Adaptivity and Non-stationarity: Problem-dependent Dynamic Regret for Online Convex Optimization

Peng Zhao ZHAOP@LAMDA.NJU.EDU.CN
Yu-Jie Zhang ZHANGYJ@LAMDA.NJU.EDU.CN
Lijun Zhang ZHANGLJ@LAMDA.NJU.EDU.CN
Zhi-Hua Zhou ZHOUZH@LAMDA.NJU.EDU.CN
National Key Laboratory for Novel Software Technology
Nanjing University, Nanjing 210023, China

Abstract

We investigate online convex optimization in non-stationary environments and choose the *dynamic regret* as the performance measure, defined as the difference between cumulative loss incurred by the online algorithm and that of any feasible comparator sequence. Let T be the time horizon and P_T be the path length that essentially reflects the non-stationarity of environments, the state-of-the-art dynamic regret is $\mathcal{O}(\sqrt{T(1+P_T)})$. Although this bound is proved to be minimax optimal for convex functions, in this paper, we demonstrate that it is possible to further enhance the guarantee for some easy problem instances, particularly when online functions are smooth. Specifically, we introduce novel online algorithms that can exploit smoothness and replace the dependence on T in dynamic regret with *problem-dependent* quantities: the variation in gradients of loss functions, the cumulative loss of the comparator sequence, and the minimum of these two terms. These quantities are at most $\mathcal{O}(T)$ while could be much smaller in benign environments. Therefore, our results are adaptive to the intrinsic difficulty of the problem, since the bounds are tighter than existing results for easy problems and meanwhile guarantee the same rate in the worst case. Notably, our proposed algorithms can achieve favorable dynamic regret with only *one* gradient per iteration, sharing the same gradient query complexity as the static regret minimization methods. To accomplish this, we introduce the framework of *collaborative online ensemble*. The proposed framework employs a two-layer online ensemble to handle non-stationarity, and uses optimistic online learning and further introduces crucial correction terms to enable effective collaboration within the meta-base two layers, thereby attaining adaptivity. We believe the framework can be useful for broader problems.

Keywords: Online Learning, Online Convex Optimization, Dynamic Regret, Problem-dependent Bounds, Gradient Variation, Optimistic Mirror Descent, Online Ensemble

1. Introduction

In many real-world applications, data are inherently accumulated over time, and thus it is of great importance to develop a learning system that updates in an online fashion. Online Convex Optimization (OCO) is a powerful paradigm for learning in such scenarios, which can be regarded as an iterative game between a player and an adversary. At iteration t, the player chooses a decision vector $\mathbf{x}_t$ from a convex set $\mathcal{X} \subseteq \mathbb{R}^d$. Subsequently, the adversary discloses a convex function $f_t : \mathcal{X} \mapsto \mathbb{R}$, and the player incurs a loss denoted by $f_t(\mathbf{x}_t)$. The standard performance measure is the (static) *regret* (Zinkevich, 2003),

$$\text{S-Regret}_T = \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \min_{\mathbf{x}\in\mathcal{X}} \sum_{t=1}^{T} f_t(\mathbf{x}), \tag{1}$$

which is the difference between cumulative loss incurred by the online algorithm and that of the best decision in hindsight. The rationale behind such a metric is that the best fixed decision in hindsight is reasonably good over all the iterations. However, this might be too optimistic and may not hold in changing environments, where data are evolving and the optimal decision is drifting over time. To address this limitation, *dynamic regret* is proposed to compete with changing comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$,

$$\text{D-Regret}_T(\mathbf{u}_1, \ldots, \mathbf{u}_T) = \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t), \tag{2}$$

which draws considerable attention recently (Zhang et al., 2018a; Zhao et al., 2020b; Cutkosky, 2020; Zhao et al., 2021a; Baby and Wang, 2021). The measure is also called the *universal* dynamic regret (or *general* dynamic regret), in the sense that it gives a universal guarantee that holds against *any* comparator sequence. Note that the static regret (1) can be viewed as its special form by choosing comparators as the fixed best decision in hindsight. Moreover, a variant appeared frequently in the literature is called the *worst-case dynamic regret* (Besbes et al., 2015; Jadbabaie et al., 2015; Mokhtari et al., 2016; Yang et al., 2016; Wei et al., 2016; Zhang et al., 2017; Baby and Wang, 2019; Yuan and Lamperski, 2020; Zhao et al., 2020a; Zhang et al., 2020a,b; Zhao and Zhang, 2021), defined as

$$\text{D-Regret}_T(\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*) = \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{x}_t^*), \tag{3}$$

which specializes the general form (2) with comparators $\mathbf{u}_t = \mathbf{x}_t^* \in \arg\min_{\mathbf{x}\in\mathcal{X}} f_t(\mathbf{x})$. Therefore, universal dynamic regret is very general and can include the static regret (1) and the worst-case dynamic regret (3) as special cases by different instantiations of comparators. We further remark that the worst-case dynamic regret is often too pessimistic, whereas the universal one is more adaptive to non-stationary environments. Actually, the changes of online functions usually come from two aspects: one is the sampling randomness and the other one is the environmental non-stationarity, and clearly the latter one is the main focus of non-stationary online learning. Optimizing the worst-case dynamic regret can be problematic in some cases. For example, considering the stochastic optimization task where f_t's are independently randomly sampled from the same distribution, then minimizing the worst-case dynamic regret is evidently inappropriate and will eventually lead to overfitting (Zhang et al., 2018a) because the minimizer of online loss function could be dramatically different from the minimizer of the expected loss function due to the sampling randomness.

There are many studies on the worst-case dynamic regret (Besbes et al., 2015; Jadbabaie et al., 2015; Mokhtari et al., 2016; Yang et al., 2016; Zhang et al., 2017, 2018b; Baby and Wang, 2019; Zhang et al., 2020b; Zhao and Zhang, 2021), but only few results are known for the universal dynamic regret. Zinkevich (2003) shows that online gradient descent (OGD) with a step size $\eta > 0$ achieves an $\mathcal{O}((1 + P_T)/\eta + \eta T)$ universal dynamic regret, where $P_T =$

$\sum_{t=2}^{T}\|\mathbf{u}_{t-1}-\mathbf{u}_t\|_2$ is the path length of comparators $\mathbf{u}_1,\ldots,\mathbf{u}_T$ and thus reflects the non-stationarity of the environments. When the path length P_T was known, one could choose the optimal step size $\eta_* = \Theta(\sqrt{(1+P_T)/T})$ and attain an $\mathcal{O}(\sqrt{T(1+P_T)})$ dynamic regret. However, this path length quantity is hard to know since the universal dynamic regret aims to provide guarantees against any feasible comparator sequence. The step size $\eta = \Theta(1/\sqrt{T})$ commonly used in static regret would lead to an inferior $\mathcal{O}(\sqrt{T}(1+P_T))$ bound, which exhibits a large gap from the favorable bound with an oracle step size tuning. Zhang et al. (2018a) resolve the issue by proposing a novel online algorithm to search the optimal step size η_*, attaining an $\mathcal{O}(\sqrt{T(1+P_T)})$ universal dynamic regret, and they also establish an $\Omega(\sqrt{T(1+P_T)})$ lower bound to show the minimax optimality.

Although the rate is minimax optimal for convex functions, we would like to design algorithms with more adaptive bounds. Specifically, we aim to enhance the guarantee for some easy problem instances, particularly when the online functions are smooth, by replacing the dependence on T by certain *problem-dependent* quantities that are $\mathcal{O}(T)$ in the worst case while could be much smaller in benign environments. In the study of static regret, we can attain such results like small-loss bounds (Srebro et al., 2010) and gradient-variation bounds (Chiang et al., 2012). Thus, a natural question arises *whether it is possible to achieve similar problem-dependent guarantees for the universal dynamic regret?*

Our results. In this paper, extending our preliminary work (Zhao et al., 2020b), we provide an affirmative answer by designing online algorithms with problem-dependent dynamic regret bounds. Specifically, we focus on the following two adaptive quantities: the gradient variation of online functions V_T, and the cumulative loss of the comparator sequence F_T, defined as

$$V_T = \sum_{t=2}^{T} \sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2, \text{ and } \quad F_T = \sum_{t=1}^{T} f_t(\mathbf{u}_t). \tag{4}$$

The two problem-dependent quantities are both at most $\mathcal{O}(T)$ under standard assumptions of online learning, while could be much smaller in easier problem instances. We propose two novel online algorithms called Sword and Sword++ ("Sword" is short for $\underline{\text{S}}$moothness-a$\underline{\text{w}}$are $\underline{\text{o}}$nline lea$\underline{\text{r}}$ning with $\underline{\text{d}}$ynamic regret) that are suitable for different feedback models. Our algorithms are online ensemble methods (Zhou, 2012; Zhao, 2021), which admit a two-layer structure with a meta-algorithm running over a group of base-learners. We prove that they enjoy an $\mathcal{O}(\sqrt{(1+P_T+\min\{V_T,F_T\})(1+P_T)})$ dynamic regret, achieving gradient-variation and small-loss bounds simultaneously. Comparing to the $\mathcal{O}(\sqrt{T(1+P_T)})$ minimax rate, our result replaces the dependence on T by the problem-dependent quantity $P_T + \min\{V_T, F_T\}$. Our bounds become much tighter when the problem is easy (for example when P_T and V_T/F_T are sublinear in T), and meanwhile safeguard the same guarantee in the worst case. Hence, our results are adaptive to the intrinsic difficulty of problem instances as well as the non-stationarity of environments.

Our first algorithm, Sword, achieves the favorable problem-dependent guarantees under the multi-gradient feedback model, namely, the player can query gradient information multiple times at each round. This algorithm is conceptually simple, but the gradient query complexity is $\mathcal{O}(\log T)$ at each round. Our second algorithm, Sword++, is an improved version that requires only *one* gradient per iteration, despite using a two-layer online ensemble structure. As a result, Sword++ is not only computationally efficient but also more attractive due to its reduced feedback requirements — Sword++ can be applied to the more challenging one-gradient feedback model, where the player only receives gradient $\nabla f_t(\mathbf{x}_t)$ as the feedback after submitting the decision $\mathbf{x}_t$. Furthermore, it is worth mentioning that Sword++ has the potential to be extended to more constrained feedback models, such as the two-point bandit convex optimization, by further leveraging the technique for gradient-variation static regret minimization presented in (Chiang et al., 2013).

Technical contributions. Note that there exist studies showing that the worst-case dynamic regret can benefit from smoothness (Yang et al., 2016; Zhang et al., 2017; Zhao and Zhang, 2021). However, their analyses do not apply to our universal dynamic regret, as we cannot exploit the optimality condition of comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T$, in stark contrast with the worst-case dynamic regret analysis. As a result, we propose an adaptive online ensemble method to hedge non-stationarity while extracting adaptivity. Specifically, we employ the meta-base two-layer ensemble to hedge the non-stationarity and utilize optimistic online learning to reuse the historical gradient information adaptively. Two crucial novel ingredients are designed in order to achieve favorable problem-dependent guarantees.

- We introduce optimistic mirror descent (OMD) as a unified building block for the algorithm design of dynamic regret minimization in both meta and base levels. We present generic and completely modular analysis for the dynamic regret of OMD, where the *negative term* is essential for achieving problem-dependent dynamic regret.

- We propose the *collaborative online ensemble* framework. In addition to employing optimistic online learning to attain adaptivity and meta-base structure to hedge non-stationarity, we incorporate a novel decision-deviation *correction term*, which facilitates effective collaborations within the two layers and is crucial for achieving the desired problem-dependent bound while requiring only one gradient per iteration.

We emphasize that these ingredients are particularly important for achieving gradient-variation dynamic regret, which we will demonstrate to be more fundamental than the small-loss bound. In particular, our overall solution (especially Sword++) effectively utilizes negative terms and introduces correction terms to ensure successful collaboration within the two-layer online ensemble. The overall framework of collaborative online ensemble is summarized in Section 5, and we believe that the proposed framework has the potential for broader online learning problems.

Organization. The rest is structured as follows. Section 2 briefly reviews the related work. In Section 3, we introduce the problem setup and algorithmic framework, where a generic dynamic regret analysis of optimistic mirror descent is provided. Section 4 presents our main results, in which the gradient-variation dynamic regret bounds are established. Section 5 illustrates a generic framework called collaborative online ensemble that is highly useful for attaining problem-dependent dynamic regret. Section 6 provides some additional results. The major proofs are presented in Section 7. Furthermore, Section 8 reports the experiments to empirically support our theoretical findings. Finally, we conclude the paper in Section 9. Some omitted details and proofs are provided in the appendix.

2. Related Work

In this section, we present a brief review of static regret and dynamic regret minimization for online convex optimization.

2.1 Static Regret

Static regret has been extensively studied in online convex optimization. Let T be the time horizon and d be the dimension, there exist online algorithms with static regret bounded by $\mathcal{O}(\sqrt{T})$, $\mathcal{O}(d \log T)$, and $\mathcal{O}(\log T)$ for convex, exponentially concave, and strongly convex functions, respectively (Zinkevich, 2003; Hazan et al., 2007). These results are proved to be minimax optimal (Abernethy et al., 2008). More results can be found in the seminal books (Shalev-Shwartz, 2012; Hazan, 2016) and references therein.

In addition to exploiting the convexity of functions, there are studies improving static regret by incorporating smoothness, whose main proposal is to replace the dependence on T by problem-dependent quantities. Such problem-dependent bounds enjoy many benign properties, in particular, they can safeguard the worst-case minimax rate yet can be much tighter in easier problem instances. There are usually two kinds of such bounds — small-loss bounds (Srebro et al., 2010) and gradient-variation bounds (Chiang et al., 2012).

Small-loss bounds are first introduced in the context of prediction with expert advice (Littlestone and Warmuth, 1994; Freund and Schapire, 1997), which replace the dependence on T by cumulative loss of the best expert. Later, Srebro et al. (2010) show that in the online convex optimization setting, OGD with a certain step size scheme can achieve an $\mathcal{O}(\sqrt{1+F_T^*})$ small-loss regret bound when the online convex functions are smooth and non-negative, where F_T^* is the cumulative loss of the best decision in hindsight, namely, $F_T^* = \sum_{t=1}^T f_t(\mathbf{x}^*)$ with $\mathbf{x}^*$ chosen as the offline minimizer. The key technical ingredient in the analysis is to exploit the self-bounding properties of smooth functions. Gradient-variation bounds are introduced by Chiang et al. (2012), rooting in the development of second-order bounds for prediction with expert advice (Cesa-Bianchi et al., 2005) and online convex optimization (Hazan and Kale, 2008). For convex and smooth functions, Chiang et al. (2012) establish an $\mathcal{O}(\sqrt{1+V_T})$ gradient-variation regret bound, where $V_T = \sum_{t=2}^T \sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$ measures the cumulative gradient variation. Gradient-variation bounds are particularly favored in slowly changing environments where online functions evolve gradually.

In addition, problem-dependent static regret bounds are also studied in the bandit online learning setting, including the gradient-variation bounds for two-point bandit convex optimization (Chiang et al., 2013), as well as small-loss bounds for multi-armed bandits (Allenberg et al., 2006; Wei and Luo, 2018; Lee et al., 2020b), linear bandits (Lee et al., 2020b), semi-bandits (Neu, 2015), graph bandits (Lykouris et al., 2018; Lee et al., 2020a), and contextual bandits (Allen-Zhu et al., 2018; Foster and Krishnamurthy, 2021), etc.

2.2 Dynamic Regret

Dynamic regret enforces the player to compete with time-varying comparators and thus is favored in online learning in open and non-stationary environments (Sugiyama and Kawanabe, 2012; Zhao et al., 2021b; Zhou, 2022). The notion of dynamic regret is sometimes referred to as tracking regret or shifting regret in the prediction with expert advice setting (Herbster and Warmuth, 1998, 2001; Bousquet and Warmuth, 2002; Cesa-Bianchi et al., 2012; György and Szepesvári, 2016). It is known that in the worst case, a sublinear dynamic regret is not attainable unless imposing certain regularities on the comparator sequence or the function sequence (Besbes et al., 2015; Jadbabaie et al., 2015). This paper focuses the most common regularity called the *path length* introduced by Zinkevich (2003), which measures fluctuation of the comparators defined by

$$P_T = \sum_{t=2}^{T} \|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2. \tag{5}$$

Note that we simply focus on the Euclidean norm throughout this paper, and it is straightforward to extend the notions and results to general primal-dual norms. Other regularities include the squared path length introduced by Zhang et al. (2017), which is defined as $S_T = \sum_{t=2}^{T} \|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2^2$, and the function variation introduced by Besbes et al. (2015) that measures the cumulative variation with respect to the function value and is defined as $V_T^f = \sum_{t=2}^{T} \sup_{\mathbf{x}\in\mathcal{X}} |f_{t-1}(\mathbf{x}) - f_t(\mathbf{x})|$.

There are two kinds of dynamic regret notions in the previous studies. The universal dynamic regret, as defined in (2), aims to compare with any feasible comparator sequence, while the worst-case dynamic regret defined in (3) specifies the comparator sequence to be the sequence of minimizers of online functions. In the following, we present the related works respectively. Notice that we will use notations P_T and S_T for path length and squared path length of the comparator sequence $\{\mathbf{u}_t\}_{t=1,\ldots,T}$, while adopt the notations P_T^* and S_T^* for that of the sequence $\{\mathbf{x}_t^*\}_{t=1,\ldots,T}$ where $\mathbf{x}_t^*$ is the minimizer of the online function f_t, namely, $P_T^* = \sum_{t=2}^{T} \|\mathbf{x}_{t-1}^* - \mathbf{x}_t^*\|_2$ and $S_T^* = \sum_{t=2}^{T} \|\mathbf{x}_{t-1}^* - \mathbf{x}_t^*\|_2^2$.

Universal dynamic regret. The seminal work of Zinkevich (2003) demonstrates that online gradient descent (OGD) enjoys an $\mathcal{O}(\sqrt{T}(1+P_T))$ universal dynamic regret, which holds against any feasible comparator sequence. Nevertheless, the result is far from the $\Omega(\sqrt{T(1+P_T)})$ lower bound established by Zhang et al. (2018a), who further close the gap by proposing a novel online algorithm that attains an optimal rate of $\mathcal{O}(\sqrt{T(1+P_T)})$ for convex functions (Zhang et al., 2018a). Our work further exploits the easiness of the problem instances and achieve problem-dependent regret guarantees, hence better than the minimax rate. Zhao et al. (2021a) study the universal dynamic regret for bandit convex optimization under both one-point and two-point feedback models. The universal dynamic regret is also studied for variants of the standard OCO model such as OCO with memory (Zhao et al., 2022) and OCO with switching cost (Zhang et al., 2021). We note that the aforementioned works and ours are all building on the two-layer meta-base structure. Concurrent to our conference version paper (Zhao et al., 2020b), Cutkosky (2020) proposes a novel online algorithm that achieves the same minimax optimal dynamic regret for convex functions as (Zhang et al., 2018a), without relying on meta-base aggregation. Instead, their method employs the combination strategy developed in parameter-free online learning (Cutkosky and Orabona, 2018; Cutkosky, 2019). We note that it may be possible to modify the algorithm of Cutkosky (2020) to achieve small-loss bounds; however, it is generally challenging to attain gradient-variation bounds, especially under the one-gradient feedback model. More specifically, it is not hard to modify their framework to be compatible with optimistic online learning, but one usually needs to exploits additional negative terms to convert the optimistic quantity $\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2$ to the gradient variation $\sup_{\mathbf{x}\in\mathcal{X}}\|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$, in order to eliminate the difference between decisions $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$. Our proposed algorithms are built upon the meta-base two-layer framework, which involves a careful exploitation of negative terms in the regret analysis of both meta and base algorithms, as well as the introduction of additional correction terms. However, as far as we can see, with only one gradient feedback per round, it is challenging for the framework of Cutkosky (2020) to achieve the gradient-variation bound due to the lack of negative terms in the regret analysis.

Worst-case dynamic regret. There are many efforts devoted to studying the worst-case dynamic regret. Yang et al. (2016) prove that OGD enjoys an $\mathcal{O}(\sqrt{T(1+P_T^*)})$ worst-case dynamic regret for convex functions when the path length P_T^* is known. For strongly convex and smooth functions, Mokhtari et al. (2016) show that an $\mathcal{O}(P_T^*)$ dynamic regret is achievable, and Zhang et al. (2017) further propose the online multiple gradient descent algorithm with an $\mathcal{O}(\min\{P_T^*, S_T^*\})$ guarantee. Yang et al. (2016) show that $\mathcal{O}(P_T^*)$ rate is attainable for convex and smooth functions, provided that all the minimizers $\mathbf{x}_t^*$'s lie in the interior of the domain $\mathcal{X}$. The above results mainly use the (squared) path length as the non-stationarity measure, which measures the cumulative variation of the comparator sequence. In another line of research, researchers use the variation with respect to the function values as the measure. Besbes et al. (2015) show that OGD with a restarting strategy attains an $\mathcal{O}(T^{2/3}V_T^{f\,1/3})$ regret for convex functions when the function variation V_T^f is available, which is improved to $\mathcal{O}(T^{1/3}V_T^{f\,2/3})$ for 1-dim square loss (Baby and Wang, 2019). Chen et al. (2019) extend the results of Besbes et al. (2015) to more general function-variation measures capable of capturing local temporal and spatial changes. To take advantage of variations in both comparator sequences and function values, (Zhao and Zhang, 2021) provide an improved analysis for online multiple gradient descent and prove an $\mathcal{O}(\min P_T^*, S_T^*, V_T^f)$ worst-case dynamic regret for strongly convex and smooth functions. For convex and smooth functions, they also demonstrate that the simple greedy strategy (i.e., $\mathbf{x}_{t+1} = \mathbf{x}_t^* \in \arg\min_{\mathbf{x}\in\mathcal{X}} f_t(\mathbf{x})$) can effectively optimize the worst-case dynamic regret (Zhao and Zhang, 2021, Section 4.2).

3. Problem Setup and Algorithmic Framework

In this section, we first formally state the problem setup, then introduce the foundational algorithmic framework for dynamic regret minimization, and finally list several assumptions that might be used in the theoretical analysis.

3.1 Problem Setup

Online Convex Optimization (OCO) can be modeled as an iterated game between the player and the environments. At iteration $t \in [T]$, the player first chooses the decision $\mathbf{x}_t$ from a convex feasible set $\mathcal{X} \subseteq \mathbb{R}^d$, then the environments reveal the loss function $f_t : \mathcal{X} \mapsto \mathbb{R}$ and the player suffers the loss $f_t(\mathbf{x}_t)$ and observes a certain information about the function $f_t(\cdot)$. According to the revealed information, the online learning problems are typically classified into *full-information* online learning and *partial-information* online learning (or called *bandit* online learning). In this paper, we focus on the full-information one, which can be further categorized into the following two setups:

(i) **multi-gradient feedback**: the player can access the entire gradient function $\nabla f_t(\cdot)$ and thus can evaluate the gradient multiple times;

(ii) **one-gradient feedback**: the player can observe the gradient information $\nabla f_t(\mathbf{x}_t)$ after submitting the decision $\mathbf{x}_t$.

