_t + \delta\mathbf{u}_t, \widehat{\mathbf{w}}_t^2 = \mathbf{w}_t - \delta\mathbf{u}_t$.

Remark: The dynamic regret averaged over two decisions with two-point bandit feedback is in the same order of $\sqrt{\max(V_T^p, 1)T}$ to that in Theorem 7 with stochastic gradient feedback.

Finally, we present an upper bound for smooth loss functions by leveraging the gradient variation, which leads to an improved dynamic regret bound compared to Lipschitz continuous loss functions. The updates are based on the META algorithm proposed in (Chiang et al., 2013), which is presented in Algorithm 1. It was proved to achieve a better static regret of $O(\sqrt{V_T^g})$ than $O(\sqrt{T})$. Below, we show that the same policy but with a different step size can achieve an improved dynamic regret, i.e., $O(\max(\sqrt{V_T^g \max(V_T^p, 1)}, V_T^p))$ for a sequence of smooth loss functions from the following set

$$\mathcal{V}_{p,g} = \{\{f_1, \ldots, f_T\} : V_T^p \leq B_T, V_T^g \leq S_T\}$$

Algorithm 1 META algorithm

Initialize solution $\mathbf{w}_1 = \widehat{\mathbf{w}}_1 = 0$ and $\widehat{\mathbf{g}}_0 = 0$
for $t = 1, \ldots, T$ **do**
  Choose i_t uniformly from $[d]$
  Submit $\widehat{\mathbf{w}}_t^1 = \widehat{\mathbf{w}}_t + \delta\mathbf{e}_{i_t}$ and $\widehat{\mathbf{w}}_t^2 = \widehat{\mathbf{w}}_t - \delta\mathbf{e}_{i_t}$
  Receive the feedback $f_t(\widehat{\mathbf{w}}_t^1)$ and $f_t(\widehat{\mathbf{w}}_t^2)$ and let $v_{t,i_t} = \frac{1}{2\delta}(f_t(\widehat{\mathbf{w}}_t^1) - f(\widehat{\mathbf{w}}_t^2))$
  Compute

$$\mathbf{g}_t = d(v_{t,i_t} - \widehat{\mathbf{g}}_{t-1,i_t})\mathbf{e}_{i_t}, \text{ and}$$
$$\widehat{\mathbf{g}}_t = d(v_{t,i_t} - \widehat{\mathbf{g}}_{t-1,i_t})\mathbf{e}_{i_t} + \widehat{\mathbf{g}}_{t-1}$$

  Update

$$\mathbf{w}_{t+1} = \Pi_{(1-\xi)\Omega}[\mathbf{w}_t - \eta\mathbf{g}_t]$$
$$\widehat{\mathbf{w}}_{t+1} = \Pi_{(1-\xi)\Omega}[\mathbf{w}_{t+1} - \eta\widehat{\mathbf{g}}_t]$$

end for

The theorem below states the result.

Theorem 9. *Assume $\{f_1, \ldots, f_T\} \in \mathcal{V}_{p,g}$ and $f_t(\mathbf{w})$ is L-smooth for any $t \geq 1$. By the policy in Algorithm 1 with $\xi = \frac{1}{T}$, $\delta = \xi R_1$ and $\eta = \min\left(\sqrt{\frac{(2rB_T+r^2)}{8S_Td^4}}, \frac{1}{4Ld^{3/2}\sqrt{\ln T}}\right)$, we have*

$$\mathrm{E}\left[\sum_{t=1}^{T}\frac{1}{2}(f_t(\widehat{\mathbf{w}}_t^1) + f(\widehat{\mathbf{w}}_t^2))\right] - \sum_{t=1}^{T} f_t(\mathbf{w}_t^*) \leq O\left(\max\left\{d^2\sqrt{S_T\max(B_T, 1)}, d^{3/2}\max(B_T, 1)\right\}\right).$$

where $\widehat{\mathbf{w}}_t^1 = \mathbf{w}_t + \delta\mathbf{u}_t, \widehat{\mathbf{w}}_t^2 = \mathbf{w}_t - \delta\mathbf{u}_t$.

Remark: When the gradient variation is small such that the upper bound is dominated by $O(V_T^p)$, it matches the lower bound established in Proposition 1. Finally, we note that a similar upper bound can be achieved for linear loss functions by extending the static regret analysis in (Chiang et al., 2013) to the dynamic regret similarly to the proof of Theorem 9.

