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

Editor: Csaba Szepesvari

Abstract

We investigate online convex optimization in non-stationary environments and choose *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 safeguard 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 *collaborative online ensemble* framework. 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 Online 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) (Hazan, 2016; Orabona, 2019) 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 non-stationary 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, 2020a; Zhao et al., 2021a; Baby and Wang, 2021; Zhang et al., 2021; Zhao et al., 2022c, 2023). 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, changes of online functions usually come from two sources: sampling randomness and environmental non-stationarity, with the latter being the primary concern in non-stationary online learning. Optimizing the worst-case dynamic regret can be problematic in certain scenarios. For instance, consider a stochastic optimization task where f_t's are independently randomly sampled from the same distribution. Then, minimizing the worst-case dynamic regret is not suitable and can lead to overfitting (Zhang et al., 2018a), as the minimizer of online function may significantly deviate from the minimizer of the expected function due to the sampling randomness. By contrast, since universal dynamic regret can accommodate any feasible comparator sequence, it can automatically adapt to underlying distribution shifts.

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

$$P_T = \sum_{t=2}^{T} \|\mathbf{u}_t - \mathbf{u}_{t-1}\|_2 \tag{4}$$

is the path length of comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T$ and thus reflects the non-stationarity of environments. If the path length P_T were 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 *problem-dependent* regret guarantees beyond the worst-case analysis (Roughgarden, 2021). 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. For static regret mininimization, existing studies 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 universal dynamic regret?*

Our results. In this paper, extending our preliminary conference version (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 problem-dependent quantities: the gradient variation of online functions V_T, and the cumulative loss of the comparator sequence F_T, defined as

$$V_T = \sum_{t=2}^{T} \sup_{\mathbf{x} \in \mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2, \text{ and } F_T = \sum_{t=1}^{T} f_t(\mathbf{u}_t). \tag{5}$$

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 <u>S</u>moothness-a<u>w</u>are <u>o</u>nline lea<u>r</u>ning with <u>d</u>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. Compared 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, such as when

both P_T and V_T (or F_T) exhibit sublinear growth in T. Meanwhile, our regret bounds can 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*, where the player can query gradient information multiple times at each round. This algorithm is conceptually simple, yet it requires a gradient query complexity of $\mathcal{O}(\log T)$ at each round. Our second algorithm, Sword++, improves upon this by necessitating only *one* gradient per iteration, despite using a two-layer online ensemble structure. Therefore, Sword++ is not only computationally efficient but also more attractive due to its reduced feedback requirements — it is particularly suitable for the *one-gradient feedback model*, in which the player receives only the gradient $\nabla f_t(\mathbf{x}_t)$ after submitting the decision $\mathbf{x}_t$. Therefore, Sword++ has the potential to be extended to more constrained bandit feedback models.

Technical contributions. Note that existing studies have demonstrated 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 concerned universal dynamic regret, because 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. To address this, we propose an adaptive *online ensemble* method to hedge non-stationarity while extracting adaptivity. Our method incorporates a meta-base two-layer ensemble to hedge the non-stationarity and employs optimistic online learning for adaptive reuse of historical gradient information. Two crucial novel ingredients are designed to achieve favorable problem-dependent guarantees.

- We introduce optimistic online mirror descent (OPTIMISTIC OMD) as a unified building block for the algorithm design of dynamic regret minimization at both meta and base levels.[1] We present generic and completely modular analysis for the dynamic regret of Optimistic OMD, where the *negative term* is essential especially for achieving problem-dependent dynamic regret guarantees.

- We implement an adaptive online ensemble method that combines optimistic online learning for attaining adaptivity with a meta-base structure to hedge non-stationarity. A key innovation is the emphasis on *collaboration* within the online ensemble. In our *collaborative online ensemble* framework, we introduce a novel decision-deviation correction term in algorithm design and simultaneously exploit the negative term in regret analysis, facilitating effective collaboration within two layers, which is crucial for achieving desired problem-dependent bounds with 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 proposed Sword++ algorithm effectively utilizes negative terms and introduces correction terms to ensure effective collaboration within the two layers.

---

1. For the meta-algorithm, we only care about its static regret, which is essentially a special case of the universal dynamic regret. When the meta-algorithm implements Hedge-style updates with changing learning rates, it aligns more with optimistic FTRL (Follow-The-Regularized-Leader) instead of optimistic OMD. However, we prioritize discussing optimistic OMD (hence using a fixed learning rate), intentionally to better illustrate the core ideas of our regret analysis and algorithm design.