In Section 4.2, we develop an online algorithm called Sword with gradient-variation dynamic regret bounds under the multi-gradient feedback model. In Section 4.3, we present an improved algorithm called Sword++ that can achieve the same guarantee under the more challenging one-gradient feedback model.

The typical performance measure is the static regret, which benchmarks the algorithm with a fixed comparator. To handle non-stationary environments, we focus on the strengthened measure called *dynamic regret*, which compares the online algorithm to a sequence of time-varying comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$, as defined in (2). An upper bound of dynamic regret should be a function of comparators, and typically the bound depends on some regularities that measure the fluctuation of the comparator sequence, such as the path length $P_T = \sum_{t=2}^{T} \|\mathbf{u}_t - \mathbf{u}_{t-1}\|_2$. Throughout the paper, we focus on the Euclidean norm for simplicity, and it is straightforward to extend our results to general primal-dual norms.

In addition to the regret measure, we further consider the *gradient query complexity*. Note that algorithms designed for the multi-gradient feedback model may query the gradients for multiple times at each round. However, most algorithms designed for the static regret minimization only require *one gradient per iteration*, namely, using $\nabla f_t(\mathbf{x}_t)$ for the next update only. Therefore, it is more desirable to achieve the favorable regret guarantees under the one-gradient feedback model. In other words, our aim is to develop first-order methods for dynamic regret minimization that require only one gradient query per iteration.

3.2 Optimistic Mirror Descent

We employ the algorithmic framework of Optimistic Mirror Descent (OMD) (Chiang et al., 2012; Rakhlin and Sridharan, 2013) as a generic building block for designing algorithms for non-stationary online learning. OMD is a methodology for optimistic online learning. Compared to the standard online learning setup, the player will now additionally receive an optimistic vector $M_t \in \mathbb{R}^d$ at each round, which serves as a hint of the future gradient. OMD starts from the initial point $\widehat{\mathbf{x}}_1 \in \mathcal{X}$ and performs the following two-step updates:

$$\begin{aligned} \mathbf{x}_t &= \operatorname*{arg\,min}_{\mathbf{x} \in \mathcal{X}} \ \eta_t \langle M_t, \mathbf{x} \rangle + \mathcal{D}_\psi(\mathbf{x}, \widehat{\mathbf{x}}_t), \\ \widehat{\mathbf{x}}_{t+1} &= \operatorname*{arg\,min}_{\mathbf{x} \in \mathcal{X}} \ \eta_t \langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle + \mathcal{D}_\psi(\mathbf{x}, \widehat{\mathbf{x}}_t), \end{aligned} \tag{6}$$

which firstly updates by the optimism and then updates by the received gradient. In above, $\eta_t > 0$ is a (potentially) time-varying step size, and $\mathcal{D}_\psi(\cdot, \cdot)$ denotes the Bregman divergence associated with the regularizer ψ defined as $\mathcal{D}_\psi(\mathbf{x}, \mathbf{y}) = \psi(\mathbf{x}) - \psi(\mathbf{y}) - \langle \nabla \psi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$. We have the following general result regarding dynamic regret of optimistic mirror descent.

Theorem 1. *Suppose that the regularizer $\psi : \mathcal{X} \mapsto \mathbb{R}$ is 1-strongly convex with respect to the norm $\|\cdot\|$, and let $\|\cdot\|_*$ be the dual norm of $\|\cdot\|$. The dynamic regret of Optimistic Mirror Descent whose update rule is specified in (6) is bounded as follows:*

$$\begin{aligned}\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \sum_{t=1}^{T} \eta_t \|\nabla f_t(\mathbf{x}_t) - M_t\|_*^2 &+ \sum_{t=1}^{T} \frac{1}{\eta_t}\Big(\mathcal{D}_\psi(\mathbf{u}_t, \widehat{\mathbf{x}}_t) - \mathcal{D}_\psi(\mathbf{u}_t, \widehat{\mathbf{x}}_{t+1})\Big) \\ &- \sum_{t=1}^{T} \frac{1}{\eta_t}\Big(\mathcal{D}_\psi(\widehat{\mathbf{x}}_{t+1}, \mathbf{x}_t) + \mathcal{D}_\psi(\mathbf{x}_t, \widehat{\mathbf{x}}_t)\Big),\end{aligned} \tag{7}$$

which holds for any comparator sequence $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$.

Remark 1. The dynamic regret upper bound in Theorem 1 consists of three terms:

(i) the first term $\sum_{t=1}^{T} \eta_t \|\nabla f_t(\mathbf{x}_t) - M_t\|_*^2$ is the *adaptivity* term that measures the deviation between gradient and optimism;

(ii) the second term $\sum_{t=1}^{T} \frac{1}{\eta_t}\big(\mathcal{D}_\psi(\mathbf{u}_t, \widehat{\mathbf{x}}_t) - \mathcal{D}_\psi(\mathbf{u}_t, \widehat{\mathbf{x}}_{t+1})\big)$ reflects the *non-stationarity* of environments and will be converted to the path length of comparators;

(iii) the last negative term $-\sum_{t=1}^{T} \frac{1}{\eta_t}\big(\mathcal{D}_\psi(\widehat{\mathbf{x}}_{t+1}, \mathbf{x}_t) + \mathcal{D}_\psi(\mathbf{x}_t, \widehat{\mathbf{x}}_t)\big)$ is crucial and will be greatly useful for problem-dependent bounds, particularly the gradient-variation one.

Moreover, we emphasize that the above regret guarantee is very general due to the flexibility in choosing the regularizer ψ and comparators $\{\mathbf{u}_t\}_{t=1,\ldots,T}$. For example, by choosing the negative-entropy regularizer and competing with the best fixed prediction, the result recovers the static regret bound of Optimistic Hedge (Syrgkanis et al., 2015); by choosing the Euclidean regularizer and competing with time-varying compactors, it recovers the dynamic regret bound of Online Gradient Descent (Zinkevich, 2003). The versatility of this optimistic mirror descent framework motivates us to use it as a unified building block for both algorithm design and theoretical analysis. ¶

3.3 Assumptions

In this part, we list several common assumptions that might be used in the theorems.

Assumption 1. The norm of the gradients of online functions over the domain $\mathcal{X}$ is bounded by G, i.e., $\|\nabla f_t(\mathbf{x})\|_2 \leq G$, for all $\mathbf{x} \in \mathcal{X}$ and $t \in [T]$.

Assumption 2. The domain $\mathcal{X} \subseteq \mathbb{R}^d$ contains the origin $\mathbf{0}$, and the diameter of the domain $\mathcal{X}$ is at most D, i.e., $\|\mathbf{x} - \mathbf{x}'\|_2 \leq D$ for any $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$.

Assumption 3. All the online functions are L-smooth, i.e., $\|\nabla f_t(\mathbf{x}) - \nabla f_t(\mathbf{x}')\|_2 \leq L\|\mathbf{x} - \mathbf{x}'\|_2$ for any $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ and $t \in [T]$.

Assumption 4. All the online functions are non-negative over $\mathbb{R}^d$.

We have the following remarks regarding the assumptions. The generic dynamic regret of OMD (Theorem 1) does not require the smoothness assumption (Assumption 3). Nevertheless, it is required in attaining the problem-dependent dynamic regret bounds. In fact, even in the static regret analysis, smoothness is demonstrated to be necessary for the first-order methods to achieve gradient-variation bounds (cf. Lemma 9 of Chiang et al. (2012) and Theorem 1 of Yang et al. (2014)). Therefore, throughout the paper we focus on the problem-dependent dynamic regret of convex and smooth functions. Moreover, note that in Assumption 4 we require the online functions to be non-negative outside the domain $\mathcal{X}$, which is a precondition for establishing the self-bounding property for smooth functions and is commonly used to establish small-loss bounds (Srebro et al., 2010; Zhang et al., 2019; Zhang and Zhou, 2019). Meanwhile, we treat double logarithmic factors in T as a constant, following previous studies (Adamskiy et al., 2012; Luo and Schapire, 2015).

4. Gradient-Variation Dynamic Regret

Our paper aims to develop online algorithms that can *simultaneously* achieve problem-dependent dynamic regret bounds, which scale with two problem-dependent quantities: the gradient-variation term V_T and the small-loss term F_T, as defined in (4). As we will demonstrate in the next section, the gradient-variation bound is more fundamental than the small-loss bound. Consequently, we start by focusing on the gradient-variation dynamic regret in this section. In Section 6, we will then present the small-loss bound and the best-of-both-worlds bound as implications of the results obtained in this section.

4.1 A Gentle Start

In the study of static regret, Chiang et al. (2012) propose the online extra-gradient descent (OEGD) algorithm and prove that the algorithm enjoys gradient-variation static regret. Specifically, OEGD starts from $\widehat{\mathbf{x}}_1, \mathbf{x}_1 \in \mathcal{X}$ and then updates by

$$\mathbf{x}_t = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_t - \eta \nabla f_{t-1}(\mathbf{x}_{t-1})\right], \quad \widehat{\mathbf{x}}_{t+1} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_t - \eta \nabla f_t(\mathbf{x}_t)\right]. \tag{8}$$

We here consider the algorithm with a fixed step size $\eta > 0$ for simplicity, and $\Pi_{\mathcal{X}}[\cdot]$ denotes the Euclidean projection onto the nearest point in $\mathcal{X}$. For convex and smooth functions, Chiang et al. (2012) prove that OEGD can achieve an $\mathcal{O}(\sqrt{1+V_T})$ static regret bound, where $V_T = \sum_{t=2}^T \sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$ is the gradient variation.

In the following, we further demonstrate that OEGD also enjoys the gradient-variation dynamic regret guarantee. Actually, OEGD can be viewed as a specialization of the optimistic mirror descent (6) presented in Section 3.2, by choosing the regularizer $\psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$ and the optimism $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$ as well as a fixed step size $\eta > 0$. Then, the general result of Theorem 1 directly implies the following dynamic regret bound for OEGD, and the proof can be found in Appendix A.

Lemma 1. *Under Assumptions 1, 2, and 3, by choosing $\eta \le \frac{1}{4L}$, the dynamic regret of OMD with a regularizer $\psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$ and optimism $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$ is bounded as*

$$\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) \le \eta(G^2 + 2V_T) + \frac{1}{2\eta}(D^2 + 2DP_T), \tag{9}$$

where $V_T = \sum_{t=2}^T \sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$ is the gradient variation and $P_T = \sum_{t=2}^T \|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2$ is the path length. The result holds for any comparator sequence $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$.

Lemma 1 immediately implies a static regret bound. Specifically, by choosing comparators as the best decision in hindsight $\mathbf{u}_1 = \ldots = \mathbf{u}_T \in \arg\min_{\mathbf{x}\in\mathcal{X}} \sum_{t=1}^T f_t(\mathbf{x})$ such that the path length $P_T = 0$, we obtain the existing result (Chiang et al., 2012, Theorem 11):

$$\sum_{t=1}^T f_t(\mathbf{x}_t) - \min_{\mathbf{x}\in\mathcal{X}} \sum_{t=1}^T f_t(\mathbf{x}) \leq \eta(G^2 + 2V_T) + \frac{D^2}{2\eta} = \mathcal{O}\Big(\sqrt{1+V_T}\Big),$$

where the last step holds by setting the step size $\eta = \eta^* = \min\{\sqrt{\frac{D^2}{G^2+2V_T}}, \frac{1}{4L}\}$. Note that the requirement on knowing V_T in the optimal tuning can be removed by either doubling trick (Cesa-Bianchi et al., 1997) or self-confident tuning (Auer et al., 2002).

However, it is more complicated when competing with a sequence of time-varying comparators. The dynamic regret bound exhibited in Lemma 1 suggests that it is crucial to tune the step size to balance the non-stationarity (path length P_T) and the adaptivity (gradient-variation V_T) in order to achieve a tight dynamic regret bound. Clearly, the optimal tuning is $\eta^* = \sqrt{(D^2 + 2DP_T)/(2G^2 + 2V_T)}$, which unfortunately requires the prior information of P_T and V_T that are generally unavailable. We emphasize that V_T is empirically observable in the sense that at round $t \in [T]$ one can observe its internal estimate $V_t = \sum_{s=2}^t \sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_s(\mathbf{x}) - \nabla f_{s-1}(\mathbf{x})\|_2^2$. By contrast, $P_T = \sum_{t=2}^T \|\mathbf{u}_t - \mathbf{u}_{t-1}\|_2$ remains unknown even after all iterations, since the comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T$ can be chosen arbitrarily as long as they are feasible in the domain and thus are entirely unknown. Consequently, we may use self-confident tuning to remove the dependence on the unknown gradient variation V_T, but the method cannot address the unknown path length P_T. In fact, this is the fundamental problem of non-stationary online learning — how to deal with uncertainty due to unknown environmental non-stationarity, captured by path length of comparators P_T in the language of dynamic regret minimization.

To simultaneously handle the uncertainty arising from adaptivity and non-stationarity, in addition to using optimistic online learning to reuse the historical gradients, we further design an adaptive online ensemble method that can hedge the non-stationarity while extracting the adaptivity. Our approach deploys a two-layer meta-base structure, in which multiple base-learners are maintained simultaneously and a meta-algorithm is used to track the best one. More concretely, we first construct a pool of candidate step sizes to discretize value range of the optimal step size, and then initialize multiple base-learners simultaneously, denoted by $\mathcal{B}_1, \ldots, \mathcal{B}_N$. Each base-learner $\mathcal{B}_i$ returns her own prediction $\mathbf{x}_{t,i}$ by running the base-algorithm with a particular step size η_i from the pool. Finally, the intermediate predictions of all the base-learners are combined by a meta-algorithm to produce the final output $\mathbf{x}_t = \sum_{i=1}^N p_{t,i}\mathbf{x}_{t,i}$, where $\boldsymbol{p}_t \in \Delta_N$ is the weight distribution.

According to the above procedure, we can decompose dynamic regret into two parts because of the two-layer meta-base structure.

$$\text{D-Regret}_T = \sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) = \underbrace{\sum_{t=1}^T f_t(\mathbf{x}_t) - f_t(\mathbf{x}_{t,i})}_{\texttt{meta-regret}} + \underbrace{\sum_{t=1}^T f_t(\mathbf{x}_{t,i}) - f_t(\mathbf{u}_t)}_{\texttt{base-regret}}, \tag{10}$$

where $\{\mathbf{x}_t\}_{t=1,\ldots,T}$ denotes the final output sequence, and $\{\mathbf{x}_{t,i}\}_{t=1,\ldots,T}$ is the prediction sequence of base-learner $\mathcal{B}_i$. Notably, the decomposition holds for any base-learner's index

$i \in [N]$. The first part is the difference between cumulative loss of the final output sequence and that of the prediction sequence of base-learner $\mathcal{B}_i$, which is introduced by the meta-algorithm and thus named as *meta-regret*; the second part is the dynamic regret of base-learner $\mathcal{B}_i$ and therefore called *base-regret*. As a result, we need to make the meta-regret and base-regret scaling with V_T to achieve the desired gradient-variation dynamic regret.

In the following, we present two solutions. The first solution, developed in our conference paper (Zhao et al., 2020b), is conceptually simpler but requires $N = \mathcal{O}(\log T)$ gradient queries at each round, making it suitable only for the multi-gradient feedback model. The second solution is an improved algorithm based on a refined analysis of the problem's structure, which attains the same dynamic regret guarantee with only one gradient per iteration and hence suits for the more challenging one-gradient feedback model. Recall that the definitions of the multi/one-gradient feedback models are presented in Section 3.1.

4.2 Multi-Gradient Feedback: Sword

Our approach, Sword, implements a meta-base online ensemble structure, in which multiple base-learners are initiated simultaneously (denoted by $\mathcal{B}_1, \ldots, \mathcal{B}_N$) and the intermediate predictions of all the base-learners are combined by a meta-algorithm to produce the final output. Below, we describe the specific settings of the base-algorithm and meta-algorithm.

For the base-algorithm, we simply employ the OEGD algorithm (Chiang et al., 2012), where the base-learner $\mathcal{B}_i$ shall update her local decision $\{\mathbf{x}_{t,i}\}_{t=1,\ldots,T}$ by

$$\mathbf{x}_{t,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_{t-1}(\mathbf{x}_{t-1,i})\right], \quad \widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_{t,i})\right], \tag{11}$$

where $\eta_i > 0$ is the associated step size from the step size pool $\mathcal{H} = \{\eta_1, \ldots, \eta_N\}$ and the number of base-learner is chosen as $N = \mathcal{O}(\log T)$. Lemma 1 ensures an upper bound of base-regret scaling with the gradient variation, i.e., $\sum_{t=1}^T f_t(\mathbf{x}_{t,i}) - \sum_{t=1}^T f_t(\mathbf{u}_t) \leq \mathcal{O}(\eta_i(1+V_T) + P_T/\eta_i)$, whenever the step size satisfies $\eta_i \leq 1/(4L)$. The caveat is that each base-learner requires her own gradient direction for the update, so we need the gradient information of $\{\nabla f_t(\mathbf{x}_{t,i})\}_{i=1,\ldots,N}$ at round $t \in [T]$. Notably, the gradient query complexity is $N = \mathcal{O}(\log T)$ per round instead of one as was desired, consequently, the method developed in this part only suits for the multi-gradient feedback model. We leave the gradient query complexity issue for a moment and will resolve it and design an improved algorithm applicable for the one-gradient feedback model in Section 4.3.

The main difficulty lies in the design and analysis of an appropriate meta-algorithm. In order to be compatible to the gradient-variation base-regret, the meta-algorithm is required to incur a problem-dependent meta-regret of order $\mathcal{O}(\sqrt{V_T \ln N})$. However, the meta-algorithms used in existing studies (van Erven and Koolen, 2016; Zhang et al., 2018a) cannot satisfy the requirements. For example, the vanilla Hedge suffers an $\mathcal{O}(\sqrt{T \ln N})$ meta-regret, which is problem-independent and thus not suitable for us. To this end, we design a novel variant of Hedge by leveraging the technique of optimistic online learning with carefully designed optimism, specifically for our problem. Consider the problem of prediction with expert advice. At the beginning of iteration $(t+1)$, in addition to the loss vector $\boldsymbol{\ell}_t \in \mathbb{R}^N$ returned by the experts, the player can receive an optimism $\boldsymbol{m}_{t+1} \in \mathbb{R}^N$.

Optimistic Hedge updates the weight vector $\boldsymbol{p}_{t+1} \in \Delta_N$ by

$$p_{t+1,i} \propto \exp\left(-\varepsilon\Big(\sum_{s=1}^{t} \ell_{s,i} + m_{t+1,i}\Big)\right), \quad \forall i \in [N]. \tag{12}$$

Here, $\varepsilon > 0$ is the learning rate of the meta-algorithm and we consider a fixed learning rate for simplicity.[1] The optimism $\boldsymbol{m}_{t+1} \in \mathbb{R}^N$ can be interpreted as an optimistic guess of the loss of round $t+1$, and we thus incorporate it into the cumulative loss for update. It turns out that Optimistic Hedge can be regarded as an instance of OMD with the negative-entropy regularizer, as mentioned in Remark 1. Therefore, the general result of Theorem 1 implies the following static regret bound of Optimistic Hedge, and the proof can be found in Appendix A. Notably, the negative term shown in (13) will be of great importance in the algorithm design and regret analysis.

Lemma 2. *The regret of Optimistic Hedge with a fixed learning rate $\varepsilon > 0$ to any expert $i \in [N]$ is at most*

$$\sum_{t=1}^{T} \langle \boldsymbol{p}_t, \boldsymbol{\ell}_t \rangle - \sum_{t=1}^{T} \ell_{t,i} \leq \varepsilon \sum_{t=1}^{T} \|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2 + \frac{\ln N}{\varepsilon} - \frac{1}{4\varepsilon} \sum_{t=2}^{T} \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2. \tag{13}$$

Let $D_T = \sum_{t=1}^{T} \|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2$ measure the deviation between optimism and gradient. With a proper learning rate tuning scheme, Optimistic Hedge enjoys an $\mathcal{O}(\sqrt{D_T \ln N})$ meta-regret.

The framework of optimistic online learning is very powerful for designing adaptive methods, in that the adaptivity quantity $D_T = \sum_{t=1}^{T} \|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2$ is very general and can be specialized flexibly with different configurations of feedback loss $\boldsymbol{\ell}_t$ and optimism $\boldsymbol{m}_t$. To achieve the desired gradient-variation dynamic regret, we need to investigate the online ensemble structure carefully. To this end, we specialize Optimistic Hedge in the following way to make the meta-regret compatible with the desired gradient-variation quantity.

- The feedback loss $\boldsymbol{\ell}_t \in \mathbb{R}^N$ is set as the linearized surrogate loss: for $t \in [T]$ and each $i \in [N]$,
$$\ell_{t,i} = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t,i} \rangle. \tag{14}$$
- The optimism $\boldsymbol{m}_t \in \mathbb{R}^N$ is set with a careful design: $\boldsymbol{m}_1 = \mathbf{0}$ and for $t \geq 2$ and each $i \in [N]$,
$$m_{t,i} = \langle \nabla f_{t-1}(\bar{\mathbf{x}}_t), \mathbf{x}_{t,i} \rangle, \text{ where } \bar{\mathbf{x}}_t = \sum_{i=1}^{N} p_{t-1,i} \mathbf{x}_{t,i}. \tag{15}$$

We will explain the motivation of such designs in Remark 2. Note that this construction of optimism is legitimate as the instrumental variable $\bar{\mathbf{x}}_t$ only uses the information of $\boldsymbol{p}_{t-1}$ as well as the local decisions $\{\mathbf{x}_{t,i}\}_{i=1,\ldots,N}$ at time t. Thus, the meta-algorithm of Sword updates the weight $\boldsymbol{p}_{t+1} \in \mathbb{R}^N$ by

$$p_{t+1,i} \propto \exp\left(-\varepsilon\Big(\sum_{s=1}^{t} \langle \nabla f_s(\mathbf{x}_s), \mathbf{x}_{s,i} \rangle + \langle \nabla f_t(\bar{\mathbf{x}}_{t+1}), \mathbf{x}_{t+1,i} \rangle\Big)\right), \quad \forall i \in [N]. \tag{16}$$

[1] We adopt the terminology "learning rate" for the meta-algorithm of our approach following the convention in the prediction with expert advice, and use "step size" for the general online convex optimization.