4. Conclusions

In this paper, we have considered dynamic regret for online learning under true and noisy gradient feedback. We have developed several lower and upper bounds of the dynamic regret based on the path variation that measures the temporal changes in the optimal solutions. In light of the presented lower bounds, the achieved upper bounds are optimal for non-strongly convex loss functions when the clairvoyant moves slowly. An interesting question that remains open is that what is the optimal dynamic regret bound for strongly convex loss functions in terms of the path variation.

Acknowledgements

The authors would like to thank the anonymous reviewers for their helpful comments. T. Yang was supported in part by NSF (1463988, 1545995).

References

Agarwal, Alekh, Dekel, Ofer, and Xiao, Lin. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Proceedings of the 23rd Conference on Learning Theory (COLT)*, pp. 28–40, 2010.

Besbes, Omar, Gur, Yonatan, and Zeevi, Assaf J. Non-stationary stochastic optimization. *CoRR*, abs/1307.5449, 2013.

Buchbinder, Niv, Chen, Shahar, Naor, Joseph, and Shamir, Ohad. Unified algorithms for online learning and competitive analysis. In *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*, pp. 5.1–5.18, 2012.

Cesa-Bianchi, Nicol, Gaillard, Pierre, Lugosi, Gbor, and Stoltz, Gilles. A new look at shifting regret. *CoRR*, abs/1202.3323, 2012.

Chiang, Chao-Kai, Yang, Tianbao, Lee1, Chia-Jung, Mahdavi, Mehrdad, Lu, Chi-Jen, Jin, Rong, and Zhu, Shenghuo. Online optimization with gradual variations. In *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*, 2012.

Chiang, Chao-Kai, Lee, Chia-Jung, and Lu, Chi-Jen. Beating bandits in gradually evolving worlds. In *Proceedings of the 26th Annual Conference on Learning Theory (COLT)*, pp. 210–227, 2013.

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

Hall, Eric C. and Willett, Rebecca M. Online optimization in dynamic environments. *CoRR*, abs/1307.5944, 2013.

Herbster, Mark and Warmuth, Manfred K. Tracking the best expert. *Machine Learning*, 32:151–178, 1998.

Jadbabaie, Ali, Rakhlin, Alexander, Shahrampour, Shahin, and Sridharan, Karthik. Online optimization : Competing with dynamic comparators. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2015.

Rakhlin, Alexander and Sridharan, Karthik. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems 26 (NIPS)*, pp. 3066–3074, 2013.

Yang, Tianbao, Mahdavi, Mehrdad, Jin, Rong, and Zhu, Shenghuo. Regret bounded by gradual variation for online convex optimization. *Machine Learning*, 95(2):183–223, 2014.

A. Proof of Lemma 4

Let $\mathbf{w}'_t = \mathbf{w}_t - \eta \mathbf{g}_t$. Thus $\mathbf{w}_{t+1} = \Pi_\Omega[\mathbf{w}'_t]$.

$$\frac{1}{2}\|\mathbf{w}_{t+1} - \mathbf{w}_t^*\|_2^2 \leq \frac{1}{2}\|\mathbf{w}'_t - \mathbf{w}_t^*\|_2^2 = \frac{1}{2}\|\mathbf{w}_t - \eta\mathbf{g}_t - \mathbf{w}_t^*\|_2^2$$
$$= \frac{1}{2}\|\mathbf{w}_t - \mathbf{w}_t^*\|_2^2 - \eta\mathbf{g}_t^\top(\mathbf{w}_t - \mathbf{w}_t^*) + \frac{1}{2}\eta^2\|\mathbf{g}_t\|_2^2$$