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 related works. In Section 3, we introduce the problem setup and the optimistic online mirror descent framework, where a general dynamic regret analysis is provided. Section 4 establishes the gradient-variation dynamic regret bounds under two different gradient feedback models. Section 5 illustrates our proposed collaborative online ensemble framework, which is very general and useful for attaining problem-dependent dynamic regret. Section 6 provides some additional results regarding implications, significance and a lower bound. The major proofs are presented in Section 7. Furthermore, Section 8 reports the experiments. Finally, we conclude the paper in Section 9. Omitted proofs are provided in the appendix.

2. Related Work

In this section, we present a brief overview of both static and dynamic regret minimization in the context of online convex optimization. Additionally, we provide more discussions on the subsequent studies after the preprint of our manuscript is publicly available.

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 two representative problem-dependent bounds — small-loss bound (Srebro et al., 2010) and gradient-variation bound (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 is to exploit the self-bounding property 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. Furthermore, the techniques developed for gradient-variation regret bounds have a profound connection to many other learning problems, including repeated games (Syrgkanis et al., 2015) and stochastic optimization (Sachs et al., 2022).

In addition, problem-dependent static regret bounds are also studied in the bandit online learning setting, including 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., 2020a), linear bandits (Lee et al., 2020a), semibandits (Neu, 2015), graph bandits (Lykouris et al., 2018; Lee et al., 2020b), and contextual bandits (Allen-Zhu et al., 2018; Foster and Krishnamurthy, 2021), etc.

Finally, we mention that problem-dependent regret minimization falls under the wider umbrella of *adaptive online convex optimization* (McMahan and Streeter, 2010; Duchi et al., 2010), with more recent explorations discussed in (McMahan, 2017; Joulani et al., 2020; Cutkosky, 2020b) and the monograph (Orabona, 2019). However, in addition to developing problem-dependent bounds, this field also covers *data-dependent* bounds. A caveat is that these data-dependent bounds (or called "*algorithm-dependent* bounds") might be influenced not only by the complexity of the problem instance but also by the dynamics of the algorithm itself. This can be sometimes undesired, particularly when the data-dependent quantity is not appropriated defined, potentially leading to a misleading representation of the learning problem's difficulty. For instance, if the regret upper bound depends on a data-dependent quantity like $\sum_{t=2}^{T}\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2$ (rather than the problem-dependent quantity $\sum_{t=2}^{T}\sup_{\mathbf{x}\in\mathcal{X}}\|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$), the regret bound becomes affected by the algorithm's decision sequence $\mathbf{x}_1, \ldots, \mathbf{x}_T$. This will misleadingly lead to large bounds in scenarios where the function sequence is constant ($f_1 = \ldots = f_T = f$) but the decision sequence is unstable.

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/switching regret/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 on the most common regularity called the *path length* $P_T = \sum_{t=2}^{T}\|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2$ introduced by Zinkevich (2003), which measures fluctuation of the comparators. 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 $S_T = \sum_{t=2}^{T}\|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2^2$ introduced by Zhang et al. (2017), 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 one of minimizers 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 pioneering 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 achieves 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. Concurrent to our conference version paper (Zhao et al., 2020b), Cutkosky (2020a) proposes a novel online algorithm that achieves the same order of minimax optimal dynamic regret for convex functions as (Zhang et al., 2018a), yet without relying on using a meta-algorithm hedging over a group of base learners. Instead, their method employs the combination strategy developed in parameter-free online learning (Cutkosky and Orabona, 2018; Cutkosky, 2019). Note that it may be possible to modify the algorithm of Cutkosky (2020a) to achieve small-loss bounds; however, attaining gradient-variation bounds would be generally challenging, especially under the one-gradient feedback model. More specifically, it is not hard to modify their framework to incorporate optimistic online learning, but one usually needs to exploit 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 gradient variation $\sup_{\mathbf{x}\in\mathcal{X}}\|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$, to eliminate the difference between decisions $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$. Our algorithms, based on the collaborative online ensemble framework, involve a careful exploitation of negative terms in the regret analysis of both meta and base algorithms, alongside introducing correction terms in the algorithm design. However, as far as we can see, with only one gradient feedback per round, it is challenging for the framework of Cutkosky (2020a) to achieve the gradient-variation bound due to the lack of negative terms in their 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 $\mathcal{O}(T^{2/3}V_T^{f1/3})$ regret for convex functions when the function varia-

tion V_T^f is available, which is improved to $\mathcal{O}(T^{1/3}V_T^{f2/3})$ 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, it is also demonstrated that a 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, as shown in (Zhao and Zhang, 2021, Section 4.2).