Algorithm 1 Sword: meta-algorithm

```
Input: step size pool $\mathcal{H}$; learning rate $\varepsilon$
Initialization: $\forall i \in [N], p_{0,i} = 1/N$
for $t = 1$ to $T$ do
    Receive $\mathbf{x}_{t+1,i}$ from base-learner $\mathcal{B}_i$
    Update weight $p_{t+1,i}$ by (16)
    Predict $\mathbf{x}_{t+1} = \sum_{i=1}^{N} p_{t+1,i}\mathbf{x}_{t+1,i}$
end for
```

Algorithm 2 Sword: base-algorithm

```
Input: step size $\eta_i \in \mathcal{H}$
Let $\widehat{\mathbf{x}}_{1,i}, \mathbf{x}_{1,i}$ be any point in $\mathcal{X}$
for $t = 1$ to $T$ do
    $\widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_{t,i})\right]$
    $\mathbf{x}_{t+1,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t+1,i} - \eta_i \nabla f_t(\mathbf{x}_{t,i})\right]$
    Send $\mathbf{x}_{t+1,i}$ to the meta-algorithm
end for
```

Algorithm 1 summarizes detailed procedures of the meta-algorithm, which in conjunction with the base-algorithm of Algorithm 2 yields the Sword algorithm.

Remark 2 (design of optimism)**.** The design of optimism in (15), particularly the construction of the instrumental variable $\bar{\mathbf{x}}_t$, is crucial and is the most challenging part in this method. Our design carefully leverages the structure of two-layer online ensemble methods, specifically, the goal of designing optimism is to approximate the current gradient $\nabla f_t(\mathbf{x}_t)$ (which is unknown) via the available knowledge till round t. We propose to use $\nabla f_{t-1}(\bar{\mathbf{x}}_t)$ as the approximation, and the difference of online functions delivers the gradient-variation term $\sup_{\mathbf{x}\in\mathcal{X}}\|f_t(\mathbf{x})-f_{t-1}(\mathbf{x})\|_2^2$, while the difference between $\mathbf{x}_t$ and $\bar{\mathbf{x}}_t$ can be upper bounded by the decision variation of the meta-algorithm,

$$\|\mathbf{x}_t - \bar{\mathbf{x}}_t\|_2^2 = \left\|\sum_{i=1}^{N}(p_{t,i} - p_{t-1,i})\mathbf{x}_{t,i}\right\|_2^2 \leq \Big(\sum_{i=1}^{N}|p_{t,i} - p_{t-1,i}|\|\mathbf{x}_{t,i}\|_2\Big)^2 \leq D^2\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2, \quad (17)$$

which can be eliminated by the *negative term* in the regret bound of Optimistic Hedge as shown in (13), providing with a suitable setting for the learning rate of the meta-algorithm. Summarizing, the aforementioned configurations of feedback loss and optimism will convert the adaptive quantity D_T to the desired gradient variation V_T plus the decision variation of the meta-algorithm, concretely,

$$\begin{aligned}
\|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2 &\overset{(15)}{=} \max\nolimits_{i\in[N]}\langle \nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\bar{\mathbf{x}}_t), \mathbf{x}_{t,i}\rangle^2 \\
&\leq D^2\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\bar{\mathbf{x}}_t)\|_2^2 \\
&\leq 2D^2(\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_t)\|_2^2 + \|\nabla f_{t-1}(\mathbf{x}_t) - \nabla f_{t-1}(\bar{\mathbf{x}}_t)\|_2^2) \\
&\leq 2D^2 \sup\nolimits_{\mathbf{x}\in\mathcal{X}}\|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2 + 2D^2L^2\|\mathbf{x}_t - \bar{\mathbf{x}}_t\|_2^2 \\
&\leq 2D^2 \sup\nolimits_{\mathbf{x}\in\mathcal{X}}\|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2 + 2D^4L^2\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2,
\end{aligned}$$

where the derivation makes use of the boundedness of the feasible domain, triangle inequality, and the smoothness of online functions. The last term will be canceled by the negative term in the meta-regret, then we obtain the desired gradient-variation regret guarantee. ¶

The following theorem shows that the meta-regret is at most $\mathcal{O}(\sqrt{(1+V_T)\ln N})$, which is nicely compatible to the attained base-regret. The proof can be found in Section 7.2.

Theorem 2. *Under Assumptions 1, 2, and 3, by setting the learning rate of the meta-algorithm* (16) *optimally as* $\varepsilon = \min\{1/(4D^2L), \sqrt{(\ln N)/(2D^2(G^2+V_T))}\}$, *the meta-regret of Sword (Algorithm 1) is at most*

$$\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{x}_{t,i}) \le 2D\sqrt{2(G^2+V_T)\ln N} + 8D^2L\ln N = \mathcal{O}\Big(\sqrt{(1+V_T)\ln N}\Big).$$

Note that the optimal learning rate tuning of the meta-algorithm requires the knowledge of gradient variation $V_T = \sum_{t=2}^T \sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$. The undesired demand can be removed by the self-confident tuning method (Auer et al., 2002), which employs a time-varying learning rate scheme for the meta-algorithm's update based on internal estimates, roughly, $p_{t+1,i} \propto \exp\big(-\varepsilon_t(\sum_{s=1}^t \ell_{s,i} + m_{t+1,i})\big), \forall i \in [N]$ with $\varepsilon_t = \mathcal{O}(1/\sqrt{1+V_t})$ with an internal estimate $V_t = \sum_{s=1}^t \sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_s(\mathbf{x}) - \nabla f_{s-1}(\mathbf{x})\|_2^2$. Besides, notice that this V_t is actually not easy to calculate due to the computation of instantaneous variation $\sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$, which is a difference of convex functions programming and is not easy to solve even with the explicit form of functions f_t and f_{t-1}. Fortunately, we can use an alternative twisted quantity $\bar{V}_T = \sum_{t=2}^T \|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2$ for the learning rate configuration and also achieve the same regret bound via a refined analysis. Then, it suffices to perform the self-confident tuning over $\bar{V}_T$ by monitoring the corresponding internal estimate $\bar{V}_t = \sum_{s=2}^t \|\nabla f_s(\mathbf{x}_s) - \nabla f_{s-1}(\mathbf{x}_{s-1})\|_2^2$, which avoids the burdensome calculations of inner optimization problems and thereby significantly streamlines the computational efforts paid for the adaptive learning rate tuning.

So far, the obtained base-regret bound (Lemma 1) and meta-regret bound (Theorem 2) are both adaptive to the gradient variation, and we can simply combine them to achieve the final gradient-variation dynamic regret as stated in Theorem 3, providing with an appropriate candidate step size pool. The proof is provided in Section 7.3.

Theorem 3. *Under Assumptions 1, 2, and 3, set the pool of candidate step sizes* $\mathcal{H}$ *as*

$$\mathcal{H} = \left\{ \eta_i = \min\left\{\frac{1}{4L}, 2^{i-1}\sqrt{\frac{D^2}{8G^2T}}\right\} \mid i \in [N] \right\}, \tag{18}$$

where $N = \lceil 2^{-1}\log_2(G^2T/(2D^2L^2))\rceil + 1$ *is the number of candidate step sizes; and set the learning rate of the meta-algorithm as* $\varepsilon = \min\{1/(4D^2L), \sqrt{(\ln N)/(2D^2(G^2+V_T))}\}$. *Then, Sword satisfies that*

$$\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) \le \mathcal{O}\Big(\sqrt{(1+P_T+V_T)(1+P_T)}\Big),$$

which holds for any comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$. *In above,* $V_T = \sum_{t=2}^T \sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$ *is the gradient variation, and* $P_T = \sum_{t=2}^T \|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2$ *is the path length.*

Remark 3. Compared with the existing $\mathcal{O}(\sqrt{T(1+P_T)})$ dynamic regret (Zhang et al., 2018a), our result is more adaptive in the sense that it replaces T by the *problem-dependent* quantity $P_T + V_T$. Therefore, the bound will be much tighter in easy problems, for example when both V_T and P_T are $o(T)$. Meanwhile, it safeguards the same minimax rate, since both quantities are at most $\mathcal{O}(T)$. Furthermore, because the *universal* dynamic regret studied in this paper holds against any comparator sequence, it specializes the static regret by setting all comparators as the best fixed decision in hindsight, i.e., $\mathbf{u}_1 = \ldots = \mathbf{u}_T = \mathbf{x}^* \in \arg\min_{\mathbf{x}\in\mathcal{X}} \sum_{t=1}^T f_t(\mathbf{x})$. Under such a circumstance, the path length $P_T = \sum_{t=2}^T \|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2$ becomes zero, so the regret bound in Theorem 3 implies an $\mathcal{O}(\sqrt{1+V_T})$ gradient-variation static regret bound, recovering the result of Chiang et al. (2012). ¶

4.3 One-Gradient Feedback: Sword++

So far, we have designed an online algorithm (Sword) with the gradient-variation dynamic regret. While it achieves a favorable regret guarantee, one caveat is that Sword runs $N = \mathcal{O}(\log T)$ base-learners simultaneously and each base-learner requires her own gradient direction for the update. Consequently, the overall algorithm necessitates $\mathcal{O}(\log T)$ gradient queries at each iteration, making it time-consuming and only applicable to the multi-gradient feedback model. In contrast, algorithms designed for static regret minimization typically work well under the more challenging one-gradient feedback model, namely, they only require the gradient information $\nabla f_t(\mathbf{x}_t)$ for updates. Given this, it is natural to ask whether it is possible to design online algorithms that can achieve favorable dynamic regret while only using one gradient query per iteration, making them applicable to the one-gradient feedback online learning.

We resolve the question affirmatively by designing an algorithm that requires only one gradient query per iteration and provably enjoys the same gradient-variation dynamic regret as Sword. The new algorithm, called Sword++, also implements an online ensemble structure. Comparing to Sword presented in Section 4.2, the key novel ingredient is the framework of *collaborative online ensemble*. We carefully introduce correction terms to the online loss and optimism, forming a biased surrogate loss and a surrogate optimism, which are then fed to the meta-algorithm. By further exploiting the negative terms in the meta and base levels, the overall algorithm ensures effective *collaboration* within the meta and base two layers, thereby achieving the favorable gradient-variation dynamic regret with only one gradient query per iteration.

In the following, we describe the details of Sword++. The algorithm maintains multiple base-learners denoted by $\mathcal{B}_1, \ldots, \mathcal{B}_N$, which are performed with different step sizes and then combined by a meta-algorithm to track the best one. An exponential step size grid is adopted as the schedule, denoted by $\mathcal{H} = \{\eta_i = c \cdot 2^i \mid i \in [N]\}$ with $N = \mathcal{O}(\log T)$ for some constant $c > 0$ (usually scaling with $\mathrm{poly}(1/T)$), whose specific setting will be given later.

Base-algorithm. Instead of performing updates over the original loss f_t as shown in (11), all the base-learners of Sword++ update over the *linearized surrogate loss* $g_t : \mathcal{X} \mapsto \mathbb{R}$ defined $g_t(\mathbf{x}) = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle$, and moreover, the optimism is chosen as $M_t = \nabla g_{t-1}(\mathbf{x}_{t-1,i})$ for each base-learner $\mathcal{B}_i$ with $i \in [N]$. By definition, we have $\nabla g_t(\mathbf{x}_{t,i}) = \nabla f_t(\mathbf{x}_t)$, so each base-learner $\mathcal{B}_i$ essentially performs the following update at each iteration:

$$\mathbf{x}_{t,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_{t-1}(\mathbf{x}_{t-1})\right], \quad \widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_t)\right]. \tag{19}$$

Using above update steps, we no longer need to evaluate the gradient $\nabla f_t(\mathbf{x}_{t,i})$ over the local decisions for every base-learner, as was done by Sword (see its update rule in (11)). Instead, a single call of $\nabla f_t(\mathbf{x}_t)$ is sufficient at each round for the update in Sword++.

We note that although the linearized trick has previously been employed in the meta-base structure for achieving the minimax dynamic regret $\mathcal{O}(\sqrt{T(1+P_T)})$ with one gradient per iteration (Zhang et al., 2018a), this modification alone is far from enough to obtain problem-dependent dynamic regret. To see this, we can check the regret of the base-learner updated with the surrogate loss $g_t(\mathbf{x})$. A similar argument of Lemma 1 shows that the base-regret over the linearized loss $(\#) \triangleq \sum_{t=1}^T g_t(\mathbf{x}_{t,i}) - \sum_{t=1}^T g_t(\mathbf{u}_t)$ satisfies

$$(\#) \leq \eta_i(G^2 + 2V_T) + \frac{D^2 + 2DP_T}{2\eta_i} + 4L^2 \sum_{t=2}^T \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 - \frac{1}{\eta_i}\sum_{t=2}^T \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2.$$

In the analysis of Sword, because the gradient $\nabla f_t(\mathbf{x}_{t,i})$ is evaluated on every base-learner's own decision $\mathbf{x}_{t,i}$, the additional positive term (the third one) is $4L^2 \sum_{t=2}^T \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$, which can be cancelled by the negative term $-\sum_{t=2}^T \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2/\eta_i$ whenever the step size is set appropriately. However, when the base-learner updates her decision over the surrogate loss, the additional positive term becomes $4L^2\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2$, which *cannot* be handled by the negative term of any base-learner. Thus, more advanced mechanisms are required to achieve the problem-dependent dynamic regret under the one-gradient query model.

To tackle the difficulty, our primary idea is to facilitate *collaboration* between the meta and base levels. Specifically, we aim to leverage negative terms from both levels to handle the positive term. However, it turns out that the positive term cannot be entirely offset by the combined negative terms from meta and base levels. To address this issue, we introduce correction terms to the feedback loss and optimism in the meta-algorithm. This generates a new negative term that, together with the negative term from the meta level, effectively cancels out the positive term. Nevertheless, another new positive term emerges due to the injected correction, which we ensure can be managed by the negative term from the base level. As a result, the overall undesired positive term is finally addressed.

The above forms the main idea of our proposed *collaborative online ensemble* framework. In the following, we outline the specific setup of the meta-algorithm. We will provide a brief explanation of the design of corrections in Remark 4 and offer a more comprehensive elaboration on the general framework of collaborative online ensemble in Section 5.

Meta-algorithm. We still employ Optimistic Hedge (OMD with the negative-entropy regularizer) as the meta-algorithm, but nevertheless require innovative design in the feedback loss and optimism. Specifically, instead of simply using the linearized surrogate loss $\ell_{t,i} \triangleq \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t,i}\rangle$ as the feedback loss like Sword (see the update rule in (14)), we carefully construct the surrogate loss in the following way and send it to the meta-algorithm.

- The feedback loss $\boldsymbol{\ell}_t \in \mathbb{R}^N$ is constructed as follows: for each $i \in [N]$, $\ell_{1,i} = \langle \nabla f_1(\mathbf{x}_1), \mathbf{x}_{1,i}\rangle$ and for $t \geq 2$, it composes the linearized surrogate loss $\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t,i}\rangle$ with a *decision-deviation correction* term, namely,
$$\ell_{t,i} = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t,i}\rangle + \lambda\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2. \tag{20}$$

- The optimism $\boldsymbol{m}_t \in \mathbb{R}^N$ is similarly configured as follows: $\boldsymbol{m}_1 = \mathbf{0}$ and for $t \geq 2$ and $i \in [N]$, the optimism also admits a *decision-deviation correction* term, namely,
$$m_{t,i} = \langle M_t, \mathbf{x}_{t,i}\rangle + \lambda\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 = \langle \nabla f_{t-1}(\mathbf{x}_{t-1}), \mathbf{x}_{t,i}\rangle + \lambda\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2. \tag{21}$$

Both feedback loss and optimism admit an additional correction term $\lambda\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$ that measures the stability of the local decisions returned by the base-learner, where $\lambda > 0$ is the correction coefficient to be determined later. Overall, the meta-algorithm of Sword++ updates the weight $\boldsymbol{p}_{t+1} \in \mathbb{R}^N$ as follows: for any $i \in [N]$,

$$p_{t+1,i} \propto \exp\left(-\varepsilon\Big(\sum_{s=1}^{t} \ell_{s,i} + m_{t+1,i}\Big)\right), \tag{22}$$

where $\varepsilon > 0$ is (for simplicity) chosen as a fixed learning rate of the meta-algorithm, and the feedback loss $\boldsymbol{\ell}_t$ and optimism $\boldsymbol{m}_t$ are defined in (20) and (21), respectively. Notably, the meta-algorithm only requires the gradient information of $\nabla f_t(\mathbf{x}_t)$ at round t and thus is feasible in the one-gradient feedback model.

Remark 4 (design of correction term)**.** We emphasize that the correction term $\lambda\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$, appearing in the construction of both feedback loss and optimism, is crucial for the design and is the most challenging part in this method. We briefly explain the motivation. As we mentioned earlier, the use of linearized surrogate loss $g_t(\mathbf{x})$ will introduce an additional term $\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2$, which *cannot* be canceled by the negative term of any base-regret, namely, $-\sum_{t=2}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$. To address the difficulty, we scrutinize the positive term and find that actually it can be further expanded as:

$$\begin{aligned}
\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 &= \left\|\sum_{i=1}^{N} p_{t,i}\mathbf{x}_{t,i} - \sum_{i=1}^{N} p_{t-1,i}\mathbf{x}_{t-1,i}\right\|_2^2 \\
&\le 2\left\|\sum_{i=1}^{N} p_{t,i}\mathbf{x}_{t,i} - \sum_{i=1}^{N} p_{t,i}\mathbf{x}_{t-1,i}\right\|_2^2 + 2\left\|\sum_{i=1}^{N} p_{t,i}\mathbf{x}_{t-1,i} - \sum_{i=1}^{N} p_{t-1,i}\mathbf{x}_{t-1,i}\right\|_2^2 \\
&\le 2\left(\sum_{i=1}^{N} p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2\right)^2 + 2\left(\sum_{i=1}^{N} |p_{t,i} - p_{t-1,i}|\|\mathbf{x}_{t-1,i}\|_2\right)^2 \\
&\le 2\sum_{i=1}^{N} p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 + 2D^2\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2,
\end{aligned}$$

which concludes that

$$\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 \le 2\sum_{t=2}^{T}\sum_{i=1}^{N} p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 + 2D^2\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2. \tag{23}$$

Consequently, it is crucial to exploit negative terms in both meta-regret and base-regret to cancel out the positive term. The second positive term (deviation of meta-algorithm's weights) can be readily canceled by the negative term of meta-regret. However, addressing the first positive term is more intricate, as it is essentially a weighted combination of the base-learners' decision variation. We tackle this positive term by algorithmically adding the decision-variation correction term in the feedback loss and optimism of the meta-algorithm, as well as leveraging the negative term of base-regret. The underlying intuition is to penalize base-learners with large decision variations, so as to ensure a small enough variation of final decisions. As such, we have facilitated the collaborations between the base and meta levels

Algorithm 3 Sword++: meta-algorithm

```
Input: step size pool $\mathcal{H}$; learning rate $\varepsilon$
Initialization: $\forall i \in [N], p_{0,i} = 1/N$
for $t = 1$ to $T$ do
    Receive $\mathbf{x}_{t+1,i}$ from base-learner $\mathcal{B}_i$
    Update weight $p_{t+1,i}$ by (20)–(22)
    Predict $\mathbf{x}_{t+1} = \sum_{i=1}^{N} p_{t+1,i}\mathbf{x}_{t+1,i}$
end for
```

Algorithm 4 Sword++: base-algorithm

```
Input: step size $\eta_i \in \mathcal{H}$
Let $\widehat{\mathbf{x}}_{1,i}, \mathbf{x}_1$ be any point in $\mathcal{X}$
for $t = 1$ to $T$ do
    $\widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_t)]$
    $\mathbf{x}_{t+1,i} = \Pi_{\mathcal{X}}[\widehat{\mathbf{x}}_{t+1,i} - \eta_i \nabla f_t(\mathbf{x}_t)]$
    Send $\mathbf{x}_{t+1,i}$ to the meta-algorithm
end for
```

— the overall positive term ($\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2$) is jointly cancelled out by the negative term of meta-regret ($-\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2$) and the one due to the injected corrections ($-\sum_{t=2}^{T}\sum_{i=1}^{N} p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$); and meanwhile, the injected corrections will introduce a new positive term ($\sum_{t=2}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$), which can be further tackled by the negative term of base-regret ($-\sum_{t=2}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$). Only through such collaborations within the two-layer online ensembles can the proposed Sword++ algorithm attain the desired gradient-variation dynamic regret, utilizing only one gradient per iteration. A more in-depth discussion on this collaborative online ensemble will be presented in Section 5. ¶

We summarize the procedures of Sword++ in Algorithm 3 (meta-algorithm) and Algorithm 4 (base-algorithm). Though multiple base-learners are performed with different step sizes to tackle the uncertainty of non-stationary environments, Sword++ requires the gradient information of $\nabla f_t(\mathbf{x}_t)$ only at round t and then broadcasts it to all the base-learners for local update. Therefore, Sword++ is feasible for the one-gradient feedback model. Moreover, the algorithm provably achieves the same gradient-variation dynamic regret as Sword, shown in Theorem 4, whose proof is presented in Section 7.4.

Theorem 4. *Under Assumptions 1, 2, and 3, set the pool of candidate step sizes $\mathcal{H}$ as*

$$\mathcal{H} = \left\{ \eta_i = \min\left\{ \frac{1}{8L}, \sqrt{\frac{D^2}{8G^2T}} \cdot 2^{i-1} \right\} \mid i \in [N] \right\}, \tag{24}$$

where $N = \lceil 2^{-1}\log_2(G^2T/(8D^2L^2)) \rceil + 1$ is the number of candidate step sizes; further set the correction coefficient as $\lambda = 2L$ and the learning rate of the meta-algorithm as $\varepsilon = \min\left\{1/(8D^2L), \sqrt{(\ln N)/(D^2(\|\nabla f_1(\mathbf{x}_1)\|_2^2 + \bar{V}_T))}\right\}$. Then, Sword++ satisfies

$$\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \mathcal{O}\left(\sqrt{(1 + P_T + V_T)(1 + P_T)}\right)$$

for any comparator sequence $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$. In above, $V_T = \sum_{t=2}^{T}\sup_{\mathbf{x}\in\mathcal{X}}\|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$ is the gradient variation, $\bar{V}_T = \sum_{t=2}^{T}\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2$ is the variant of V_T, and $P_T = \sum_{t=2}^{T}\|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2$ is the path length.

Note that the optimal learning rate tuning of the meta-algorithm requires the knowledge of $\bar{V}_T = \sum_{t=2}^{T}\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2$. Actually, this unpleasant dependence can be removed by performing the self-confident tuning over $\bar{V}_T$ by monitoring the internal estimate $\bar{V}_t = \sum_{s=2}^{t} \|\nabla f_s(\mathbf{x}_s) - \nabla f_{s-1}(\mathbf{x}_{s-1})\|_2^2$. Importantly, this adaptive learning rate tuning can be realized under the one-gradient feedback model, namely, only $\nabla f_t(\mathbf{x}_t)$ available at round t. To avoid clutters, we here stick to a fixed learning rate with dependence on $\bar{V}_T$ and defer an adaptive learning rate version to Appendix B.

Up to now, we have shown that it is possible to design online methods to achieve stronger guarantees than static methods (recall that universal dynamic regret can immediately imply static regret) under the challenging one-gradient feedback online learning, and meanwhile suffer no computational overhead in terms of the gradient query complexity.