Then

$$\mathbf{g}_t^\top(\mathbf{w}_t - \mathbf{w}_t^*) \leq \frac{1}{2}\|\mathbf{w}_t - \mathbf{w}_t^*\|_2^2 - \frac{1}{2}\|\mathbf{w}_{t+1} - \mathbf{w}_t^*\|_2^2$$
$$+ \frac{1}{2}\eta^2\|\mathbf{g}_t\|_2^2$$
$$= \frac{1}{2\eta}\|\mathbf{w}_t - \mathbf{w}_t^*\|_2^2 - \frac{1}{2\eta}\|\mathbf{w}_{t+1} - \mathbf{w}_{t+1}^* + \mathbf{w}_{t+1}^* - \mathbf{w}_t^*\|_2^2$$
$$+ \frac{1}{2}\eta\|\mathbf{g}_t\|_2^2$$
$$= \frac{1}{2\eta}\|\mathbf{w}_t - \mathbf{w}_t^*\|_2^2 - \frac{1}{2\eta}\|\mathbf{w}_{t+1} - \mathbf{w}_{t+1}^*\|_2^2$$
$$- \frac{1}{2\eta}\|\mathbf{w}_{t+1}^* - \mathbf{w}_t^*\|_2^2 + \frac{1}{\eta}(\mathbf{w}_{t+1}^* - \mathbf{w}_{t+1})^\top(\mathbf{w}_t^* - \mathbf{w}_{t+1}^*)$$
$$+ \frac{1}{2}\eta\|\mathbf{g}_t\|_2^2$$
$$\leq \frac{1}{2\eta}\|\mathbf{w}_t - \mathbf{w}_t^*\|_2^2 - \frac{1}{2\eta}\|\mathbf{w}_{t+1} - \mathbf{w}_{t+1}^*\|_2^2$$
$$- \frac{1}{2\eta}\|\mathbf{w}_{t+1}^* - \mathbf{w}_t^*\|_2^2 + \frac{1}{\eta}r\|\mathbf{w}_t^* - \mathbf{w}_{t+1}^*\|_2 + \frac{1}{2}\eta\|\mathbf{g}_t\|_2^2$$

B. Proof of Theorem 8

Define $\hat{f}_t(\mathbf{w})$ as

$$\hat{f}_t(\mathbf{w}) = \mathrm{E}_{\mathbf{u}}[f_t(\mathbf{w} + \delta\mathbf{u})]$$

where $\mathbf{u}$ is a random unit vector. We first give the following lemma.

Lemma 10. *Let* $\widehat{\mathbf{w}}_t^* = (1-\xi)\mathbf{w}_t^*$.

$$\sum_{t=1}^T \frac{1}{2}(f_t(\widehat{\mathbf{w}}_t^1) + f_t(\widehat{\mathbf{w}}_t^2)) - \sum_{t=1}^T f_t(\mathbf{w}_t^*)$$
$$\leq \sum_{t=1}^T \hat{f}_t(\mathbf{w}_t) - \sum_{t=1}^T \hat{f}_t(\widehat{\mathbf{w}}_t^*) + 3TG\delta + TGR_2\xi$$

Note that the updating rule is OGD with a noisy gradient applied to a sequence of functions $\hat{f}_t(\mathbf{w}), t = 1, \ldots, T$. Following (Agarwal et al., 2010), $\hat{\mathbf{g}}_t$ is bounded by $\|\hat{\mathbf{g}}_t\|_2 \leq Gd$. Then following the proof of Theorem 6,

we have,

$$\sum_{t=1}^{T}\hat{f}_t(\mathbf{w}_t)-\sum_{t=1}^{T}\hat{f}_t(\widehat{\mathbf{w}}_t^*)\le\frac{\|\mathbf{w}_1-\widehat{\mathbf{w}}_1^*\|_2^2}{2\eta}$$
$$+\sum_{t=1}^{T}\frac{1}{\eta}\|\mathbf{w}_{t+1}-\widehat{\mathbf{w}}_{t+1}^*\|_2\|\widehat{\mathbf{w}}_t^*-\widehat{\mathbf{w}}_{t+1}^*\|_2+\frac{\eta}{2}G^2d^2T$$

Since $\mathbf{w}_t,\widehat{\mathbf{w}}_t^*\in(1-\xi)\Omega$, we have

$$\|\mathbf{w}_t-\widehat{\mathbf{w}}_t^*\|_2\le r$$

due to Assumption 1. In addition,

$$\|\widehat{\mathbf{w}}_t^*-\widehat{\mathbf{w}}_{t+1}^*\|_2=\|(1-\xi)\mathbf{w}_t^*-(1-\xi)\mathbf{w}_{t+1}^*\|_2$$
$$\le(1-\xi)\|\mathbf{w}_t^*-\mathbf{w}_{t+1}^*\|_2\le\|\mathbf{w}_t^*-\mathbf{w}_{t+1}^*\|_2$$

Then

$$\sum_{t=1}^{T}\hat{f}_t(\mathbf{w}_t)-\sum_{t=1}^{T}\hat{f}_t(\widehat{\mathbf{w}}_t^*)\le\frac{1}{2\eta}(r^2+2rV_T^p)+\frac{\eta}{2}G^2d^2T$$
$$\le\frac{1}{2\eta}(r^2+2rB_T)+\frac{\eta}{2}G^2d^2T$$