5. A Generic Framework: Collaborative Online Ensemble

In this section, we formally introduce the proposed *collaborative online ensemble* framework, a generic algorithmic template designed to achieve (problem-dependent) dynamic regret guarantees. This framework is particularly crucial for attaining gradient-variation bounds. As will be demonstrated shortly, our proposed Sword (in Section 4.2) and Sword++ (in Section 4.3) can both be considered as specific instantiations.

5.1 Algorithmic Template

We focus on the standard OCO setup as specified in Section 3.1. At iteration $t \in [T]$, the player first chooses the decision $\mathbf{x}_t \in \mathcal{X}$, then the environments reveal the loss function $f_t : \mathcal{X} \mapsto \mathbb{R}$. Subsequently, the player suffers the loss $f_t(\mathbf{x}_t)$ and observes certain gradient information of $\nabla f_t(\cdot)$ according to the feedback model.

The overall algorithmic template implements a meta-base two-layer online ensemble. There are three crucial ingredients in collaborative online ensemble: (i) the surrogate loss, (ii) the surrogate optimism, and (iii) the correction terms. Additionally, the negative terms, hidden in the analysis, play a significant role within the framework. To better present the algorithmic template, we introduce the following notations:

- for the base-algorithm, let $g_t^{\mathtt{base}}(\cdot) : \mathcal{X} \mapsto \mathbb{R}$ be the base surrogate loss and $h_t^{\mathtt{base}}(\cdot) : \mathcal{X} \mapsto \mathbb{R}$ be the base surrogate optimism;
- for the meta-algorithm, let $g_t^{\mathtt{meta}}(\cdot) : \mathcal{X} \mapsto \mathbb{R}$ be the meta surrogate loss and $h_t^{\mathtt{meta}}(\cdot) : \mathcal{X} \mapsto \mathbb{R}$ be the meta surrogate optimism, and let $\boldsymbol{c}_t \in \mathbb{R}^d$ be the correction term.

The base-algorithm updates the decisions $\{\mathbf{x}_{t,i}\}_{i=1}^N$ by Optimistic OGD over the base surrogate loss and optimism, that is,

$$\mathbf{x}_{t,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla h_t^{\mathtt{base}}(\mathbf{x}_{t-1,i})\right], \quad \widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla g_t^{\mathtt{base}}(\mathbf{x}_{t,i})\right], \tag{25}$$

where $\eta_i > 0$ is a fixed step size specified by the step size pool $\mathcal{H} = \{\eta_1, \ldots, \eta_N\}$. Subsequently, the player makes the final decision at this round by $\mathbf{x}_t = \sum_{i=1}^N p_{t,i} \mathbf{x}_{t,i}$.

The meta-algorithm will then update the weight $\boldsymbol{p}_{t+1} \in \Delta_N$ by Optimistic Hedge over the feedback loss $\boldsymbol{\ell}_t \in \mathbb{R}^d$ and optimism $\boldsymbol{m}_t \in \mathbb{R}^d$,

$$p_{t+1,i} \propto \exp\left(-\varepsilon\Big(\sum_{s=1}^{t} \ell_{s,i} + m_{t+1,i}\Big)\right), \tag{26}$$

Table 1: Summary of three instantiations of the collaborative online ensemble framework, including Sword, Sword++, and Sword.optimism.

Algorithm	$g_t^{\mathtt{base}}(\mathbf{x})$	$h_t^{\mathtt{base}}(\mathbf{x})$	$g_t^{\mathtt{meta}}(\mathbf{x})$	$h_t^{\mathtt{meta}}(\mathbf{x})$	$\boldsymbol{c}_t$
Sword	$f_t(\mathbf{x})$	$f_{t-1}(\mathbf{x})$	$\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle$	$\langle \nabla f_{t-1}(\bar{\mathbf{x}}_t), \mathbf{x}\rangle$	$\boldsymbol{c}_t = \mathbf{0}$
Sword++	$\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle$	$\langle \nabla f_{t-1}(\mathbf{x}_{t-1}), \mathbf{x}\rangle$	$\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle$	$\langle \nabla f_{t-1}(\mathbf{x}_{t-1}), \mathbf{x}\rangle$	$c_{t,i} = \lVert \mathbf{x}_{t,i} - \mathbf{x}_{t-1,i} \rVert_2^2$
Sword.optimism	$\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle$	$\langle M_t, \mathbf{x}\rangle$	$\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle$	$\langle M_t, \mathbf{x}\rangle$	$c_{t,i} = \lVert \mathbf{x}_{t,i} - \mathbf{x}_{t-1,i} \rVert_2^2$

where $\varepsilon > 0$ is (for simplicity) chosen as a fixed learning rate of the meta-algorithm and the feedback loss $\ell_t \in \mathbb{R}^d$ and optimism $\boldsymbol{m}_t \in \mathbb{R}^d$ are defined as

$$\ell_{t,i} = g_t^{\mathtt{meta}}(\mathbf{x}_{t,i}) + \lambda c_{t,i}, \text{ and } m_{t,i} = h_t^{\mathtt{meta}}(\mathbf{x}_{t,i}) + \lambda c_{t,i}, \tag{27}$$

with $\lambda \geq 0$ being the coefficient of the correction terms.

Remark 5. The meta-base updates in (25) and (26) are quite versatile, as there are many options for constructing the surrogate (meta/base) loss, optimism, and the correction term. We remind that a feasible construction must adhere to the feedback model — in the multi-gradient feedback model, the entire gradient function $\nabla f_t(\cdot)$ is available, while in the one-gradient feedback model, only the gradient $\nabla f_t(\mathbf{x}_t)$ is available to the player. In Section 5.2, we will present several concrete instantiations of the generic algorithmic template, including the proposed Sword and Sword++ in the earlier subsections. ¶

5.2 Instantiations

In this part, we present three instantiations of the generic algorithmic template, including Sword, Sword++, and another important instantiation, which we refer to as Sword.optimism. For clarity, we provide a summary of these instantiations in Table 1.

Recovering Sword. We instantiate the algorithmic template as follows: setting

- base surrogate loss as $g_t^{\mathtt{base}}(\mathbf{x}) = f_t(\mathbf{x})$ and base optimism as $h_t^{\mathtt{base}}(\mathbf{x}) = f_{t-1}(\mathbf{x})$;
- meta surrogate loss as $g_t^{\mathtt{meta}}(\mathbf{x}) = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle$ and meta optimism as $h_t^{\mathtt{meta}}(\mathbf{x}) = \langle \nabla f_{t-1}(\bar{\mathbf{x}}_t), \mathbf{x}\rangle$, as well as correction term as $\boldsymbol{c}_t = \mathbf{0}$.

Then, the template updates in the following way: the base-algorithm updates by

$$\mathbf{x}_{t,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_{t-1}(\mathbf{x}_{t-1,i})\right], \quad \widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_{t,i})\right],$$

and the meta-algorithm updates by

$$p_{t+1,i} \propto \exp\left(-\varepsilon\Big(\sum_{s=1}^{t} \langle \nabla f_s(\mathbf{x}_s), \mathbf{x}_{s,i}\rangle + \langle \nabla f_t(\bar{\mathbf{x}}_{t+1}), \mathbf{x}_{t,i}\rangle\Big)\right).$$

The update procedures precisely recover Sword as presented in Algorithms 1 and 2. Note that in Sword, there is no correction terms, since the gradient-variation dynamic regret bound is attained by guaranteeing gradient-variation meta-regret for the meta-algorithm and gradient-variation base-regret for the base-algorithm, respectively.

Recovering Sword++. We instantiate the algorithmic template as follows: setting

- base surrogate loss as $g_t^{\texttt{base}}(\mathbf{x}) = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle$ and base optimism as $h_t^{\texttt{base}}(\mathbf{x}) = \langle \nabla f_{t-1}(\mathbf{x}_{t-1}), \mathbf{x} \rangle$;
- meta surrogate loss as $g_t^{\texttt{meta}}(\mathbf{x}) = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle$ and meta optimism as $h_t^{\texttt{meta}}(\mathbf{x}) = \langle \nabla f_{t-1}(\mathbf{x}_{t-1}), \mathbf{x} \rangle$, as well as correction term $\boldsymbol{c}_t$ as $c_{t,i} = \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$ with $\mathbf{x}_{0,1} = \mathbf{0}$.

Then, the template updates in the following way: the base-algorithm updates by

$$\mathbf{x}_{t,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_{t-1}(\mathbf{x}_{t-1})\right], \quad \widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_t)\right],$$

and the meta-algorithm updates by

$$p_{t+1,i} \propto \exp\left(-\varepsilon\Big(\sum_{s=1}^{t} \langle \nabla f_s(\mathbf{x}_s), \mathbf{x}_{s,i} \rangle + \lambda \sum_{s=1}^{t+1} \|\mathbf{x}_{s,i} - \mathbf{x}_{s-1,i}\|_2^2 + \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t+1,i} \rangle \Big) \right).$$

The update procedures precisely correspond to Sword++ as presented in Algorithms 3 and 4. We emphasize once more that the algorithmic updates only necessitate querying the gradient $\nabla f_t(\mathbf{x}_t)$ at each round $t \in [T]$.

Another important instantiation. We further present another instantiation of the template that can be of independent interest. The resulting algorithm can achieve an optimistic dynamic regret bound of order $\mathcal{O}(\sqrt{A_T(1+P_T)})$, where $A_T = \sum_{t=1}^{T} \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2$ measures the quality of the optimistic vectors $\{M_t\}_{t=1}^{T}$. We instantiate the algorithmic template as follows: setting

- base surrogate loss as $g_t^{\texttt{base}}(\mathbf{x}) = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle$ and base optimism as $h_t^{\texttt{base}}(\mathbf{x}) = \langle M_t, \mathbf{x} \rangle$;
- meta surrogate loss as $g_t^{\texttt{meta}}(\mathbf{x}) = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle$ and meta optimism as $h_t^{\texttt{meta}}(\mathbf{x}) = \langle M_t, \mathbf{x} \rangle$, as well as correction term $\boldsymbol{c}_t$ as $c_{t,i} = \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$ with $\mathbf{x}_{0,1} = \mathbf{0}$.

Then, the template updates in the following way: the base-algorithm updates by

$$\mathbf{x}_{t,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i M_t\right], \quad \widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_t)\right], \tag{28}$$

and the meta-algorithm updates by

$$p_{t+1,i} \propto \exp\left(-\varepsilon\Big(\sum_{s=1}^{t} \langle \nabla f_s(\mathbf{x}_s), \mathbf{x}_{s,i} \rangle + \lambda \sum_{s=1}^{t+1} \|\mathbf{x}_{s,i} - \mathbf{x}_{s-1,i}\|_2^2 + \langle M_{t+1}, \mathbf{x}_{t+1,i} \rangle \Big) \right). \tag{29}$$

We refer to the above meta-base updates, (28) and (29), as Sword.optimism. Its dynamic regret analysis detailed in Section 5.3. Notice that by setting the optimism as the last-round gradient, specifically, $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$, Sword.optimism recovers Sword++ exactly.

5.3 Theoretical Guarantee

In this part, we present the dynamic regret analysis for Sword.optimism, which is arguably the most general instantiation of the collaborative online ensemble template. It is straightforward to extend Theorem 5 for the generic template presented in Section 5.1, specifically, the meta-base updates in (25) and (26). However, the various variables in the generic template may somewhat obscure the core ideas. Therefore, we choose to showcase the dynamic regret analysis for Sword.optimism, as its analysis effectively captures the essence and its algorithm is also sufficient general (for instance, it can specialize Sword++).

Theorem 5. *Under Assumptions 1 and 2, set the pool of candidate step sizes $\mathcal{H}$ as*

$$\mathcal{H} = \left\{ \eta_i = \min\left\{ \bar{\eta}, \sqrt{\frac{D^2}{8G^2T}} \cdot 2^{i-1} \right\} \mid i \in [N] \right\}, \tag{30}$$

where $N = \lceil 2^{-1}\log_2((8G^2T\bar{\eta}^2)/D^2)\rceil + 1$ is the number of candidate step sizes; further set the learning rate of the meta-algorithm as

$$\varepsilon = \min\left\{ \bar{\varepsilon}, \sqrt{\frac{\ln N}{D^2 \sum_{t=1}^T \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2}} \right\}. \tag{31}$$

Then, decisions returned by Sword.optimism (namely, meta-algorithm as (29) and base-algorithm as (28)) satisfy that for any comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$,

$$\begin{aligned} \sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) \le{}& 2\sqrt{D^2(\ln N)A_T} + 2\sqrt{(D^2 + 2DP_T)A_T} \\ &+ \frac{2\ln N}{\bar{\varepsilon}} + \frac{2(D^2 + 2DP_T)}{\bar{\eta}} + \left(\lambda - \frac{1}{4\bar{\eta}}\right) S_{x,i} - \frac{1}{4\bar{\varepsilon}} S_p - \lambda S_{\text{mix}}. \end{aligned} \tag{32}$$

In above, $A_T = \sum_{t=1}^T \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2$ is the adaptivity term measuring the quality of optimistic gradient vectors $\{M_t\}_{t=1}^T$, and $P_T = \sum_{t=2}^T \|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2$ is the path length of comparators. The terms $S_{x,i} = \sum_{t=2}^T \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$, $S_p = \sum_{t=2}^T \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2$, and $S_{\text{mix}} = \sum_{t=2}^T \sum_{i=1}^N p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$ measures the stability of the decisions returned by the base-algorithm, meta-algorithm, and overall algorithm, respectively.

The proof of Theorem 5 is in Section 7.5. Notice that by setting the correction coefficient $\lambda = 0$ and setting clipped parameters $\bar{\eta}$ and $\bar{\varepsilon}$ as appropriate constants, Theorem 5 directly implies an $\mathcal{O}(\sqrt{A_T(1 + P_T)})$ dynamic regret for Sword.optimism.

As aforementioned, when setting the optimism as $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$, Sword.optimism recovers Sword++. Consequently, Theorem 5 serves as a preliminary analysis for Sword++ by substituting $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$ in the upper bound (32). By further combining the analysis of (23) in Remark 4, we can then prove the gradient-variation bound of Theorem 4, see the detailed argument in Section 7.4. The key element is to effectively cancel out the additional positive term using negative terms and correction terms jointly, which are strategically introduced due to the collaboration between the meta and base levels.

6. Implication, Significance, and Lower Bound

In this section, we present several additional results, including the implication to small-loss dynamic regret, the implication to the worst-case dynamic regret, the significance of problem-dependent bounds, and a lower bound.

6.1 Implication to Small-Loss Dynamic Regret

In this part, we investigate another problem-dependent quantity — the cumulative loss of comparators defined as $F_T = \sum_{t=1}^T f_t(\mathbf{u}_t)$.

In the conference version, we propose the Sword algorithm (presented in Section 4.2) to achieve the gradient-variation dynamic regret, and then propose a variant to attain the small-loss bound, which employs OGD as the base-algorithm and uses the vanilla Hedge with linearized surrogate loss as the meta-algorithm (i.e., choosing the optimistic vector $M_t = \mathbf{0}$ for both meta- and base-algorithms). In the current paper, we demonstrate that the improved algorithm Sword++ designed in Section 4.3 itself provably achieves the small-loss dynamic regret *without* any algorithmic modification. In fact, we have the following theorem regarding the small-loss bound of Sword++, whose proof is in Section 7.6.

Theorem 6. *Set the parameters the same as those in Theorem 4. Under Assumptions 1, 2, 3, and 4, Sword++ satisfies that*

$$\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) \leq \mathcal{O}\left(\sqrt{(1+P_T+F_T)(1+P_T)}\right),$$

and hence achieves the best-of-both-worlds guarantee:

$$\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) \leq \mathcal{O}\left(\sqrt{(1+P_T+\min\{V_T, F_T\})(1+P_T)}\right).$$

The bounds hold for any comparator sequence $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$. In above, $F_T = \sum_{t=1}^T f_t(\mathbf{u}_t)$ is the cumulative loss of comparators, $V_T = \sum_{t=2}^T \sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$ is the gradient variation, and $P_T = \sum_{t=2}^T \|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2$ is the path length.

Comparing with Theorem 4, one more assumption (Assumption 4) is required. This non-negativity assumption is a precondition for establishing the self-bounding property for convex and smooth functions (Srebro et al., 2010), and thus is commonly used in the small-loss analysis of online learning and stochastic optimization (Srebro et al., 2010; Cotter et al., 2011; Zhang et al., 2013, 2019; Zhang and Zhou, 2019).

Remark 6. Our conference version (Zhao et al., 2020b) achieves the best-of-both-worlds bound in a different way, in which we use a heterogeneous model selection method of learning an optimism (Rakhlin and Sridharan, 2013) since different optimistic vectors are used for the small-loss and gradient-variation bounds. As such, three algorithms (Sword$_{\text{var}}$, Sword$_{\text{small}}$, and Sword$_{\text{best}}$) are designed to achieve the three different bounds (gradient-variation, small-loss, and best-of-both-worlds bounds) respectively. By contrast, Theorem 6 indicates that the Sword++ algorithm can achieve *all* the three problem-dependent dynamic regret bounds without any modifications, owing to its one-gradient query complexity property. ¶

Remark 7. Comparing to the $\mathcal{O}(\sqrt{T(1+P_T)})$ minimax rate, Theorem 6 replaces the dependence on T by the problem-dependent quantity $P_T + \min\{V_T, F_T\}$ and thus achieves dual adaptivity in terms of both gradient variation V_T and the small-loss quantity F_T. Furthermore, one may wonder whether it is possible to replace T by $\min\{V_T, F_T\}$ only. This requires a lower bound argument and we only have a partial answer. Specifically, we prove an $\Omega(P_T)$ lower bound (see Theorem 8 in Section 6.4) by constructing a problem instance via probabilistic methods, in which the small-loss quantity satisfies $F_T = 0$. As a result, an $\mathcal{O}(\sqrt{(1+F_T)(1+P_T)})$ upper bound will contradict with the lower bound, hence eliminating the general possibility of attaining this result. Nevertheless, we fail to provide a similar reasoning for the gradient-variation bound. Indeed, we have the following conjectures on the tightness of the gradient-variation dynamic regret. For the multi-gradient feedback model, we lean to believe that the $\mathcal{O}(\sqrt{(1+P_T+V_T)(1+P_T)})$ rate is *not* optimal but the optimal one might be $\mathcal{O}(\sqrt{(1+V_T)(1+P_T)})$. For the one-gradient feedback model, we conjecture that our obtained rate is already optimal. Note that the latter setup is actually the one we are mostly concerned with, and we will study the optimality in the future. ¶

6.2 Implication to Worst-Case Dynamic Regret

In this part, we present the implication of the universal dynamic regret to the worst-case dynamic regret. As discussed in Section 2.2, there are two kinds of worst-case dynamic regret bounds, with different regularities: the path-length bound with $P_T^* = \sum_{t=2}^T \|\mathbf{x}_{t-1}^* - \mathbf{x}_t^*\|_2$ and the function-variation bound with $V_T^f = \sum_{t=2}^T \sup_{\mathbf{x}\in\mathcal{X}} |f_{t-1}(\mathbf{x}) - f_t(\mathbf{x})|$. The following theorem provides a unified reduction to both of them.

Theorem 7. *Let $A_T \in \mathbb{R}_+$ be a certain adaptivity term. Suppose there exists an algorithm $\mathcal{A}$ that guarantees*

$$\text{D-Regret}_T(\mathbf{u}_1, \ldots, \mathbf{u}_T) \leq \sqrt{A_T(D + P_T)}, \tag{33}$$

for any comparator sequence $\mathbf{u}_1, \ldots \mathbf{u}_T \in \mathcal{X}$ with path length $P_T = \sum_{t=2}^T \|\mathbf{u}_t - \mathbf{u}_{t-1}\|_2$, then the algorithm $\mathcal{A}$ enjoys the following worst-case dynamic regret bounds:

$$\text{D-Regret}_T(\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*) \leq 3\sqrt{DA_T} + \min\left\{\sqrt{A_T P_T^*}, 5D^{1/3}T^{1/3}A_T^{1/3}(V_T^f)^{1/3}\right\}. \tag{34}$$

Theorem 7 demonstrates that an $\mathcal{O}(\sqrt{A_T(1+P_T)})$ universal dynamic regret bound can *directly* imply an $\mathcal{O}(\sqrt{A_T} + \min\{\sqrt{A_T P_T^*}, (TA_T V_T^f)^{1/3}\})$ worst-case dynamic regret bound. A typical choice of this adaptivity term is $A_T = \sum_{t=1}^T \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2$ that measures the quality of optimistic gradient vectors $\{M_t\}_{t=1}^T$. Then, the implication matches the best-known optimistic worst-case dynamic regret bound presented in (Jadbabaie et al., 2015; Zhang et al., 2020b), taking the best of the path-length and function-variation regularities. It is worth noting that Jadbabaie et al. (2015) achieve this result through a novel and sophisticated doubling trick scheme, which will introduce a potentially non-convex inner optimization $\sup_{\mathbf{x}\in\mathcal{X}}|f_t(\mathbf{x}) - f_{t-1}(\mathbf{x})|$ at iteration $t \in [T]$. In contrast, our Theorem 7 demonstrates that when the algorithm achieves an $\mathcal{O}(\sqrt{A_T(1+P_T)})$ universal dynamic regret, it automatically obtains the desired worst-case dynamic regret bounds. Notably, our proposed Sword.optimism algorithm (see the last instantiation in Section 5.2) already satisfies this requirement using the collaborative online ensemble framework.

The proof of Theorem 7 can be found in Section 7.7. Given the universal dynamic regret bound (33), one can immediately derive an $\mathcal{O}(\sqrt{A_T} + \sqrt{A_T P_T^*})$ worst-case path-length bound by setting $\mathbf{u}_t = \mathbf{x}_t^*$ for any $t \in [T]$, but it is less straightforward to obtain the $\mathcal{O}(\sqrt{A_T} + T^{1/3} A_T^{1/3} (V_T^f)^{1/3})$ function-variation bound. To achieve so, we need to introduce a reference comparator sequence that exhibits piecewise-stationary behavior. The desired function-variation bound is then achievable by optimally tuning the stationary length of the sequence during the analysis. The idea was introduced in Zhang et al. (2020b, Appendix A.2), but an explicit reduction was not provided. We offer a clear presentation of the results.

Moreover, in Theorem 7, we focus on the $\mathcal{O}(\sqrt{A_T(1+P_T)})$ universal dynamic regret bound, which incorporates the general adaptivity term A_T. Employing a similar proof methodology, we can also convert the gradient-variation/small-loss universal dynamic regret bounds, attained by Sword and Sword++, into the context of worst-case dynamic regret. We omit the details, as the universal dynamic regret has already been established as a more reasonable performance measure for non-stationary online convex optimization.

6.3 Significance of Problem-Dependent Bounds

In this part, we justify the significance of our problem-dependent dynamic regret bounds. We present two concrete problem instances to demonstrate that it is possible to achieve a *constant* dynamic regret bound instead of the minimax rate $\mathcal{O}(\sqrt{T(1+P_T)})$ by exploiting the problem's structure.

We consider the quadratic loss function of the form $f_t(x) = \frac{1}{2}(a_t \cdot x - b_t)^2$, where $a_t \neq 0$ and $x \in \mathcal{X} = [-1, 1]$. Clearly, the online function $f_t : \mathbb{R} \mapsto \mathbb{R}$ is convex and smooth. Denote by T the time horizon. The coefficients a_t and b_t will be specified below in each instance.

Instance 1 ($V_T \ll F_T$)**.** Let the time horizon $T = 2K+1$ be odd with $K > 2$. We set the coefficients $a_t = 0.5 - \frac{t-1}{T}$ and $b_t = 1$ for all $t \in [T]$.

We set the comparator u_t to be the minimizer of f_t, i.e, $u_t = x_t^* = \arg\min_{x \in \mathcal{X}} f_t(x)$. Clearly, $u_t = 1$ for $t \in [K+1]$, and $u_t = -1$ for $t = K+2, \ldots, T$. A direct calculation shows

$$\begin{aligned} V_T &= \sum_{t=2}^{T} \sup_{x \in \mathcal{X}} |(a_{t-1}^2 - a_t^2)x - (a_{t-1} - a_t)|^2 = \sum_{t=2}^{T} \sup_{x \in \mathcal{X}} \left| \left(\frac{T - 2t + 3}{T^2} \right) \cdot x - \frac{1}{T} \right|^2 \\ &= \sum_{t=2}^{K+2} \left(\frac{2T - (2t-3)}{T^2} \right)^2 + \sum_{t=K+3}^{T} \left(\frac{2t-3}{T^2} \right)^2 \leq \sum_{t=2}^{T} \left(\frac{2}{T} \right)^2 = \mathcal{O}(1). \end{aligned}$$

$$F_T = \sum_{t=1}^{T} \frac{1}{2}(a_t u_t - b_t)^2 = \sum_{t=1}^{K+1} \frac{1}{2} \left(0.5 - \frac{t-1}{T} - 1 \right)^2 + \sum_{t=K+2}^{T} \frac{1}{2} \left(-0.5 + \frac{t-1}{T} - 1 \right)^2 = \Theta(T).$$

We can observe that the gradient variation $V_T = \mathcal{O}(1)$ is significantly smaller than the small-loss quantity $F_T = \Theta(T)$ (as well as the problem-independent quantity T) in this instance; and meanwhile, the path length is $P_T = \mathcal{O}(1)$. Then, the minimax dynamic regret bound is $\mathcal{O}(\sqrt{T(1+P_T)}) = \mathcal{O}(\sqrt{T})$; the small-loss bound is $\mathcal{O}(\sqrt{(1+P_T+F_T)(1+P_T)}) = \mathcal{O}(\sqrt{T})$;

and the gradient-variation bound is $\mathcal{O}(\sqrt{(1+P_T+V_T)(1+P_T)}) = \mathcal{O}(1)$. As a result, by exploiting the problem's structure, Sword++ can enjoy a *constant* dynamic regret in this scenario, significantly improving upon the problem-independent bound of order $\mathcal{O}(\sqrt{T})$.

Instance 2 ($F_T \ll V_T$)**.** Let the time horizon $T = 2K$ be even. During the first half iterations, (a_t, b_t) is set as $(1,1)$ on odd rounds and $(0.5, 0.5)$ on even rounds. During the remaining iterations, (a_t, b_t) is set as $(1,-1)$ on odd rounds and $(0.5, -0.5)$ on even rounds.

We set the comparator u_t to be the minimizer of f_t, i.e, $u_t = x_t^* = \arg\min_{x \in \mathcal{X}} f_t(x)$. Clearly, $u_t = 1$ for $t \in [K]$, and $u_t = -1$ for $t = K+1, \ldots, T$. A direct calculation shows

$$V_T = \sum_{t=2}^{T} \sup_{x \in \mathcal{X}} |(a_{t-1}^2 - a_t^2)x - (a_{t-1}b_{t-1} - a_t b_t)|^2 = \Theta(T), \qquad F_T = 0.$$

We can see that the small-loss quantity $F_T = 0$ is considerably smaller than the gradient variation $V_T = \Theta(T)$ (as well as the problem-independent quantity T) in this scenario; and meanwhile, the path length is $P_T = \mathcal{O}(1)$. Then, the minimax dynamic regret bound is $\mathcal{O}(\sqrt{T(1+P_T)}) = \mathcal{O}(\sqrt{T})$; the gradient-variation bound is $\mathcal{O}(\sqrt{(1+P_T+V_T)(1+P_T)}) = \mathcal{O}(\sqrt{T})$; and the small-loss bound is $\mathcal{O}(\sqrt{(1+P_T+F_T)(1+P_T)}) = \mathcal{O}(1)$. As a result, by exploiting the problem's structure, Sword++ can enjoy a *constant* dynamic regret in this scenario, significantly improving upon the problem-independent bound of order $\mathcal{O}(\sqrt{T})$.

6.4 A Lower Bound

We here present a lower bound for dynamic regret of convex and smooth functions.

Theorem 8. *For any online algorithm $\mathcal{A}$, there always exists a sequence of convex and smooth functions $f_1, \ldots, f_T$ and a sequence of comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T$, such that*

$$\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) = \Omega(P_T(\mathbf{u}_1, \ldots, \mathbf{u}_T)). \tag{35}$$

The proof of Theorem 8 can be found in Section 7.8. The theorem is proved by the probabilistic method, and in the constructed problem instance, the small-loss quantity is $F_T = \sum_{t=1}^{T} f_t(\mathbf{u}_t) = 0$. As a result, an $\mathcal{O}(\sqrt{(1+F_T)(1+P_T)})$ upper bound would contradict with the $\Omega(P_T)$ lower bound, as one can verify that $\mathcal{O}(\sqrt{(1+F_T)(1+P_T)}) = \mathcal{O}(\sqrt{P_T})$, which can be smaller than the $\Omega(P_T)$ lower bound. Hence, the lower bound eliminates the general possibility of attaining a better small-loss dynamic regret. Nevertheless, as the gradient variation $V_T = \sum_{t=2}^{T} \sup_{\mathbf{x} \in \mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$ is larger than 0 in this instance, we cannot rule out the possibility of the $\mathcal{O}(\sqrt{(1+V_T)(1+P_T)})$ upper bound.

7. Proofs

This section presents the proofs of main results, including Theorem 2 and Theorem 3 of Section 4.2, as well as Theorem 4 and Theorem 6 of Section 4.3.

7.1 Proof of Theorem 1

Proof The instantaneous dynamic regret can be decomposed in the following way:

$$\begin{aligned} f_t(\mathbf{x}_t) - f_t(\mathbf{u}_t) &\leq \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{u}_t \rangle \\ &= \underbrace{\langle \nabla f_t(\mathbf{x}_t) - M_t, \mathbf{x}_t - \widehat{\mathbf{x}}_{t+1} \rangle}_{\texttt{term (a)}} + \underbrace{\langle M_t, \mathbf{x}_t - \widehat{\mathbf{x}}_{t+1} \rangle}_{\texttt{term (b)}} + \underbrace{\langle \nabla f_t(\mathbf{x}_t), \widehat{\mathbf{x}}_{t+1} - \mathbf{u}_t \rangle}_{\texttt{term (c)}}. \end{aligned}$$

In the following, we will bound the three terms respectively. In brief, we use the stability lemma (Lemma 5) to bound term (a) and appeal to the Bregman proximal inequality (Lemma 4) to bound term (b) and term (c). Below we present the precise arguments.

We first investigate term (a). Intuitively, the prediction $\mathbf{x}_t$ should be close the $\widehat{\mathbf{x}}_{t+1}$ when the optimistic vector M_t is close to the gradient of the next iteration $\nabla f_t(\mathbf{x}_t)$. The intuition is formalized in the stability lemma (Chiang et al., 2012, Propostion 7), as restated in Lemma 5 of Appendix C. Indeed, the stability lemma implies $\|\mathbf{x}_t - \widehat{\mathbf{x}}_{t+1}\| \leq \eta_t \|\nabla f_t(\mathbf{x}_t) - M_t\|_*$ and consequently,

$$\begin{aligned} \texttt{term (a)} &= \langle \nabla f_t(\mathbf{x}_t) - M_t, \mathbf{x}_t - \widehat{\mathbf{x}}_{t+1} \rangle \\ &\leq \|\nabla f_t(\mathbf{x}_t) - M_t\|_* \|\mathbf{x}_t - \widehat{\mathbf{x}}_{t+1}\| \leq \eta_t \|\nabla f_t(\mathbf{x}_t) - M_t\|_*^2. \end{aligned}$$

We now analyze term (b) and term (c). By the Bregman proximal inequality (Lemma 4) and the OMD update step $\mathbf{x}_t = \arg\min_{\mathbf{x}\in\mathcal{X}} \eta_t \langle M_t, \mathbf{x} \rangle + \mathcal{D}_\psi(\mathbf{x}, \widehat{\mathbf{x}}_t)$, we have

$$\texttt{term (b)} = \langle M_t, \mathbf{x}_t - \widehat{\mathbf{x}}_{t+1} \rangle \leq \frac{1}{\eta_t}\Big(\mathcal{D}_\psi(\widehat{\mathbf{x}}_{t+1}, \widehat{\mathbf{x}}_t) - \mathcal{D}_\psi(\widehat{\mathbf{x}}_{t+1}, \mathbf{x}_t) - \mathcal{D}_\psi(\mathbf{x}_t, \widehat{\mathbf{x}}_t)\Big).$$

Similarly, the OMD update step $\widehat{\mathbf{x}}_{t+1} = \arg\min_{\mathbf{x}\in\mathcal{X}} \eta_t \langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle + \mathcal{D}_\psi(\mathbf{x}, \widehat{\mathbf{x}}_t)$ implies

$$\texttt{term (c)} = \langle \nabla f_t(\mathbf{x}_t), \widehat{\mathbf{x}}_{t+1} - \mathbf{u}_t \rangle \leq \frac{1}{\eta_t}\Big(\mathcal{D}_\psi(\mathbf{u}_t, \widehat{\mathbf{x}}_t) - \mathcal{D}_\psi(\mathbf{u}_t, \widehat{\mathbf{x}}_{t+1}) - \mathcal{D}_\psi(\widehat{\mathbf{x}}_{t+1}, \widehat{\mathbf{x}}_t)\Big).$$

Combining the three upper bounds completes the proof. ■

7.2 Proof of Theorem 2

Proof Recall the definitions of feedback loss $\boldsymbol{\ell}_t \in \mathbb{R}^N$ in (14) and optimism $\boldsymbol{m}_t \in \mathbb{R}^N$ in (15), which are restated below for ease of reading: $\ell_{t,i} = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t,i} \rangle$ and $m_{t,i} = \langle \nabla f_{t-1}(\bar{\mathbf{x}}_t), \mathbf{x}_{t,i} \rangle$ where $\bar{\mathbf{x}}_t = \sum_{i=1}^N p_{t-1,i}\mathbf{x}_{t,i}$. Substituting them into Lemma 2 yields

$$\sum_{t=1}^T \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_{t,i} \rangle \leq \varepsilon \sum_{t=1}^T \|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2 + \frac{\ln N}{\varepsilon} - \frac{1}{4\varepsilon}\sum_{t=2}^T \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2. \tag{36}$$

The adaptivity term can be further upper bounded as follows:

$$\begin{aligned} \|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2 &\overset{(15)}{=} \max\nolimits_{i\in[N]} \langle \nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\bar{\mathbf{x}}_t), \mathbf{x}_{t,i} \rangle^2 \\ &\leq D^2 \|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\bar{\mathbf{x}}_t)\|_2^2 \\ &\leq 2D^2 (\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_t)\|_2^2 + \|\nabla f_{t-1}(\mathbf{x}_t) - \nabla f_{t-1}(\bar{\mathbf{x}}_t)\|_2^2) \\ &\leq 2D^2 \sup\nolimits_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2 + 2D^2L^2\|\mathbf{x}_t - \bar{\mathbf{x}}_t\|_2^2 \\ &\leq 2D^2 \sup\nolimits_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2 + 2D^4L^2\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2, \end{aligned} \tag{37}$$

where the inequality (37) holds because of the following fact

$$\|\mathbf{x}_t - \bar{\mathbf{x}}_t\|_2^2 = \Big\| \sum_{i=1}^N (p_{t,i} - p_{t-1,i})\mathbf{x}_{t,i} \Big\|_2^2 \leq \Big(\sum_{i=1}^N |p_{t,i} - p_{t-1,i}| \|\mathbf{x}_{t,i}\|_2 \Big)^2 \leq D^2 \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2.$$

Substituting (37) into (36) and exploiting the boundedness of the gradient norm, we get

$$\sum_{t=1}^T \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_{t,i} \rangle \leq 2\varepsilon D^2 (G^2 + V_T) + \frac{\ln N}{\varepsilon} + \left(2D^4 L^2 \varepsilon - \frac{1}{4\varepsilon} \right) \sum_{t=2}^T \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2.$$

By setting $\varepsilon = \min\{1/(4D^2 L), \sqrt{(\ln N)/(2D^2(G^2 + V_T))}\}$, by Lemma 7, we obtain

$$\sum_{t=1}^T \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_{t,i} \rangle \leq 2D\sqrt{2(G^2 + V_T) \ln N} + 8D^2 L \ln N,$$

which completes the proof as $f_t(\mathbf{x}_t) - f_t(\mathbf{x}_{t,i}) \leq \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_{t,i} \rangle$ holds for any $t \in [T]$ due to the convexity of loss functions. ■

7.3 Proof of Theorem 3

Proof As stated in (10), dynamic regret can be decomposed into two parts:

$$\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) \leq \underbrace{\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{x}_{t,i})}_{\texttt{meta-regret}} + \underbrace{\sum_{t=1}^T f_t(\mathbf{x}_{t,i}) - \sum_{t=1}^T f_t(\mathbf{u}_t)}_{\texttt{base-regret}}, \tag{38}$$

which holds for any base-learner's index $i \in [N]$. We now presents the upper bounds of the meta-regret and base-regret, respectively.

Upper bound of meta-regret. Theorem 2 shows that the meta-regret against any base-learner's index $i \in [N]$ can be bounded by

$$\texttt{meta-regret} = \sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{x}_{t,i}) \leq 2D\sqrt{2(4G^2 + V_T) \ln N} + 8D^2 L \ln N. \tag{39}$$

Upper bound of base-regret. Lemma 1 indicates that for any index $i \in [N]$, the dynamic regret of the base-learner is at most

$$\texttt{base-regret} = \sum_{t=1}^T f_t(\mathbf{x}_{t,i}) - \sum_{t=1}^T f_t(\mathbf{u}_t) \leq \eta_i (G^2 + 2V_T) + \frac{1}{2\eta_i}(D^2 + 2DP_T), \tag{40}$$

where $\eta_i \in \mathcal{H}$ is the step size associated with the i-th base-learner. Recall in Lemma 1, we require the step size $\eta_i \leq 1/(4L)$ to leverage the negative term in the regret analysis. Denote by $\eta^* = \sqrt{(D^2 + 2DP_T)/(G^2 + 2V_T)}$ the optimal step size without considering the constraint and by $\eta^\dagger = \min\{1/(4L), \eta^*\}$ the clipped one. Notice that we have

$\eta_1 = \sqrt{D^2/(8G^2T)}$, $\eta_N = 1/(4L)$, and $\eta_1 \le \eta^\dagger \le \eta_N$, due to path length $P_T = \sum_{t=2}^T \|\mathbf{u}_t - \mathbf{u}_{t-1}\|_2 \in [0, DT]$ and gradient variation $V_T = \sum_{t=2}^T \sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2 \le 4G^2(T-1)$ by Assumption 1 and Assumption 2. More importantly, owing to the construction of the step size pool $\mathcal{H}$ in (18), we can assure that there exists an index $i^* \in [N]$ such that $\eta_{i^*} \le \eta^\dagger \le \eta_{i^*+1} = 2\eta_{i^*}$. As a result, we pick $i = i^*$ in (40) and get

$$\begin{aligned}
\texttt{base-regret} &\le \eta_{i^*}(G^2 + 2V_T) + \frac{D^2 + 2DP_T}{2\eta_{i^*}} \\
&\le \eta^\dagger(G^2 + 2V_T) + \frac{D^2 + 2DP_T}{\eta^\dagger} \qquad (41) \\
&\le 2\sqrt{(G^2 + 2V_T)(D^2 + 2DP_T)} + 8L(D^2 + 2DP_T) \qquad (42) \\
&\le \mathcal{O}\Big(\sqrt{(1 + P_T + V_T)(1 + P_T)}\Big). \qquad (43)
\end{aligned}$$

In above, (42) holds because $\eta^\dagger$ is either η^* or $1/(4L)$ and

- when $\eta^\dagger = \eta^*$, R.H.S of (41) $= 2\sqrt{(G^2 + 2V_T)(D^2 + 2DP_T)}$;
- when $\eta^\dagger = 1/(4L)$, we have $\eta^* = \sqrt{(D^2 + 2DP_T)/(G^2 + 2V_T)} \ge \frac{1}{4L}$, which implies that $\frac{1}{4L}(G^2 + V_T) \le 4L(D^2 + 2DP_T)$. Under such a case, R.H.S of (41) $= 4L(D^2 + 2DP_T) + \frac{1}{4L}(G^2 + 2V_T) \le 8L(D^2 + 2DP_T)$.

Combining the two upper bounds yields (42). Moreover, (43) follows from $\sqrt{a} + \sqrt{b} \le \sqrt{2(a+b)}$, $\forall a, b > 0$.

Upper bound of overall dynamic regret. Note that the meta-base regret decomposition (38), meta-regret upper bound (39), and base-regret upper bound (40) all hold for any index $i \in [N]$. Hence, we can choose the index as i^* as specified above and get

$$\begin{aligned}
&\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) \\
&\le 2D\sqrt{2(G^2 + V_T)\ln N} + 8D^2 L \ln N + \eta_{i^*}(G^2 + 2V_T) + \frac{D^2 + 2DP_T}{2\eta_{i^*}} \\
&\le \mathcal{O}(\sqrt{1 + V_T}) + \mathcal{O}(\sqrt{(1 + P_T + V_T)(1 + P_T)}) \\
&= \mathcal{O}\Big(\sqrt{(1 + P_T + V_T)(1 + P_T)}\Big),
\end{aligned}$$

where the second inequality is by (43). Hence, we complete the proof of Theorem 3. ■

7.4 Proof of Theorem 4

Proof It is important to note that Sword++ is essentially an instantiation of the unified online ensemble framework presented in Section 5, as specified in Section 5.2. Therefore, for simplicity, we will prove the dynamic regret of Sword++ by building upon the general theorem for the collaborative online ensemble (namely, Theorem 5).

Indeed, we can obtain a dynamic regret upper bound of Sword++ by substituting $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$ into (32) of Theorem 5. And we focus on the adaptivity term A_T, which can be further expanded as

$$\begin{aligned}
A_T &\leq G^2 + \sum_{t=2}^{T} \|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2 \\
&\leq G^2 + 2\sum_{t=2}^{T} \|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_t)\|_2^2 + 2\sum_{t=2}^{T} \|\nabla f_{t-1}(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2 \\
&\leq G^2 + 2\sup_{\mathbf{x}\in\mathcal{X}} \sum_{t=2}^{T} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2 + 2L^2 \sum_{t=2}^{T} \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 \\
&\leq G^2 + 2V_T + 4L^2 S_{\mathrm{mix}} + 4D^2L^2 S_p, \qquad (44)
\end{aligned}$$

where the third inequality is by smoothness of online functions and the last inequality holds by the same argument of deriving (23) and definitions of V_T, S_{mix} and S_p. As a result, the first term of (32) can be further bounded by

$$\begin{aligned}
& 2\sqrt{D^2(\ln N)A_T} \\
\leq{}& 2\sqrt{D^2(\ln N)\left(G^2 + 2V_T + 4L^2 S_{\mathrm{mix}} + 4D^2L^2 S_p\right)} \\
\leq{}& 2\sqrt{D^2(\ln N)\left(G^2 + 2V_T\right)} + 2\sqrt{D^2(\ln N)(4L^2 S_{\mathrm{mix}} + 4D^2L^2 S_p)} \\
\leq{}& 2\sqrt{D^2(\ln N)\left(G^2 + 2V_T\right)} + \frac{2\ln N}{\bar{\varepsilon}} + 8\bar{\varepsilon}D^2L^2 S_{\mathrm{mix}} + 8\bar{\varepsilon}D^4L^2 S_p, \qquad (45)
\end{aligned}$$

where the second inequality is due to the fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for any $a, b > 0$ and the last inequality is a consequence of the AM-GM inequality. Using a similar argument, we can bound the second term of (32) by

$$\begin{aligned}
& 2\sqrt{(D^2 + 2DP_T)A_T} \\
\leq{}& 2\sqrt{(D^2 + 2DP_T)(G^2 + 2V_T + 4L^2 S_{\mathrm{mix}} + 4D^2L^2 S_p)} \\
\leq{}& 2\sqrt{(D^2 + 2DP_T)(G^2 + 2V_T)} + \frac{2D^2 + 4DP_T}{\bar{\eta}} + 8\bar{\eta}L^2 S_{\mathrm{mix}} + 8\bar{\eta}D^2L^2 S_p. \qquad (46)
\end{aligned}$$

Plugging (45) and (46) into (32), we get the following dynamic regret bound,

$$\begin{aligned}
& \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \\
\leq{}& 2\sqrt{\ln N\left(G^2D^2 + 2D^2V_T\right)} + 2\sqrt{(D^2 + 2DP_T)(G^2 + 2V_T)} + \frac{4\ln N}{\bar{\varepsilon}} + \frac{4(D^2 + 2DP_T)}{\bar{\eta}} \\
& + \left(\lambda - \frac{1}{4\bar{\eta}}\right) S_{x,i} + \left(8\bar{\eta}D^2L^2 + 8\bar{\varepsilon}D^4L^2 - \frac{1}{4\bar{\varepsilon}}\right) S_p + \left(8\bar{\eta}L^2 + 8\bar{\varepsilon}D^2L^2 - \lambda\right) S_{\mathrm{mix}}.
\end{aligned}$$

We complete the proof by dropping the last three negative terms, which is guaranteed by the parameter configurations $\lambda = 2L$, $\bar{\eta} = 1/(8L)$ and $\bar{\varepsilon} = 1/(8D^2L)$. We finally mention that in above derivations, the term $\ln N = \mathcal{O}(\log\log T)$ is a double logarithmic factor in T, which is treated as a constant throughout the paper (see the statement at the end of Section 3.3) following previous studies (Adamskiy et al., 2012; Luo and Schapire, 2015).[2] ■

7.5 Proof of Theorem 5

Proof The proof shares the same spirit with that of Theorem 3, where we decompose the overall dynamic regret into the meta-regret and base-regret. The difference is that we now use a linearized surrogate loss function to substitute the original loss function. Indeed,

$$\begin{aligned}\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq {} & \sum_{t=1}^{T}\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{u}_t\rangle \\ = {} & \underbrace{\sum_{t=1}^{T}\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_{t,i}\rangle}_{\texttt{meta-regret}} - \underbrace{\sum_{t=1}^{T}\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t,i} - \mathbf{u}_t\rangle}_{\texttt{base-regret}}. \end{aligned} \tag{47}$$

Notably, the above meta-base regret decomposition holds for any base-learner's index $i \in [N]$. In the following, we upper bound these two terms respectively.