By plugging the value of η, we have

$$\sum_{t=1}^{T}\hat{f}_t(\mathbf{w}_t)-\sum_{t=1}^{T}\hat{f}_t(\widehat{\mathbf{w}}_t^*)\le\sqrt{(r^2+2rB_T)G^2d^2T}$$

Then combining the above inequality with Lemma 13, we have

$$\sum_{t=1}^{T}\frac{1}{2}(f_t(\widehat{\mathbf{w}}_t^1)+f_t(\widehat{\mathbf{w}}_t^2))-\sum_{t=1}^{T}f_t(\mathbf{w}_t^*)$$
$$\le\sqrt{(r^2+2rB_T)G^2d^2T}+3TG\delta+TGR_2\xi$$
$$\le\sqrt{(r^2+2rB_T)G^2d^2T}+G(3R_1+R_2).$$

B.1. Proof of Lemma ??

The proof is almost identical to that of Lemma 2 in (Agarwal et al., 2010). By the Lipschitz property of $f_t(\mathbf{w})$, we have

$$f_t(\widehat{\mathbf{w}}_t^1)=f_t(\mathbf{w}_t+\delta\mathbf{u}_t)\le f_t(\mathbf{w}_t)+G\delta\|\mathbf{u}_t\|_2$$
$$f_t(\widehat{\mathbf{w}}_t^2)=f_t(\mathbf{w}_t-\delta\mathbf{u}_t)\le f_t(\mathbf{w}_t)+G\delta\|\mathbf{u}_t\|_2$$

Since $\|\mathbf{u}_t\|_2=1$, thus

$$\frac{1}{2}(f_t(\widehat{\mathbf{w}}_t^1)+f_t(\widehat{\mathbf{w}}_t^2))\le f_t(\mathbf{w}_t)+G\delta$$

By the Lipschitz property and $\Omega\subset R_2\mathbf{B}$, we have for any $\mathbf{w}\in\Omega$

$$f_t((1-\xi)\mathbf{w})\le f_t(\mathbf{w})+GR_2\xi$$

Further for any $\mathbf{w}\in(1-\xi)\Omega$,

$$|f_t(\mathbf{w})-\hat{f}_t(\mathbf{w})|=|f_t(\mathbf{w})-\mathrm{E}_tf_t(\mathbf{w}+\delta\mathbf{u})|$$
$$\le\mathrm{E}_t|f_t(\mathbf{w})-f_t(\mathbf{w}+\delta\mathbf{u})|\le G\delta$$

Then

$$f_t(\mathbf{w}_t)\le\hat{f}_t(\mathbf{w}_t)+G\delta,\quad\text{and }\hat{f}_t((1-\xi)\mathbf{w}_t^*)\le f_t((1-\xi)\mathbf{w}_t^*)+G\delta$$

Combining the above inequalities, we get

$$\frac{1}{2}(f_t(\widehat{\mathbf{w}}_t^1)+f_t(\widehat{\mathbf{w}}_t^2))+\hat{f}_t((1-\xi)\mathbf{w}_t^*)$$
$$\le f_t(\mathbf{w}_t)+G\delta+f_t((1-\xi)\mathbf{w}_t^*)+G\delta$$
$$\le f_t(\mathbf{w}_t)+f_t(\mathbf{w}_t^*)+GR_2\xi+2G\delta$$
$$\le\hat{f}_t(\mathbf{w}_t)+f_t(\mathbf{w}_t^*)+GR_2\xi+3G\delta$$

As a result,

$$\sum_{t=1}^{T}\frac{1}{2}(f_t(\widehat{\mathbf{w}}_t^1)+f_t(\widehat{\mathbf{w}}_t^2))-\sum_{t=1}^{T}f_t(\mathbf{w}_t^*)$$
$$\le\sum_{t=1}^{T}\hat{f}_t(\mathbf{w}_t)-\sum_{t=1}^{T}\hat{f}_t((1-\xi)\mathbf{w}_t^*)+3TG\delta+TGR_2\xi$$

C. Proof of Theorem 9

The proof follows similarly to the analysis in (Chiang et al., 2013). We present a series of Lemmas with some of the Lemmas' proof omitted due to that they are identical to that in (Chiang et al., 2013). To simply the presentation, we denote by $O(1)$ any constant independent of T.

Lemma 11.

$$\sum_{i=1}^{T}\frac{1}{2}(f_t(\widehat{\mathbf{w}}_t^1)+f_t(\widehat{\mathbf{w}}_t^2))-\sum_{t=1}^{T}f_t(\mathbf{w}_t^*)$$
$$\le\sum_{t=1}^{T}f_t(\mathbf{w}_t)-\sum_{t=1}^{T}f_t((1-\xi)\mathbf{w}_t^*)+O(1)$$

The proof of this lemma is presented later.