Upper bound of meta-regret. According to the definitions of the feedback loss $\boldsymbol{\ell}_t$ and the optimism $\boldsymbol{m}_t$, see the definition below (26), we can rewrite the meta-regret as

$$\begin{aligned}\texttt{meta-regret} = {} & \sum_{t=1}^{T}\sum_{i=1}^{N} p_{t,i}\cdot\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t,i}\rangle - \sum_{t=1}^{T}\langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t,i}\rangle \\ = {} & \sum_{t=1}^{T}\langle \boldsymbol{p}_t, \boldsymbol{\ell}_t\rangle - \sum_{t=1}^{T}\ell_{t,i} - \lambda\sum_{t=1}^{T}\sum_{i=1}^{N} p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 + \lambda\sum_{t=1}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2. \end{aligned}$$

As the meta-algorithm is an instance of OMD, we can exploit Lemma 2 to get that

$$\begin{aligned}\sum_{t=1}^{T}\langle \boldsymbol{p}_t, \boldsymbol{\ell}_t\rangle - \sum_{t=1}^{T}\ell_{t,i} \leq {} & \varepsilon\sum_{t=1}^{T}\|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2 + \frac{\ln N}{\varepsilon} - \frac{1}{4\varepsilon}\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2 \\ \leq {} & \varepsilon D^2\sum_{t=1}^{T}\|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2 + \frac{\ln N}{\varepsilon} - \frac{1}{4\varepsilon}\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2 \\ \leq {} & 2\sqrt{D^2(\ln N)A_T} + \frac{2\ln N}{\bar{\varepsilon}} - \frac{1}{4\bar{\varepsilon}}\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2, \end{aligned}$$

where $A_T = \sum_{t=1}^{T}\|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2$ is the adaptivity term measuring the quality of optimistic gradient vectors. The last inequality is true due to the setting of step size $\varepsilon =$

[2] Actually, this $\mathcal{O}(\log\log T)$ term can be improved to $\log\log P_T$ by imposing a non-uniform prior over the base-learners, and then can be strictly omitted in the $\mathcal{O}(\cdot)$-notation. We here omit the details and a similar argument can be found in (Zhang et al., 2018a, Section 4.3 Proof of Theorem 3).

$\min\{\bar{\varepsilon}, \sqrt{(\ln N)/(D^2 A_T)}\}$ and Lemma 7. Combining above two inequalities, we obtain

$$\begin{aligned}\texttt{meta-regret} \leq{} & 2\sqrt{D^2(\ln N)A_T} + \frac{2\ln N}{\bar{\varepsilon}} - \frac{1}{4\bar{\varepsilon}}\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2 \\ & - \lambda\sum_{t=1}^{T}\sum_{i=1}^{N} p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 + \lambda\sum_{t=1}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 \end{aligned} \tag{48}$$

Upper bound of base-regret. Since the base-algorithm can be seen as an instance of OMD running over linearized surrogate loss $g_t(\mathbf{x}) = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle$ with $\psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$, we can apply Lemma 1 to obtain the base-regret for any index $i \in [N]$ as

$$\begin{aligned}\texttt{base-regret} &= \sum_{t=1}^{T} g_t(\mathbf{x}_{t,i}) - \sum_{t=1}^{T} g_t(\mathbf{u}_t) \\ &\leq \eta_i \sum_{t=1}^{T}\|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2 + \frac{D^2 + 2DP_T}{2\eta_i} - \frac{1}{4\eta_i}\sum_{t=2}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 \\ &\leq \eta_i A_T + \frac{D^2 + 2DP_T}{2\eta_i} - \frac{1}{4\bar{\eta}}\sum_{t=2}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2, \end{aligned} \tag{49}$$

where the first inequality is a direct consequence of Lemma 1 with a fixed step size.

Upper bound of overall dynamic regret. Combining the meta-regret (48) and the base-regret (49) with (47), for any $i \in [N]$, we arrive at the following result:

$$\begin{aligned}&\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \\ &\leq 2\sqrt{D^2(\ln N)A_T} + \eta_i A_T + \frac{D^2 + 2DP_T}{2\eta_i} + \frac{2\ln N}{\bar{\varepsilon}} \\ &+ \left(\lambda - \frac{1}{4\bar{\eta}}\right)\sum_{t=2}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 - \frac{1}{4\bar{\varepsilon}}\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2 - \lambda\sum_{t=1}^{T}\sum_{i=1}^{N} p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2. \end{aligned} \tag{50}$$

Here, we remain to choose the best base-learner to make the term $\eta_i A_T + \frac{D^2+2DP_T}{2\eta_i}$ tightest possible. Note that the optimal step size is $\eta^* = \sqrt{(D^2 + 2DP_T)/A_T}$, but nevertheless, the step size we should identify is actually $\eta^\dagger = \{\eta^*, \bar{\eta}\}$ due to the threshold in the construction of the step size pool (30). Indeed, it can be verified that the candidate step sizes range from $\eta_1 = \sqrt{\frac{D^2}{8G^2T}}$ to $\eta_N = \bar{\eta}$. We discuss the upper bound in two cases.

- when $\eta^\dagger = \sqrt{(D^2 + 2DP_T)/A_T}$, the optimal step size η^* provably lies in the range of $\mathcal{H}$ and there must be an index i^* satisfying that $\eta_{i^*} \leq \eta^* \leq \eta_{i^*+1} = 2\eta_{i^*}$. Therefore, we will choose the compared index as $i = i^*$ and obtain that $\eta_{i^*} A_T + \frac{D^2+2DP_T}{2\eta_{i^*}} \leq \eta^* A_T + \frac{D^2+2DP_T}{\eta^*} = 2\sqrt{(D^2 + 2DP_T)A_T}$
- when $\eta^\dagger = \bar{\eta}$, we will choose the compared index as $i = N$ and obtain that $\eta_N A_T + \frac{D^2+2DP_T}{2\eta_N} = \bar{\eta} A_T + \frac{D^2+2DP_T}{2\bar{\eta}} \leq (2D^2 + 4DP_T)/\bar{\eta}$.

As a result, taking both cases into account yields

$$\begin{aligned}
&\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \\
&\le 2\sqrt{D^2 \ln N A_T} + 2\sqrt{(D^2 + 2DP_T)A_T} + \frac{2\ln N}{\bar{\varepsilon}} + \frac{2(D^2 + 2DP_T)}{\bar{\eta}} \\
&\qquad + \left(\lambda - \frac{1}{4\bar{\eta}}\right)\sum_{t=2}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 - \frac{1}{4\bar{\varepsilon}}\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2 - \lambda\sum_{t=1}^{T}\sum_{i=1}^{N} p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2,
\end{aligned}$$

which completes the proof. ■

7.6 Proof of Theorem 6

Proof The proof shares the same spirit as that of Theorem 4, whereas we upper bound the adaptivity term in a different way to achieve the small-loss bound. Specifically, we convert the adaptivity term to the cumulative loss of decisions defined by $F_T^X = \sum_{t=1}^T f_t(\mathbf{x}_t)$.

$$\begin{aligned}
A_T &= \|\nabla f_1(\mathbf{x}_1)\|_2^2 + \sum_{t=2}^{T}\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2 \\
&\le \|\nabla f_1(\mathbf{x}_1)\|_2^2 + 2\sum_{t=2}^{T}\|\nabla f_t(\mathbf{x}_t)\|_2^2 + 2\sum_{t=2}^{T}\|\nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2 \\
&\le 8L\sum_{t=1}^{T} f_t(\mathbf{x}_t) + 8L\sum_{t=2}^{T} f_{t-1}(\mathbf{x}_{t-1}) \le 16L\sum_{t=1}^{T} f_t(\mathbf{x}_t) = 16LF_T^X,
\end{aligned}$$

where the second inequality comes from the self-bounding property of smooth and non-negative functions as shown in Lemma 6. Then, a direct application of Theorem 5 with the parameter configurations $\lambda = 2L$, $\bar{\eta} = 1/(8L)$ and $\bar{\varepsilon} = 1/(8D^2L)$ indicates that the dynamic regret can be bounded by

$$\begin{aligned}
\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) &\le 2\sqrt{16LD^2 \ln N F_T^X} + 2\sqrt{16L(D^2 + 2DP_T)F_T^X} \\
&\qquad + 16D^2L\ln N + 16L(D^2 + 2DP_T).
\end{aligned}$$

It remains to convert the dynamic regret bound with respect to the cumulative loss of predictions F_T^X to that of the comparator $F_T = \sum_{t=1}^T f_t(\mathbf{u}_t)$. According to the definition of F_T and F_T^X, the above inequality implies that

$$\begin{aligned}
F_T^X - F_T &\le 2\sqrt{16L(D^2\ln N + D^2 + 2DP_T)F_T^X} + 16L(D^2\ln N + D^2 + 2DP_T) \\
&\le 2\sqrt{16L(D^2\ln N + D^2 + 2DP_T)(F_T + 16L(D^2\ln N + D^2 + 2DP_T))} \\
&\quad + 80L(D^2\ln N + D^2 + 2DP_T) \\
&= \mathcal{O}(\sqrt{(1 + P_T + F_T)(1 + P_T)}) + \mathcal{O}(1 + P_T) \\
&= \mathcal{O}(\sqrt{(1 + P_T + F_T)(1 + P_T)}), \qquad (51)
\end{aligned}$$

where the first inequality is due to the fact that $\sqrt{a}+\sqrt{b} \le \sqrt{2(a+b)}$ holds for any $a, b \ge 0$, and the second inequality comes from Lemma 9. We have completed the proof. ■

7.7 Proof of Theorem 7

Proof By choosing $\mathbf{u}_t = \mathbf{x}_t^*$ and the universal dynamic regret bound of $\text{D-Regret}_T(\mathbf{u}_1, \ldots, \mathbf{u}_T) \le \sqrt{A_T(D+P_T)}$, we can directly obtain

$$\text{D-Regret}_T(\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*) \le \sqrt{A_T(D+P_T^*)}, \tag{52}$$

which is the path-length worst-case dynamic regret bound.

In the following, we focus on the function-variation type bound. This is achieved by following the arguments of Zhang et al. (2020b), we introduce a *virtual* piece-wise stationary comparator sequence that only changes every $\Delta \in [1, T]$ iterations. Specifically, denoting by $\mathcal{I}_m = [(m-1)\Delta + 1, \min\{m\Delta, T\}] \subseteq [1, T]$ the m-th interval, we define the comparator over the interval $\mathcal{I}_m$ as $\mathbf{x}_{\mathcal{I}_m}^* = \arg\min_{\mathbf{x}\in\mathcal{X}} \sum_{t\in\mathcal{I}_m} f_t(\mathbf{x})$. There are in total $M = \lceil T/\Delta \rceil$ intervals. Then, we can decompose the worst-case dynamic regret as

$$\begin{aligned}\text{D-Regret}_T(\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*) &= \sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{x}_t^*)\\ &= \underbrace{\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{m=1}^M \sum_{t\in\mathcal{I}_m} f_t(\mathbf{x}_{\mathcal{I}_m}^*)}_{\texttt{term (a)}} + \underbrace{\sum_{m=1}^M \sum_{t\in\mathcal{I}_m} f_t(\mathbf{x}_{\mathcal{I}_m}^*) - \sum_{t=1}^T f_t(\mathbf{x}_t^*)}_{\texttt{term (b)}}.\end{aligned}$$

For term (a), since the piece-wise stationary comparator sequence only change $M-1$ times over the time horizon, its path length is at most $D(M-1)$. As a consequence, the universal dynamic regret guarantee (33) of the algorithm ensures

$$\texttt{term (a)} \le \sqrt{A_T\left(D + D(M-1)\right)} \le \sqrt{DA_T\left(1+\frac{T}{\Delta}\right)} \le \sqrt{DA_T} + \sqrt{\frac{DTA_T}{\Delta}}.$$

Moreover, the argument in Besbes et al. (2015, Proposition 2) shows that

$$\texttt{term (b)} \le 2\Delta V_T^f.$$

Combining the upper bounds for term (a) and term (b), we obtain

$$\text{D-Regret}_T(\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*) \le \sqrt{DA_T} + \sqrt{\frac{DTA_T}{\Delta}} + 2\Delta V_T^f.$$

The optimal choice of the interval length is $\Delta_* := (DTA_T)^{1/3}(V_T^f)^{-2/3}$, which will lead to an $\mathcal{O}(\sqrt{A_T} + A_T^{\frac{1}{3}} T^{\frac{1}{3}} (V_T^f)^{\frac{1}{3}})$ worst-case dynamic regret. However, a caveat is that the interval length $\Delta \in [T]$ should be a positive integer. We thus use the clipped version $\Delta_\dagger := \min\{\lceil \Delta_* \rceil, T\}$. We show that the desired bound (34) is achievable with $\Delta_\dagger$ by considering the following three cases.

- **Case 1** ($1 \leq \Delta_* \leq T$): in such a case, $\Delta_\dagger = \lceil \Delta_* \rceil$ and we have $\Delta_* \leq \Delta_\dagger \leq 2\Delta_*$. Then, the dynamic regret is bounded by

  $$\text{D-Regret}_T(\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*) \leq \sqrt{DA_T} + \sqrt{\frac{DTA_T}{\Delta_*}} + 4\Delta_* V_T^f \leq \sqrt{DA_T} + 5D^{\frac{1}{3}} A_T^{\frac{1}{3}} T^{\frac{1}{3}} (V_T^f)^{\frac{1}{3}}.$$

  We complete the proof for case 1 by combining the result with the path-length worst-case dynamic regret bound in (52).

- **Case 2** ($\Delta_* > T$): in such a case, $\Delta_\dagger = T$ and $\sqrt{DA_T} \geq TV_T^f$. Then, we have

  $$\text{D-Regret}_T(\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*) \leq \sqrt{DA_T} + \sqrt{DA_T} + 2TV_T^f \leq 3\sqrt{DA_T}.$$

  We complete the proof for case 2 by combining the result with the path-length worst-case dynamic regret bound in (52).

- **Case 3** ($\Delta_* \leq 1$): in such a case, $\Delta_\dagger = 1$ and $\sqrt{DA_T T} \leq V_T^f$. Since $P_T^* \leq DT$, we have $\sqrt{A_T P_T^*} \leq \sqrt{DA_T T} \leq D^{\frac{1}{3}} A_T^{\frac{1}{3}} T^{\frac{1}{3}} (V_T^f)^{\frac{1}{3}}$, indicating that the path-length bound (52) is tighter than the desired result (34), which completes the proof for case 3.

Overall, the proof is completed by combining the above three cases. ■

7.8 Proof of Theorem 8

Proof The theorem is proved by the probabilistic method, following the proof of Zhang et al. (2017, Theorem 5). For iterations $t = 1, \ldots, T$, we randomly sample a convex and smooth function $f_t : \mathbb{R}^d \mapsto \mathbb{R}$ from the distribution $\mathcal{P}$.

More specifically, we construct the function as $f_t(\mathbf{x}) = \|\mathbf{x} - \sigma\boldsymbol{\varepsilon}_t\|_2^2$, where $\sigma > 0$ and $\boldsymbol{\varepsilon}_t \in \mathbb{R}^d$ is a random vector with components sampled independently from the Rademacher distribution, i.e., $\boldsymbol{\varepsilon}_t(i) = 1$ or -1 with equal probability of 50%. We further set the comparator $\mathbf{u}_t = \mathbf{x}_t^* \in \arg\min_{\mathbf{x}\in\mathcal{X}} f_t(\mathbf{x}) = \sigma\boldsymbol{\varepsilon}_t$. Denote by $\mathbf{x}_t$ the decision returned by any deterministic online algorithm $\mathcal{A}$. Then the expected dynamic regret is defined as

$$\mathbb{E}[\text{D-Regret}_T] = \mathbb{E}\left[\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t)\right].$$

In the following we show that $\mathbb{E}[\text{D-Regret}_T] \geq \mathbb{E}[P_T(\mathbf{u}_1, \ldots, \mathbf{u}_T)]$. On one hand,

$$\begin{aligned}\mathbb{E}[\text{D-Regret}_T] = \mathbb{E}\left[\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t)\right] &= \sum_{t=1}^T \mathbb{E}[\|\mathbf{x}_t - \sigma\boldsymbol{\varepsilon}_t\|_2^2] \\ &= \sum_{t=1}^T \mathbb{E}[\|\mathbf{x}_t\|_2^2 + 2\sigma\langle \mathbf{x}_t, \boldsymbol{\varepsilon}_t\rangle + \sigma^2\|\boldsymbol{\varepsilon}_t\|_2^2] \geq dT\sigma^2,\end{aligned}$$

where the last inequality holds since $\sigma\langle \mathbf{x}_t, \varepsilon_t\rangle \geq 0$ and $\mathbb{E}[\sigma^2\|\varepsilon_t\|_2^2] \geq d\sigma^2$ for any $t \geq 1$. On the other hand, we have

$$\mathbb{E}[P_T(\mathbf{u}_1, \ldots, \mathbf{u}_T)] = \sigma \cdot \sum_{t=2}^T \mathbb{E}[\|\boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_{t-1}\|_2] = \sigma \cdot \sum_{t=2}^T \mathbb{E}\left[\sqrt{\sum_{i=1}^d \boldsymbol{\delta}_t^2(i)}\right] \leq 2\sqrt{d}T\sigma,$$

where $\boldsymbol{\delta}_t(i) = \boldsymbol{\varepsilon}_t(i) - \boldsymbol{\varepsilon}_{t-1}(i)$. By choosing $\sigma \geq 2/\sqrt{d}$, we can ensure that $\mathbb{E}[\text{D-Regret}_T] \geq \mathbb{E}[P_T(\mathbf{u}_1, \ldots, \mathbf{u}_T)]$. We note that the choice of σ might lead to a violation of the assumption of domain boundedness, which can be easily fixed by the rescaling. So the probabilistic argument implies that for any algorithm $\mathcal{A}$ there exists a sequence of online functions $f_1, \ldots, f_T$ such that $\text{D-Regret}_T \geq P_T(\mathbf{u}_1, \ldots, \mathbf{u}_T)$, which concludes the proof. ■

8. Experiments

In this section, we provide empirical studies to validate the effectiveness of our proposed algorithm and support the theoretical findings.

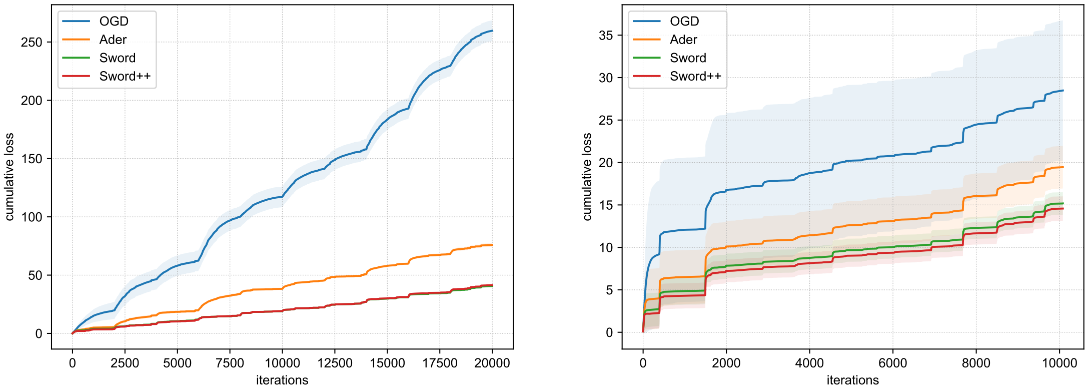
Settings. We simulate the online prediction environments as follows. The player *sequentially* receives the feature of an instance and is then required to make the prediction. We focus on the problem of online regression, where at each round an instance $(\mathbf{x}_t, y_t)$ is received with $\mathbf{x}_t \in \mathcal{X} \subseteq \mathbb{R}^d$ being the feature and $y_t \in \mathcal{Y} \subseteq \mathbb{R}$ being the corresponding label. At each round, the player first receives the feature $\mathbf{x}_t$ and is required to make the prediction by $\widehat{y}_t = \mathbf{x}_t^\top \mathbf{w}_t$ based on the learned model $\mathbf{w}_t \in \mathcal{W} \subseteq \mathbb{R}^d$; then, the ground-truth label $y_t \in \mathbb{R}$ is revealed and the player suffers a loss of $\ell(y_t, \widehat{y}_t)$, where in the simulation we choose the Huber loss defined as

$$\ell(y, \widehat{y}) = \begin{cases} \frac{1}{2}(y - \widehat{y})^2, & \text{for } |y - \widehat{y}| \leq \delta, \\ \delta(|y - \widehat{y}| - \frac{1}{2}\delta), & \text{otherwise.} \end{cases}$$

As a result, the online function can be regarded as a composition of the loss function and the data item, that is, $f_t : \mathcal{W} \mapsto \mathbb{R}$ with $f_t(\mathbf{w}) = \ell(y_t, \widehat{y}_t)$. It can be verified that the functions are convex, and satisfy the condition of non-negativity and smoothness. The player will receive the feedback of the online function and subsequently update her model.

Datasets. We compare the performance on both synthetic and real-world datasets. First, the synthetic data are generated as follows: at each round, the feature $\mathbf{x}_t \in \mathbb{R}^d$ is randomly generated from a ball with a radius of Γ, i.e., $\mathfrak{B} = \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x}\|_2 \leq \Gamma\}$; the associated label is set as $y_t = \mathbf{x}_t^\top \mathbf{w}_t^* + \varepsilon_t$, where ε_t is the random noise drawn from $[0, 0.1]$ and $\mathbf{w}_t^* \in \mathbb{R}^d$ is the underlying model. The underlying model $\mathbf{w}_t^*$ is randomly sampled from a ball with a radius of $D/2$ (recall that D is the diameter of the feasible domain throughout the paper), and it is forced to be stationary within a stage and will be changed every S rounds to simulate the non-stationary environments with abrupt changes. In our simulation, we set $\Gamma = 1$, $D = 2$, $d = 5$, $T = 50000$, $S = 1000$, and $\delta = 2$. Next, we employ a real-world dataset called Sulfur recovery unit (SRU) (Zhao et al., 2021b), which is a regression dataset with slowly evolving distribution changes. There are in total 10,081 data samples representing the records of gas diffusion, where the feature consists of five different chemical and physical indexes and the label is the concentration of SO_2.

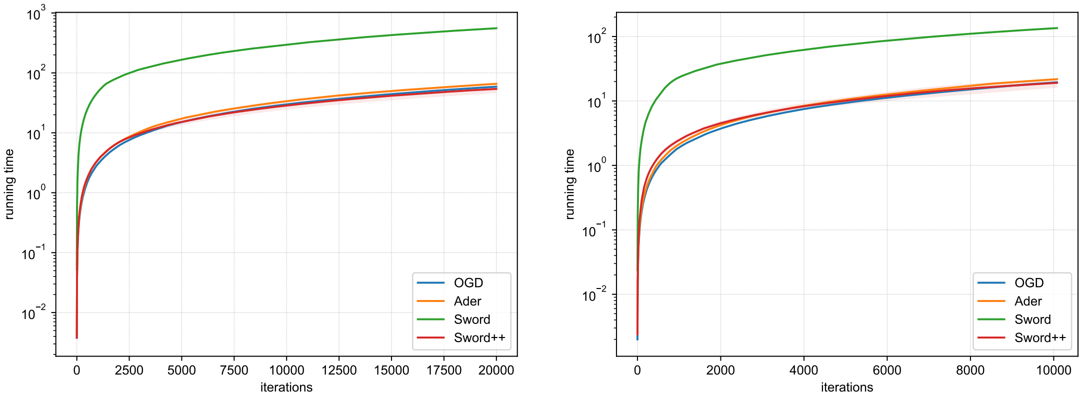
Contenders. We compare the proposed algorithm Sword++ with the following three contenders: (i) OGD (Zinkevich, 2003), online gradient descent, which is an OCO algorithm designed for optimizing static regret; (ii) Ader (Zhang et al., 2018a), an OCO algorithm designed for optimizing dynamic regret yet with only problem-independent guarantee; (iii)



(a) synthetic data

(b) real-world data

Figure 1: Performance comparisons of all algorithms (OGD, Ader, Sword, and Sword++) in terms of cumulative loss.



(a) synthetic data

(b) real-world data

Figure 2: Performance comparisons of all algorithms (OGD, Ader, Sword, and Sword++) in terms of running time (in seconds).

Sword, the algorithm proposed in Section 4.2, which achieves problem-dependent dynamic regret guarantees yet requires multiple gradients per iteration.

Results. We repeat the experiments for five times and report the mean and the standard deviation in Figure 1 and Figure 2. In Figure 1, we examine the performance in terms of cumulative loss. First, we can observe that OGD incurs a large cumulative loss over the horizon and is not able to effectively learn from the non-stationary environments. By contrast, both Ader and our approach (Sword, Sword++) achieve a satisfactory performance in present of distribution changes. Moreover, Sword and Sword++ exploit the adaptivity of the problem instance and thus achieve an encouraging empirical behavior than Ader, which demonstrates the empirical effectiveness. Figure 2 reports the running time comparison, where the y-axis uses a logarithmic scale for a better presentation. We can observe that OGD is the most computationally efficient; besides, Ader and Sword++ are also comparable. By contrast, Sword requires significantly more running time. The result accords to our theory well, in that the gradient computation is the most time-consuming in our simulations. Theoretically, both Sword++ and Ader (with linearized surrogate loss) only require one gradient query per iteration, which shares the same gradient query complexity with OGD. On the contrary, Sword needs to query $N = \mathcal{O}(\log T)$ gradients at each round and is thus much more computational inefficient. To summarize, the empirical results validate the advantage of Sword++, which behaves well and is also computationally lightweight.

9. Conclusion

In this paper, we exploit the easiness of problem instances to enhance the universal dynamic regret. We propose two novel online ensemble algorithms, Sword and Sword++, for convex and smooth online learning. Both algorithms achieve a best-of-both-worlds dynamic regret of order $\mathcal{O}(\sqrt{(1 + P_T + \min\{V_T, F_T\})(1 + P_T)})$, where V_T measures the gradient variation and F_T is the cumulative loss of comparators. These quantities are at most $\mathcal{O}(T)$ yet can be very small when the problem is easy, hence reflecting the difficulty of problem instance. Consequently, our bounds can outperform the $\mathcal{O}(\sqrt{T(1 + P_T)})$ minimax dynamic regret (Zhang et al., 2018a) by exploiting smoothness. Our results are accomplished by several crucial technical ingredients. We adopt optimistic mirror descent as a unified building block for both base and meta algorithms, and carefully exploit the negative terms in the regret analysis. Moreover, in the design of Sword++, we introduce the framework of *collaborative online ensemble*, which emphasizes the construction of surrogate loss in the algorithm design and devises a decision-deviation correction term in conjunction with the linearized loss to facilitate collaboration within the meta-base two layers. By incorporating these elements, we can finally achieve favorable problem-dependent dynamic regret guarantees under the one-gradient feedback model.

All of attained dynamic regret bounds are universal in the sense that they hold against *any* feasible comparator sequence, making the algorithms adaptive to non-stationary environments. An important future work is to investigate the optimality of the our attained problem-dependent dynamic regret bounds. We now only have very preliminary understandings for small-loss dynamic regret (see the lower bound in Theorem 8 and also discussions in Remark 7) and some conjectures regarding the gradient-variation bound (see Remark 7), but a complete understanding requires refined lower bounds that take problem-dependent quantities into account. Moreover, it is important to investigate the possibility of exploiting other function curvatures for the analysis of universal dynamic regret, such as strong convexity, exp-concavity, and self-concordance.

Acknowledgment

This research was supported by NSFC (61921006, 61976112) and the Collaborative Innovation Center of Novel Software Technology and Industrialization. The authors would like to thank Mengxiao Zhang and Ashok Cutkosky for helpful discussions. We are also grateful for anonymous reviewers for their valuable comments.

References

J. D. Abernethy, P. L. Bartlett, A. Rakhlin, and A. Tewari. Optimal stragies and minimax lower bounds for online convex games. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT)*, pages 415–424, 2008.

D. Adamskiy, W. M. Koolen, A. V. Chernov, and V. Vovk. A closer look at adaptive regret. In *Proceedings of the 23rd International Conference on Algorithmic Learning Theory (ALT)*, pages 290–304, 2012.

Z. Allen-Zhu, S. Bubeck, and Y. Li. Make the minority great again: First-order regret bound for contextual bandits. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pages 186–194, 2018.

C. Allenberg, P. Auer, L. Györfi, and G. Ottucsák. Hannan consistency in on-line learning in case of unbounded losses under partial monitoring. In *Proceedings of the 17th International Conference on Algorithmic Learning Theory (ALT)*, pages 229–243, 2006.

P. Auer, N. Cesa-Bianchi, and C. Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64(1):48–75, 2002.

D. Baby and Y.-X. Wang. Online forecasting of total-variation-bounded sequences. In *Advances in Neural Information Processing Systems 32 (NeurIPS)*, pages 11071–11081, 2019.

D. Baby and Y.-X. Wang. Optimal dynamic regret in exp-concave online learning. In *Proceedings of the 34th Conference on Learning Theory (COLT)*, pages 359–409, 2021.

O. Besbes, Y. Gur, and A. J. Zeevi. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, 2015.

O. Bousquet and M. K. Warmuth. Tracking a small set of experts by mixing past posteriors. *Journal of Machine Learning Research*, 3:363–396, 2002.

N. Cesa-Bianchi, Y. Freund, D. Haussler, D. P. Helmbold, R. E. Schapire, and M. K. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.

N. Cesa-Bianchi, Y. Mansour, and G. Stoltz. Improved second-order bounds for prediction with expert advice. In *Proceedings of the 18th Annual Conference on Learning Theory (COLT)*, pages 217–232, 2005.

N. Cesa-Bianchi, P. Gaillard, G. Lugosi, and G. Stoltz. Mirror descent meets fixed share (and feels no regret). In *Advances in Neural Information Processing Systems 25 (NIPS)*, pages 989–997, 2012.

G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993.

X. Chen, Y. Wang, and Y.-X. Wang. Non-stationary stochastic optimization under $L_{p,q}$-variation measures. *Operations Research*, 67(6):1752–1765, 2019.

C. Chiang, C. Lee, and C. Lu. Beating bandits in gradually evolving worlds. In *Proceedings of the 26th Annual Conference on Learning Theory (COLT)*, pages 210–227, 2013.

C.-K. Chiang, T. Yang, C.-J. Lee, M. Mahdavi, C.-J. Lu, R. Jin, and S. Zhu. Online optimization with gradual variations. In *Proceedings of the 25th Conference On Learning Theory (COLT)*, pages 6.1–6.20, 2012.

A. Cotter, O. Shamir, N. Srebro, and K. Sridharan. Better mini-batch algorithms via accelerated gradient methods. In *Advances in Neural Information Processing Systems 24 (NIPS)*, pages 1647–1655, 2011.

A. Cutkosky. Combining online learning guarantees. In *Proceedings of the 32rd Conference on Learning Theory (COLT)*, pages 895–913, 2019.

A. Cutkosky. Parameter-free, dynamic, and strongly-adaptive online learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 2250–2259, 2020.

A. Cutkosky and F. Orabona. Black-box reductions for parameter-free online learning in banach spaces. In *Proceedings of the 31st Conference on Learning Theory (COLT)*, pages 1493–1529, 2018.

H. Fang, N. Harvey, V. S. Portella, and M. P. Friedlander. Online mirror descent and dual averaging: keeping pace in the dynamic case. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 3008–3017, 2020.

D. J. Foster and A. Krishnamurthy. Efficient first-order contextual bandits: Prediction, allocation, and triangular discrimination. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, pages 18907–18919, 2021.

Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

A. György and C. Szepesvári. Shifting regret, mirror descent, and matrices. In *Proceedings of the 33nd International Conference on Machine Learning (ICML)*, pages 2943–2951, 2016.

E. Hazan. Introduction to Online Convex Optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016.

E. Hazan and S. Kale. Extracting certainty from uncertainty: Regret bounded by variation in costs. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT)*, pages 57–68, 2008.

E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

M. Herbster and M. K. Warmuth. Tracking the best expert. *Machine Learning*, 32(2): 151–178, 1998.

M. Herbster and M. K. Warmuth. Tracking the best linear predictor. *Journal of Machine Learning Research*, 1:281–309, 2001.

A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan. Online optimization : Competing with dynamic comparators. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 398–406, 2015.

C. Lee, H. Luo, and M. Zhang. A closer look at small-loss bounds for bandits with graph feedback. In *Proceedings of the 33th Conference on Learning Theory (COLT)*, pages 2516–2564, 2020a.

C.-W. Lee, H. Luo, C.-Y. Wei, and M. Zhang. Bias no more: high-probability data-dependent regret bounds for adversarial bandits and MDPs. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, pages 15522–15533, 2020b.

N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

H. Luo and R. E. Schapire. Achieving all with no parameters: AdaNormalHedge. In *Proceedings of the 28th Annual Conference Computational Learning Theory (COLT)*, pages 1286–1304, 2015.

T. Lykouris, K. Sridharan, and É. Tardos. Small-loss bounds for online learning with partial information. In *Proceedings of the 31st Conference on Learning Theory (COLT)*, pages 979–986, 2018.

A. Mokhtari, S. Shahrampour, A. Jadbabaie, and A. Ribeiro. Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *Proceedings of the 55th IEEE Conference on Decision and Control (CDC)*, pages 7195–7201, 2016.

G. Neu. First-order regret bounds for combinatorial semi-bandits. In *Proceedings of the 28th Conference on Learning Theory (COLT)*, pages 1360–1375, 2015.

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

R. Pogodin and T. Lattimore. On first-order bounds, variance and gap-dependent bounds for adversarial bandits. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 894–904, 2019.

A. Rakhlin and K. Sridharan. Online learning with predictable sequences. In *Proceedings of the 26th Conference On Learning Theory (COLT)*, pages 993–1019, 2013.

S. Shalev-Shwartz. Online Learning: Theory, Algorithms and Applications. *PhD Thesis*, 2007.

S. Shalev-Shwartz. Online Learning and Online Convex Optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2012.

N. Srebro, K. Sridharan, and A. Tewari. Smoothness, low noise and fast rates. In *Advances in Neural Information Processing Systems 23 (NIPS)*, pages 2199–2207, 2010.

M. Sugiyama and M. Kawanabe. *Machine Learning in Non-stationary Environments: Introduction to Covariate Shift Adaptation*. The MIT Press, 2012.

V. Syrgkanis, A. Agarwal, H. Luo, and R. E. Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems 28 (NIPS)*, pages 2989–2997, 2015.

T. van Erven and W. M. Koolen. Metagrad: Multiple learning rates in online learning. In *Advances in Neural Information Processing Systems 29 (NIPS)*, pages 3666–3674, 2016.

C.-Y. Wei and H. Luo. More adaptive algorithms for adversarial bandits. In *Proceedings of the 31st Conference on Learning Theory (COLT)*, pages 1263–1291, 2018.

C.-Y. Wei, Y.-T. Hong, and C.-J. Lu. Tracking the best expert in non-stationary stochastic environments. In *Advances in Neural Information Processing Systems 29 (NIPS)*, pages 3972–3980, 2016.

T. Yang, M. Mahdavi, R. Jin, and S. Zhu. Regret bounded by gradual variation for online convex optimization. *Machine Learning*, 95(2):183–223, 2014.

T. Yang, L. Zhang, R. Jin, and J. Yi. Tracking slowly moving clairvoyant: Optimal dynamic regret of online learning with true and noisy gradient. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, pages 449–457, 2016.

J. Yuan and A. G. Lamperski. Trading-off static and dynamic regret in online least-squares and beyond. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, pages 6712–6719, 2020.

L. Zhang and Z.-H. Zhou. Stochastic approximation of smooth and strongly convex functions: Beyond the $O(1/T)$ convergence rate. In *Proceedings of the 32nd Conference on Learning Theory (COLT)*, pages 3160–3179, 2019.

L. Zhang, T. Yang, R. Jin, and X. He. $O(\log T)$ projections for stochastic optimization of smooth and strongly convex functions. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, volume 28, pages 1121–1129, 2013.

L. Zhang, T. Yang, J. Yi, R. Jin, and Z.-H. Zhou. Improved dynamic regret for non-degenerate functions. In *Advances in Neural Information Processing Systems 30 (NIPS)*, pages 732–741, 2017.

L. Zhang, S. Lu, and Z.-H. Zhou. Adaptive online learning in dynamic environments. In *Advances in Neural Information Processing Systems 31 (NeurIPS)*, pages 1330–1340, 2018a.

L. Zhang, T. Yang, R. Jin, and Z.-H. Zhou. Dynamic regret of strongly adaptive methods. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pages 5877–5886, 2018b.

L. Zhang, T.-Y. Liu, and Z.-H. Zhou. Adaptive regret of convex and smooth functions. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 7414–7423, 2019.

L. Zhang, S. Lu, and T. Yang. Minimizing dynamic regret and adaptive regret simultaneously. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 309–319, 2020a.

L. Zhang, W. Jiang, S. Lu, and T. Yang. Revisiting smoothed online learning. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, pages 13599–13612, 2021.

Y.-J. Zhang, P. Zhao, and Z.-H. Zhou. A simple online algorithm for competing with dynamic comparators. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 390–399, 2020b.

P. Zhao. *Online Ensemble Theories and Methods for Robust Online Learning*. PhD thesis, Nanjing University, Nanjing, China, 2021. Advisor: Zhi-Hua Zhou.

P. Zhao and L. Zhang. Improved analysis for dynamic regret of strongly convex and smooth functions. In *Proceedings of the 3rd Conference on Learning for Dynamics and Control (L4DC)*, pages 48–59, 2021.

P. Zhao, L. Zhang, Y. Jiang, and Z.-H. Zhou. A simple approach for non-stationary linear bandits. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 746–755, 2020a.

P. Zhao, Y.-J. Zhang, L. Zhang, and Z.-H. Zhou. Dynamic regret of convex and smooth functions. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, pages 12510–12520, 2020b.

P. Zhao, G. Wang, L. Zhang, and Z.-H. Zhou. Bandit convex optimization in non-stationary environments. *Journal of Machine Learning Research*, 22(125):1 – 45, 2021a.

P. Zhao, X. Wang, S. Xie, L. Guo, and Z.-H. Zhou. Distribution-free one-pass learning. *IEEE Transaction on Knowledge and Data Engineering*, 33:951–963, 2021b.

P. Zhao, Y.-X. Wang, and Z.-H. Zhou. Non-stationary online learning with memory and non-stochastic control. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 2101–2133, 2022.

Z.-H. Zhou. *Ensemble Methods: Foundations and Algorithms*. Chapman & Hall/CRC Press, 2012.

Z.-H. Zhou. Open-environment machine learning. *National Science Review*, 9(8), 07 2022.

M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML)*, pages 928–936, 2003.

Appendix A. Proofs of Lemma 1 and Lemma 2

As discussed in Section 3.2, the versatility of optimistic mirror descent (OMD) makes it very general to derive many existing results in a unified view. In this section, we provide two specific implications of the general dynamic regret results presented in Theorem 1, which essentially serves as the proofs of Lemma 1 (dynamic regret of OEGD) and Lemma 2 (static regret of Optimistic Hedge).

First, by choosing the regularizer as $\psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$, we obtain the dynamic regret bound for the OEGD algorithm as stated in Lemma 1.

Proof [of Lemma 1] We first show a general result for the OMD algorithm with arbitrary $M_t \in \mathbb{R}^d$ and the regularizer $\psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$, and then prove Lemma 1 by choosing $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$.

It is easy to verify that the regularizer $\psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$ is 1-strongly convex with respect to the Euclidean norm $\|\cdot\|_2$ and the induced Bregman divergence is $\mathcal{D}_\psi(\mathbf{x},\mathbf{y}) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2$. As a consequence, Theorem 1 gives

$$\begin{aligned}\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) \le \eta \sum_{t=1}^T \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2 + \frac{1}{2\eta}\sum_{t=1}^T \Big(\|\mathbf{u}_t - \widehat{\mathbf{x}}_t\|_2^2 - \|\mathbf{u}_t - \widehat{\mathbf{x}}_{t+1}\|_2^2\Big) \\ - \frac{1}{2\eta}\sum_{t=1}^T \Big(\|\widehat{\mathbf{x}}_{t+1} - \mathbf{x}_t\|_2^2 + \|\widehat{\mathbf{x}}_t - \mathbf{x}_t\|_2^2\Big).\end{aligned}$$

Notice that the second term can be upper bounded as follows.

$$\begin{aligned}\sum_{t=1}^T \Big(\|\mathbf{u}_t - \widehat{\mathbf{x}}_t\|_2^2 - \|\mathbf{u}_t - \widehat{\mathbf{x}}_{t+1}\|_2^2\Big) &\le \|\mathbf{u}_1 - \widehat{\mathbf{x}}_1\|_2^2 + \sum_{t=2}^T \Big(\|\mathbf{u}_t - \widehat{\mathbf{x}}_t\|_2^2 - \|\mathbf{u}_{t-1} - \widehat{\mathbf{x}}_t\|_2^2\Big) \\ &\le \|\mathbf{u}_1 - \widehat{\mathbf{x}}_1\|_2^2 + \sum_{t=2}^T \|\mathbf{u}_t - \mathbf{u}_{t-1}\|_2 \|\mathbf{u}_t - \widehat{\mathbf{x}}_t + \mathbf{u}_{t-1} - \widehat{\mathbf{x}}_t\|_2 \\ &\le D^2 + 2D\sum_{t=2}^T \|\mathbf{u}_t - \mathbf{u}_{t-1}\|_2,\end{aligned}$$

where we use the triangle inequality and the boundedness of the feasible domain. We further evaluate the last term:

$$\sum_{t=1}^T \Big(\|\widehat{\mathbf{x}}_{t+1} - \mathbf{x}_t\|_2^2 + \|\widehat{\mathbf{x}}_t - \mathbf{x}_t\|_2^2\Big) \ge \sum_{t=2}^T \Big(\|\widehat{\mathbf{x}}_t - \mathbf{x}_{t-1}\|_2^2 + \|\widehat{\mathbf{x}}_t - \mathbf{x}_t\|_2^2\Big) \ge \frac{1}{2}\sum_{t=2}^T \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2, \tag{53}$$

in which the last inequality holds due to the fact $a^2 + b^2 \ge (a+b)^2/2$. Hence, combining all the three inequalities, we get the following result for general OMD with $\psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$,

$$\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) \le \eta \sum_{t=1}^T \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2 + \frac{1}{\eta}(D^2 + 2DP_T) - \frac{1}{4\eta}\sum_{t=2}^T \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2$$

Furthermore, when choosing the optimism as the last-round gradient as $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$, the adaptivity term $\eta \sum_{t=1}^T \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2$ can be upper bounded in the following way:

$$\begin{aligned}
&\eta \sum_{t=1}^T \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2 \\
&= \eta \|\nabla f_1(\mathbf{x}_1)\|_2^2 + \eta \sum_{t=2}^T \|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2 \\
&\leq \eta G^2 + 2\eta \sum_{t=2}^T \|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_t)\|_2^2 + 2\eta \sum_{t=2}^T \|\nabla f_{t-1}(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2 \\
&\leq \eta G^2 + 2\eta \sum_{t=2}^T \sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2 + 2\eta L^2 \sum_{t=2}^T \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2,
\end{aligned}$$

where the last step exploits the L-smoothness of the online functions. Substituting the result back yields

$$\begin{aligned}
\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) &\leq \eta(G^2 + 2V_T) + \frac{1}{2\eta}(D^2 + 2DP_T) + \left(2\eta L^2 - \frac{1}{4\eta}\right) \sum_{t=2}^T \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 \\
&\leq \eta(G^2 + 2V_T) + \frac{1}{2\eta}(D^2 + 2DP_T),
\end{aligned}$$

where the setting of step size $\eta \leq 1/(4L)$ ensures the last term in the first inequality be non-positive. We hence complete the proof. ■

Next, by choosing the regularizer as $\psi(\boldsymbol{p}) = \sum_{i=1}^N p_i \ln p_i$, the loss function as $f_t(\boldsymbol{p}_t) = \langle \boldsymbol{p}_t, \boldsymbol{\ell}_t \rangle$ and optimism $M_t = \boldsymbol{m}_t$, OMD recovers Optimistic Hedge (Rakhlin and Sridharan, 2013). Here, with a slight abuse of notations, we now use $\boldsymbol{p} \in \Delta_N$ to denote the variable. Theorem 1 implies the static regret for Optimistic Hedge algorithm by choosing comparators as a fixed one in the simplex.