Lemma 12. *Let* $\widehat{\mathbf{w}}_t^*=(1-\xi)\mathbf{w}_t^*$.

$$\sum_{t=1}^{T}f_t(\mathbf{w}_t)-\sum_{t=1}^{T}f_t(\widehat{\mathbf{w}}_t^*)\le\sum_{t=1}^{T}\nabla f_t(\mathbf{w}_t)^\top(\mathbf{w}_t-\widehat{\mathbf{w}}_t^*)$$

The lemma above follows the convexity of $f_t(\mathbf{w})$.

Lemma 13. *Let* $\widehat{\mathbf{w}}_t^*=(1-\xi)\mathbf{w}_t^*$.

$$\mathrm{E}\left[\sum_{t=1}^{T}\nabla f_t(\mathbf{w}_t)^\top(\mathbf{w}_t-\widehat{\mathbf{w}}_t^*)\right]$$
$$\le\mathrm{E}\left[\sum_{t=1}^{T}\mathbf{g}_t^\top(\mathbf{w}_t-\widehat{\mathbf{w}}_t^*)\right]+O(1)$$

The proof of the above lemma follows the same to the proof of Lemma 5 in (Chiang et al., 2013).

Lemma 14. *Define*

$$S_t = \eta_t \|\mathbf{g}_t - \mathbf{g}_{t-1}\|_2^2$$
$$A_t = \frac{1}{2\eta}\|\mathbf{w}_t - \widehat{\mathbf{w}}_t^*\|_2^2 - \frac{1}{2\eta}\|\mathbf{w}_{t+1} - \widehat{\mathbf{w}}_t^*\|_2^2$$
$$C_t = \frac{1}{2}\|\mathbf{w}_{t+1} - \widehat{\mathbf{w}}_t\|_2^2 + \frac{1}{2\eta}\|\mathbf{w}_t - \widehat{\mathbf{w}}_t\|_2^2$$

Then

$$\sum_{t=1}^T \mathbf{g}_t^\top(\mathbf{w}_t - \widehat{\mathbf{w}}_t^*) \leq \sum_{t=1}^T S_t + \sum_{t=1}^T A_t - \sum_{t=1}^T C_t$$

The above lemma is a result of the Lemma 4 in (Chiang et al., 2013). Combining the above lemmas, we have

Theorem 15.

$$\mathrm{E}\left[\sum_{i=1}^T \frac{1}{2}(f_t(\widehat{\mathbf{w}}_t^1) + f_t(\widehat{\mathbf{w}}_t^2)) - \sum_{t=1}^T f_t(\mathbf{w}_t^*)\right] \leq \mathrm{E}\left[\sum_{t=1}^T S_t + \sum_{t=1}^T A_t - \sum_{t=1}^T C_t\right] + O(1)$$

To proceed we bound the three summation terms in the R.H.S..

Lemma 16.

$$\sum_{t=1}^T C_t \geq \frac{1}{4\eta}\mathrm{E}\left[\sum_{t=1}^T \|\widehat{\mathbf{w}}_t - \widehat{\mathbf{w}}_{t+1}\|_2^2\right] - O(1)$$

This is the same to the Lemma 11 in (Chiang et al., 2013).

Lemma 17.

$$\sum_{t=1}^T A_t \leq \frac{\|\mathbf{w}_1 - \widehat{\mathbf{w}}_1^*\|_2^2}{2\eta} + \frac{1}{\eta}\sum_{t=1}^T \|\mathbf{w}_{t+1} - \widehat{\mathbf{w}}_{t+1}^*\|_2 \|\widehat{\mathbf{w}}_t^* - \widehat{\mathbf{w}}_{t+1}^*\|_2 \leq \frac{1}{2\eta}(r^2 + 2rV_T^p)$$

The proof follows similarly to that of Lemma 1.

Lemma 18.

$$\sum_{t=1}^T S_t \leq 4\eta d^4 V_T^g + 4\eta d^3 L^2 \ln T \mathrm{E}\left[\sum_{t=1}^T \|\widehat{\mathbf{w}}_t - \widehat{\mathbf{w}}_{t+1}\|_2^2\right] + O(1)$$

where L is the smoothness parameter.