Proof [of Lemma 2] When choosing the negative-entropy regularizer $\psi(\boldsymbol{p}) = \sum_{i=1}^N p_i \ln p_i$, it is not hard to verify that ψ is 1-strongly convex with respect to $\|\cdot\|_1$ and the induced Bregman divergence is $\mathcal{D}_\psi(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^N p_i \ln(p_i/q_i)$. As a result, by choosing comparators as $\mathbf{e}_i$ Theorem 1 gives

$$\begin{aligned}
\sum_{t=1}^T \langle \boldsymbol{p}_t, \boldsymbol{\ell}_t \rangle - \sum_{t=1}^T \ell_{t,i} \leq \varepsilon \sum_{t=1}^T \|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2 &+ \frac{1}{2\varepsilon} \sum_{t=1}^T \Big(\mathcal{D}_\psi(\mathbf{e}_i, \widehat{\boldsymbol{p}}_t) - \mathcal{D}_\psi(\mathbf{e}_i, \widehat{\boldsymbol{p}}_{t+1})\Big) \\
&- \frac{1}{\varepsilon} \sum_{t=1}^T \Big(\mathcal{D}_\psi(\widehat{\boldsymbol{p}}_{t+1}, \boldsymbol{p}_t) + \mathcal{D}_\psi(\boldsymbol{p}_t, \widehat{\boldsymbol{p}}_t)\Big).
\end{aligned}$$

Consider the second term on the right hand side. The telescoping structure of the second term implies

$$\frac{1}{2\varepsilon} \sum_{t=1}^T \Big(\mathcal{D}_\psi(\mathbf{e}_i, \widehat{\boldsymbol{p}}_t) - \mathcal{D}_\psi(\mathbf{e}_i, \widehat{\boldsymbol{p}}_{t+1})\Big) \leq \frac{1}{\varepsilon} \mathcal{D}_\psi(\mathbf{e}_i, \widehat{\boldsymbol{p}}_1) = \frac{1}{\varepsilon} \ln(1/p_{1,i}).$$

Moreover, by Pinsker's inequality, we have $\mathcal{D}_\psi(\boldsymbol{p}, \boldsymbol{q}) = \mathrm{KL}(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^N p_i \ln(p_i/q_i) \geq \frac{1}{2}\|\boldsymbol{p} - \boldsymbol{q}\|_1^2$. Therefore, the last term can be lower bounded as

$$\sum_{t=1}^T \Big(\mathcal{D}_\psi(\widehat{\boldsymbol{p}}_{t+1}, \boldsymbol{p}_t) + \mathcal{D}_\psi(\boldsymbol{p}_t, \widehat{\boldsymbol{p}}_t)\Big) \geq \frac{1}{2}\sum_{t=1}^T \Big(\|\widehat{\boldsymbol{p}}_{t+1} - \boldsymbol{p}_t\|_1^2 + \|\boldsymbol{p}_t - \widehat{\boldsymbol{p}}_t\|_1^2\Big) \geq \frac{1}{4}\sum_{t=2}^T \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2,$$

where the last inequality is got by regrouping the sum and applying triangle inequality like (53).

Combining all above these, we achieve

$$\sum_{t=1}^T \langle \boldsymbol{p}_t, \boldsymbol{\ell}_t\rangle - \sum_{t=1}^T \ell_{t,i} \leq \varepsilon \sum_{t=1}^T \|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2 + \frac{\ln N}{\varepsilon} - \frac{1}{4\varepsilon}\sum_{t=2}^T \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2, \tag{54}$$

which ends the proof. ■

Note that the negative term in the regret bound (54) is very essential, which is quite useful in a variety of problems requiring adaptive bounds. Our analysis is based on the unified view of OMD (Theorem 1), and is simpler and easier to understand than the original proof due to Syrgkanis et al. (2015), who interpret the update from the lens of FTRL (Follow-the-Regularized-Leader) and prove the result based on mathematical induction.

Appendix B. Adaptive Learning Rate Version

In the main text, our proposed algorithms (Sword and Sword++) both employ a fixed learning rate for the meta-algorithm, which greatly simplifies the presentation and regret analysis. The learning rate configurations require the knowledge of gradient variation $V_T = \sum_{t=2}^T \sup_{\mathbf{x}\in\mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$ (for Sword) or its variant $\bar{V}_T = \sum_{t=2}^T \|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2$ (for Sword++).

In this section, we present an adaptive version using the self-confident learning rate tuning (Auer et al., 2002) such that the meta-algorithm does not require such information ahead of time.[3] Before presenting the details, we first extend the collaborative online ensemble framework in Section 5 to an adaptive version. We then use it to prove the gradient-variation and small-loss dynamic regret bounds for the adaptive version of Sword++.

B.1 Adaptive Collaborative Online Ensemble

In this part, we provide the adaptive learning rate version of the unified framework presented in Section 5. Comparing with the fixed learning rate version, the only difference is that we run the optimistic Hedge with a time-varying learning rate for the meta-algorithm,

$$p_{t+1,i} \propto \exp\left(-\varepsilon_t\Big(\sum_{s=1}^t \ell_{s,i} + m_{t+1,i}\Big)\right), \tag{55}$$

[3] For simplicity, we only present the adaptive version for Sword++, and the one for Sword can be similarly obtained (which is actually simpler). Moreover, an important note is that our adaptive version also only requires one gradient per iteration, hence still feasible for the one-gradient feedback model.

where the loss vector $\boldsymbol{\ell}_t$ and $\boldsymbol{m}_t$ share the same configurations as (26). For any i-th base-algorithm, we use the same update rule as the fixed learning rate version

$$\mathbf{x}_{t,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i M_t\right], \quad \widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_t)\right], \tag{56}$$

Then, we can generate the prediction for iteration t by $\mathbf{x}_t = \sum_{i=1}^N p_{t,i}\mathbf{x}_{t,i}$ and have the following guarantee.

Theorem 9. *Under the same assumptions and parameter configurations as Theorem 5 and setting the learning rate of the meta-algorithm as*

$$\varepsilon_t = \min\left\{\bar{\varepsilon}, \sqrt{\frac{\ln N}{D^2\sum_{s=1}^t \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2}}\right\}, \tag{57}$$

Then, decisions specified by (55) and (56) satisfy that for any comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$,

$$\sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{u}_t) \leq 5\sqrt{D^2 \ln N A_T} + 2\sqrt{(D^2 + 2DP_T)A_T} + \frac{\ln N}{\bar{\varepsilon}} + 2\bar{\varepsilon}D^2\widetilde{G}^2 \\ + \frac{2(D^2+2DP_T)}{\bar{\eta}} + \left(\lambda - \frac{1}{4\bar{\eta}}\right)S_{x,i} - \frac{1}{4\bar{\varepsilon}}S_p - \lambda S_{\text{mix}}. \tag{58}$$

In above, $\widetilde{G} = \max_{t\in[T]}\|\nabla f_t(\mathbf{x}) - M_t\|_2$, $A_T = \sum_{t=1}^T \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2$ is the adaptivity term measuring the quality of optimistic gradient vectors $\{M_t\}_{t=1}^T$, and $P_T = \sum_{t=2}^T \|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2$ is the path length of comparators. The terms $S_{x,i} = \sum_{t=2}^T \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$, $S_p = \sum_{t=2}^T \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2$ and $S_{\text{mix}} = \sum_{t=2}^T\sum_{i=1}^N p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$ measures the stability of the decisions returned by the base-algorithm, meta-algorithm, and overall algorithm, respectively.

We remark that, in the fixed learning rate case, one can show that the Optimistic Hedge (22) is identical to the optimistic OMD with the negative-entropy regularizer. However, the adaptive learning rate version (55) can only be interpreted as a follow the regularized leader (FTRL) algorithm. Thus, it is hard to directly apply Theorem 1 to obtain the meta-regret. We choose a FTRL-type meta-algorithm instead of an OMD-type algorithm, in that OMD with time-varying learning rates would suffer linear regret in the worst case when using the negative-entropy regularizer. Although one can fix this issue with the stabilization technique (Fang et al., 2020), we just use the FTRL-type update for simplicity. We present the proof of Theorem 9 as follows.

Proof [of Theorem 9] The proof is almost identical to that of Theorem 5. The main difference is that we use a counterpart of Lemma 2 to bound the meta-regret for the adaptive learning rate version (55). Specifically, since (55) is identical to the Optimistic FTRL algorithm $\boldsymbol{p}_{t+1} = \arg\min_{\boldsymbol{p}\in\Delta}\langle \boldsymbol{p}, \sum_{s=1}^t \boldsymbol{\ell}_s + \boldsymbol{m}_{t+1}\rangle + \psi_{t+1}(\boldsymbol{p})$ with the regularizer $\psi_{t+1}(\boldsymbol{p}) = \frac{1}{\varepsilon_t}(\sum_{i=1}^N p_i \ln p_i + \ln N)$,[4] a direct application of Orabona (2019, Theorem 7.35) shows that

[4] Here, we add an additional constant $\ln N$ in the regularizer, which will not effect the solution of the optimization problem and meanwhile make the regret analysis more convenient.

Lemma 3 (Theorem 7.35 of Orabona (2019))**.** *The regret of Optimistic Hedge with a time-varying learning rate $\varepsilon_t > 0$ (see the update specified in (55)) to any expert $i \in [N]$ satisfies*

$$\sum_{t=1}^{T} \langle \boldsymbol{p}_t, \boldsymbol{\ell}_t \rangle - \sum_{t=1}^{T} \ell_{t,i} \le \max_{p \in \Delta} \psi_{T+1}(\boldsymbol{p}) + \sum_{t=1}^{T} \langle \boldsymbol{\ell}_t - \boldsymbol{m}_t, \boldsymbol{p}_t - \boldsymbol{p}_{t+1} \rangle - \sum_{t=1}^{T} \frac{1}{2\varepsilon_{t-1}} \|\boldsymbol{p}_t - \boldsymbol{p}_{t+1}\|_1^2.$$

Then, based on this Lemma 3, we can bound the regret of the Optimistic Hedge by

$$\begin{aligned}
&\sum_{t=1}^{T} \langle \boldsymbol{p}_t, \boldsymbol{\ell}_t \rangle - \sum_{t=1}^{T} \ell_{t,i} \\
&\le \max_{p \in \Delta} \psi_{T+1}(\boldsymbol{p}) + \sum_{t=1}^{T} \langle \boldsymbol{\ell}_t - \boldsymbol{m}_t, \boldsymbol{p}_t - \boldsymbol{p}_{t+1} \rangle - \sum_{t=1}^{T} \frac{1}{2\varepsilon_{t-1}} \|\boldsymbol{p}_t - \boldsymbol{p}_{t+1}\|_1^2 \\
&\le \frac{\ln N}{\varepsilon_T} + \sum_{t=1}^{T} \varepsilon_{t-1} \|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2 + \sum_{t=1}^{T} \frac{1}{4\varepsilon_{t-1}} \|\boldsymbol{p}_t - \boldsymbol{p}_{t+1}\|_1^2 - \sum_{t=1}^{T} \frac{1}{2\varepsilon_{t-1}} \|\boldsymbol{p}_t - \boldsymbol{p}_{t+1}\|_1^2 \\
&\le \frac{\ln N}{\varepsilon_T} + D^2 \sum_{t=1}^{T} \varepsilon_{t-1} \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2 - \sum_{t=2}^{T} \frac{1}{4\varepsilon_{t-1}} \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2 \\
&\le 2\bar{\varepsilon} D^2 G^2 + \frac{\ln N}{\bar{\varepsilon}} + 5\sqrt{D^2 \ln N \sum_{t=1}^{T} \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2} - \sum_{t=2}^{T} \frac{1}{4\varepsilon_{t-1}} \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2,
\end{aligned}$$

where the second inequality is due to the Hölder's inequality $\langle \boldsymbol{\ell}_t - \boldsymbol{m}_t, \boldsymbol{p}_t - \boldsymbol{p}_{t+1} \rangle \le \|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty \|\boldsymbol{p}_t - \boldsymbol{p}_{t+1}\|_1$ and the fact that $ab \le \varepsilon_{t-1} a^2 + \frac{b^2}{4\varepsilon_{t-1}}$ holds for any $a, b, \varepsilon_{t-1} > 0$. The third inequality is by definitions of $\boldsymbol{\ell}_t$ and $\boldsymbol{m}_t$. The last inequality is a consequence of the inequality $\ln N/\varepsilon_T \le \ln N/\bar{\varepsilon} + \sqrt{D^2 \ln N \sum_{t=1}^{T} \|\nabla f_t(\mathbf{x}) - M_t\|_2^2}$ by learning rate configuration (60) and Lemma 11, which provides a clipped version of the self-confident tuning.

Then, by the same argument to obtain the meta-regret in the proof of Theorem 5, see (48), we have

$$\begin{aligned}
\texttt{meta-regret} \le{} & 5\sqrt{D^2 (\ln N) A_T} + \frac{\ln N}{\bar{\varepsilon}} + 2\bar{\varepsilon} D^2 G^2 - \frac{1}{4\bar{\varepsilon}} \sum_{t=2}^{T} \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2 \\
& - \lambda \sum_{t=1}^{T} \sum_{i=1}^{N} p_{t,i} \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 + \lambda \sum_{t=1}^{T} \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2, \qquad (59)
\end{aligned}$$

which holds from any base-algorithm $i \in [N]$. Subsequently, following the same arguments in the proof of Theorem 5, we can identify an optimal base-algorithm indexed by $i^* \in [N]$, whose base-regret is bounded by

$$\texttt{base-regret} \le 2\sqrt{(D^2 + 2DP_T) A_T} + \frac{2(D^2 + 2DP_T)}{\bar{\eta}}.$$

Finally, combining the meta-regret and the base-regret of the i^*-th base-learner we have

$$\begin{aligned}
\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \le{} & 5\sqrt{D^2 \ln N A_T} + 2\sqrt{(D^2 + 2DP_T) A_T} + \frac{\ln N}{\bar{\varepsilon}} + 2\bar{\varepsilon} D^2 G^2 \\
& + \frac{2(D^2 + 2DP_T)}{\bar{\eta}} + \left(\lambda - \frac{1}{4\bar{\eta}}\right) S_{x,i} - \frac{1}{4\bar{\varepsilon}} S_p - \lambda S_{\text{mix}},
\end{aligned}$$

where $S_{x,i} = \sum_{t=2}^{T} \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$, $S_p = \sum_{t=2}^{T} \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2$ and $S_{\text{mix}} = \sum_{t=2}^{T} \sum_{i=1}^{N} p_{t,i} \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$ measures the stability of the decisions. ■

B.2 Adaptive Version of Sword++

We show that the adaptive learning rate version of the framework (55) and (56) with $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$ for $t \geq 2$ ($M_1 = \mathbf{0}$) achieves the same problem-dependent dynamic regret bound as that in Theorem 6.

Theorem 10. *Under the same assumptions and parameter configurations as Theorem 6 and set the learning rate of the meta-algorithm as*

$$\varepsilon_t = \min\left\{\bar{\varepsilon}, \sqrt{\frac{\ln N}{D^2 \sum_{s=1}^{t} \|\nabla f_s(\mathbf{x}_s) - M_s\|_2^2}}\right\} \tag{60}$$

with $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$ for $t \geq 2$ ($M_1 = \mathbf{0}$). Then, decisions specified by (55) and (56) satisfy that for any comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$,

$$\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \mathcal{O}\left(\sqrt{(1 + P_T + \min\{V_T, F_T\})(1 + P_T)}\right). \tag{61}$$

Proof Under the parameter configurations $\lambda = 2L$, $\bar{\eta} = 1/(8L)$ and $\bar{\varepsilon} = 1/(8D^2L)$, the dynamic regret bound of the unified algorithm with the adaptive learning rate (c.f. Theorem 9) is almost the same as that of the fixed learning rate (c.f. Theorem 5) with difference up to constant factors. Thus, the same arguments in the proof of Theorem 4 and Theorem 6 lead to the best-of-both-worlds bound. ■

Appendix C. Technical Lemmas

This section collects several useful technical lemmas frequently used in the proofs. The first one is the Bregman proximal inequality, which is crucial in the analysis of first-order optimization methods based on Bregman divergence.

Lemma 4 (Bregman proximal inequality (Chen and Teboulle, 1993, Lemma 3.2))**.** *Let $\mathcal{X}$ be a convex set in a Banach space. Let $f : \mathcal{X} \mapsto \mathbb{R}$ be a closed proper convex function on $\mathcal{X}$. Given a convex regularizer $\psi : \mathcal{X} \mapsto \mathbb{R}$, we denote its induced Bregman divergence by $\mathcal{D}_\psi(\cdot, \cdot)$. Then, any update of the form*

$$\mathbf{x}_k = \arg\min_{\mathbf{x} \in \mathcal{X}} \{f(\mathbf{x}) + \mathcal{D}_\psi(\mathbf{x}, \mathbf{x}_{k-1})\}$$

satisfies the following inequality

$$f(\mathbf{x}_k) - f(\mathbf{u}) \leq \mathcal{D}_\psi(\mathbf{u}, \mathbf{x}_{k-1}) - \mathcal{D}_\psi(\mathbf{u}, \mathbf{x}_k) - \mathcal{D}_\psi(\mathbf{x}_k, \mathbf{x}_{k-1}) \tag{62}$$

for any $\mathbf{u} \in \mathcal{X}$.

The second one is the stability lemma, which is very useful in analyzing online algorithms based on FTRL or OMD frameworks.

Lemma 5 (stability lemma (Chiang et al., 2012, Proposition 7))**.** *Consider the following two updates: (i)* $\mathbf{x}_* = \arg\min_{\mathbf{x}\in\mathcal{X}} \langle \mathbf{a}, \mathbf{x}\rangle + \mathcal{D}_\psi(\mathbf{x}, \mathbf{c})$*, and (ii)* $\mathbf{x}'_* = \arg\min_{\mathbf{x}\in\mathcal{X}} \langle \mathbf{a}', \mathbf{x}\rangle + \mathcal{D}_\psi(\mathbf{x}, \mathbf{c})$*. When the regularizer* $\psi : \mathcal{X} \mapsto \mathbb{R}$ *is a 1-strongly convex function with respect to the norm* $\|\cdot\|$*, we have* $\|\mathbf{x}_* - \mathbf{x}'_*\| \leq \|(\nabla\psi(\mathbf{c}) - \mathbf{a}) - (\nabla\psi(\mathbf{c}) - \mathbf{a}')\|_* = \|\mathbf{a} - \mathbf{a}'\|_*$*.*

The self-bounding property of smooth functions is crucial and frequently used in proving small-loss bounds for convex and smooth functions.

Lemma 6 (self-bounding property (Srebro et al., 2010, Lemma 3.1))**.** *For an L-smooth and non-negative function* $f : \mathcal{X} \mapsto \mathbb{R}_+$*, we have* $\|\nabla f(\mathbf{x})\|_2 \leq \sqrt{4Lf(\mathbf{x})}, \ \forall \mathbf{x} \in \mathcal{X}$*.*

Notably, from the analysis of original paper (Srebro et al., 2010, Lemma 2.1 and Lemma 3.1), we can find that actually both the non-negativity and smoothness are required outside the domain $\mathcal{X}$, and this is why we require the function $f_t(\cdot)$ to be non-negative and smooth outside the domain $\mathcal{X}$.

Finally, we present several useful inequalities.

Lemma 7. *Let* $a, b > 0$ *and* $x_0 > 0$ *be three positive values. Suppose that* $L \leq ax + \frac{b}{x}$ *holds for any* $x \in (0, x_0]$*. Then, by taking* $x^* = \min\{\sqrt{b/a}, x_0\}$*, we ensure that*

$$L \leq 2\sqrt{ab} + \frac{2b}{x_0}.$$

Proof Suppose $\sqrt{b/a} \leq x_0$, then $x^* = \sqrt{b/a}$ and we have $L \leq ax^* + \frac{b}{x^*} = 2\sqrt{ab}$. Otherwise, $x^* = x_0$ and we have $L \leq ax^* + \frac{b}{x^*} = ax_0 + \frac{b}{x_0}$. Notice that in latter case $x_0 \leq \sqrt{b/a}$ holds, which implies $ax_0 \leq \frac{b}{x_0}$ and hence $ax_0 + \frac{b}{x_0} \leq \frac{2b}{x_0}$. Combining two cases finishes the proof. ■

Lemma 8 (Lemma 19 of Shalev-Shwartz (2007))**.** *For any* $x, y, a \in \mathbb{R}_+$ *that satisfy* $x - y \leq \sqrt{ax}$*,*

$$x - y \leq a + \sqrt{ay}. \tag{63}$$

Lemma 9. *For any* $x, y, a, b \in \mathbb{R}_+$ *that satisfy* $x - y \leq \sqrt{ax} + b$*,*

$$x - y \leq a + b + \sqrt{ay + ab}. \tag{64}$$

Lemma 10 (Lemma 3.5 of Auer et al. (2002))**.** *Let* $a_1, a_2, \ldots, a_T$ *be non-negative real numbers. Then*

$$\sum_{t=1}^{T} \frac{a_t}{\sqrt{\delta + \sum_{s=1}^{t} a_s}} \leq 2\sqrt{\delta + \sum_{t=1}^{T} a_t},$$

where $0/\sqrt{0} = 0$*.*

Lemma 11. *Let $a_1, a_2, \ldots, a_T, b$ and $\bar{c}$ be non-negative real numbers and $a_t \in [0, B]$ for any $t \in [T]$. Let the step size be*

$$c_t = \min\left\{\bar{c}, \sqrt{\frac{b}{\sum_{s=1}^{t} a_s}}\right\} \text{ and } c_0 = \bar{c}.$$

Then, we have

$$\sum_{t=1}^{T} c_{t-1} a_t \leq 2\bar{c}B + 4\sqrt{b \sum_{t=1}^{T} a_t}. \tag{65}$$

Proof This proof shares the same spirit with that of Pogodin and Lattimore (2019, Lemma 4.8). We assume $\sum_{t=1}^{T} a_t \leq B$, otherwise we can directly have $\sum_{t=1}^{T} c_{t-1} a_t \leq \bar{c}B$. When $\sum_{t=1}^{T} a_t > B$, let $t' = \min\{t \in [T] : \sum_{s=1}^{t-1} a_s \geq B\}$. We can decompose the target by

$$\sum_{t=1}^{T} c_{t-1} a_t = \sum_{t=1}^{t'-1} c_{t-1} a_t + \sum_{t=t'}^{T} c_{t-1} a_t.$$

For the first term we have $\sum_{t=1}^{t'-1} c_{t-1} a_t = \sum_{t=1}^{t'-2} c_{t-1} a_t + c_{t'-2} a_{t'-1} \leq 2\bar{c}B$. As for the second term, we have

$$\sum_{t=t'}^{T} c_{t-1} a_t \leq \sum_{t=t'}^{T} \frac{a_t \sqrt{b}}{\sqrt{\sum_{s=1}^{t-1} a_s}} \leq \sum_{t=t'}^{T} \frac{a_t \sqrt{b}}{\sqrt{\frac{1}{2} \sum_{s=1}^{t} a_s}} \leq \sum_{t=1}^{T} \frac{a_t \sqrt{b}}{\sqrt{\frac{1}{2} \sum_{s=1}^{t} a_s}} \leq 4\sqrt{b \sum_{t=1}^{T} a_t}$$

where the second inequality is due to $\sum_{s=1}^{t} a_s \leq B + \sum_{s=1}^{t-1} a_s \leq 2 \sum_{s=1}^{t-1} a_s$ and the last inequality comes from Lemma 10. We complete the proof by combining the two terms. ■