The lemma follows the Lemma 12 in (Chiang et al., 2013). Combining Lemma **??**, **??** and Lemma **??**, we have

$$\mathrm{E}\left[\sum_{t=1}^T S_t + \sum_{t=1}^T A_t - \sum_{t=1}^T C_t\right] \leq 4\eta d^4 V_T^g + \frac{1}{2\eta}(r^2 + 2rV_T^p) + 4\eta d^3 L^2 \ln T \mathrm{E}\left[\sum_{t=1}^T \|\widehat{\mathbf{w}}_t - \widehat{\mathbf{w}}_{t+1}\|_2^2\right] - \frac{1}{4\eta}\mathrm{E}\left[\sum_{t=1}^T \|\widehat{\mathbf{w}}_t - \widehat{\mathbf{w}}_{t+1}\|_2^2\right] + O(1)$$

Since $\eta \leq \frac{1}{4d^{3/2}L\sqrt{\ln T}}$, then

$$\begin{aligned}&\mathrm{E}\left[\sum_{t=1}^T S_t + \sum_{t=1}^T A_t - \sum_{t=1}^T C_t\right] \\ &\leq 4\eta d^4 V_T^g + \frac{1}{2\eta}(r^2 + 2rV_T^p) + O(1) \\ &\leq 4\eta d^4 S_T + \frac{1}{2\eta}(r^2 + 2rB_T) + O(1)\end{aligned}$$

Then by the value of η, we have

$$\begin{aligned}&\mathrm{E}\left[\sum_{t=1}^T S_t + \sum_{t=1}^T A_t - \sum_{t=1}^T C_t\right] \\ &\leq 4\eta d^4 S_T + \frac{1}{2\eta}(r^2 + 2rB_T) + O(1) \\ &\leq \max\left\{2\sqrt{2}\sqrt{d^4 S_T(r^2 + 2rB_T)}, 4d^{3/2}L\sqrt{\ln T}(r^2 + 2rB_T)\right\} \\ &+ O(1)\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathrm{E}\left[\sum_{i=1}^T \frac{1}{2}(f_t(\widehat{\mathbf{w}}_t^1) + f_t(\widehat{\mathbf{w}}_1^t)) - \sum_{t=1}^T f_t(\mathbf{w}_t^*)\right] \\ &\leq O\left(\max(d^2\sqrt{S_T \max(1, B_T)}, d^{3/2}\max(1, B_T))\right)\end{aligned}$$

C.1. Proof of Lemma ??

From the Lipschitz property,

$$f_t(\widehat{\mathbf{w}}_t^1) - f_t(\mathbf{w}_t) \leq G\|\widehat{\mathbf{w}}_t^1 - \mathbf{w}_t\|_2 \leq G\delta$$
$$f_t(\widehat{\mathbf{w}}_t^2) - f_t(\mathbf{w}_t) \leq G\|\widehat{\mathbf{w}}_t^2 - \mathbf{w}_t\|_2 \leq G\delta$$

Then

$$\sum_{i=1}^T \frac{1}{2}(f_t(\widehat{\mathbf{w}}_t^1) + f_t(\widehat{\mathbf{w}}_1^t)) - \sum_{t=1}^T f_t(\mathbf{w}_t) \leq G\delta T$$

To proceed,

$$\begin{aligned}&f_t((1-\xi)\mathbf{w}_t^*) \leq \xi f_t(0) + (1-\xi)f_t(\mathbf{w}_t^*) \\ &= f_t(\mathbf{w}_t^*) + \xi(f_t(0) - f_t(\mathbf{w}_t^*)) \leq f_t(\mathbf{w}_t^*) + \xi G\|\mathbf{w}_t^*\|_2 \\ &= f_t(\mathbf{w}_t^*) + \xi G R_2\end{aligned}$$

Then

$$\sum_{t=1}^{T} f_t((1-\xi)\mathbf{w}_t^*) - \sum_{t=1}^{T} f_t(\mathbf{w}_t^*) \leq \xi G R_2 T$$

Thus

$$\begin{aligned}
&\sum_{i=1}^{T} \frac{1}{2}(f_t(\widehat{\mathbf{w}}_t^1) + f_t(\widehat{\mathbf{w}}_1^t)) - \sum_{t=1}^{T} f_t(\mathbf{w}_t^*) \\
&\leq \sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t((1-\xi)\mathbf{w}_t^*) + G\delta T + \xi G R_2 T \\
&\leq \sum_{t=1}^{T} f_t(\mathbf{w}_t) - \sum_{t=1}^{T} f_t((1-\xi)\mathbf{w}_t^*) + O(1)
\end{aligned}$$