2.3 More Discussions

Subsequent works for dynamic regret minimization. There are many developments for dynamic regret minimization after our work became publicly available (Zhao et al., 2021c), and we briefly mention a few here. For exp-concave or strongly convex online functions, optimal dynamic regret can be obtained by algorithms minimizing strongly adaptive regre (Baby and Wang, 2021, 2022a). Dynamic regret of decision-theoretic online learning is substantially explored, including online non-stochastic control (Zhao et al., 2022b; Zhang et al., 2022b; Baby and Wang, 2022b), online MDPs (Fei et al., 2020; Zhao et al., 2022a; Li et al., 2023b), and online games (Zhang et al., 2022a; Yan et al., 2023; Harris et al., 2022). Furthermore, related techniques have been applied to online label shift (Bai et al., 2022) and online covariate shift (Zhang et al., 2023). The efficiency issue regarding the projection complexity of two-layer online ensemble is considered in (Zhao et al., 2022c).

Subsequent works employing the collaborative online ensemble. A pivotal technique in our paper is the collaborative online ensemble framework, which effectively facilitates the *collaboration* between meta and base layers by incorporating correction terms in the algorithm design and exploiting negative terms in the regret analysis. We have found this collaboration crucial for a variety of problems involving deploying a two-layer structure. We mention two particular examples raised in the literature after our result became available, including game theory (Zhang et al., 2022a; Yan et al., 2023) and an intermediate model for bridging stochastic and adversarial optimization (Chen et al., 2023a,b).

- Zhang et al. (2022a) investigate time-varying zero-sum games, introducing individual regret, dynamic NE-regret, and duality gap as the joint performance measures to guide algorithmic design. To handle multiple performance requirements, they deploy a two-layer algorithm for each player, demonstrating that the overall algorithm enjoys favorable regret guarantees. A vital component in their algorithm is to facilitate collaborations between the meta and base layers. This is again achieved by injecting correction terms in base level and exploiting negative terms in regret analysis, as well as leveraging the unique structure of the zero-sum games. These results are further generalized to strongly monotone games (Yan et al., 2023).

- Chen et al. (2023a) investigate the Stochastically Extended Adversarial (SEA) model, initially proposed in (Sachs et al., 2022), serving as an intermediate model to bridge stochastic and adversarial convex optimization. They enhance the theoretical guarantees of (Sachs et al., 2022) by a careful analysis using optimistic online mirror descent.

Furthermore, they generalize the results by considering dynamic regret minimization for the SEA model, accommodating potential distribution shifts. Consequently, they implement a two-layer algorithm similar to Sword++, ensuring a favorable regret. Note that in the SEA model, due to the dependence issue of the random variables, it is necessary to use the collaborative online ensemble like Sword++. As highlighted in (Chen et al., 2023b, Remark 10), deploying an algorithm similar to Sword to the SEA model will fail to yield desired regret bounds, primarily due to the dependence issue introduced by employing intermediate decisions in the meta-base structure.

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 $f_t(\cdot)$.[2] According to the revealed information, the online learning problems are typically classified into *full-information* online learning and *partial-information* online learning (or sometimes 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 at each round;

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

In Section 4.2, we develop the Sword algorithm, which achieves the gradient-variation dynamic regret under the multi-gradient feedback model. In Section 4.3, we present an improved algorithm called Sword++ that can achieve the same dynamic regret guarantee (up to constants) under the more challenging one-gradient feedback model.

To handle non-stationary environments, we focus on the *dynamic regret* measure, 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 gra-

---

2. One may also understand this by defining f_t over the entire $\mathbb{R}^d$ space while constraining the decisions to the feasible domain $\mathcal{X} \subseteq \mathbb{R}^d$.

dients 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 Online Mirror Descent

We employ the algorithmic framework of Optimistic Online Mirror Descent (OPTIMISTIC OMD) (Chiang et al., 2012; Rakhlin and Sridharan, 2013) as a general building block for designing algorithms for non-stationary online learning. Optimistic OMD is an algorithmic realization of optimistic online learning. Compared to the standard online learning setup, the player receives an additional element at each round: an optimistic vector $M_t \in \mathbb{R}^d$. This vector acts as a predictive hint or an optimistic estimate of the upcoming gradient, thereby called "optimistic vector" or simply "optimism". Optimistic OMD starts from the initial point $\widehat{\mathbf{x}}_1 \in \mathcal{X}$ and performs the following two-step updates at each round:

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

which firstly updates by the optimistic vector M_t and then updates by the received gradient $\nabla f_t(\mathbf{x}_t)$. 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 may assume the regularizer to be σ-strongly convex with respect to the norm $\|\cdot\|$, i.e., $\psi(\mathbf{y}) \geq \psi(\mathbf{x}) + \langle\nabla\psi(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + \frac{\sigma}{2}\|\mathbf{y} - \mathbf{x}\|^2$ holds for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. We have the following general result regarding dynamic regret of optimistic OMD.

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 OMD *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 the gradient and optimistic vector;

(ii) the second term can be restructured as $\sum_{t=2}^T \big(\frac{1}{\eta_t}\mathcal{D}_\psi(\mathbf{u}_t, \widehat{\mathbf{x}}_t) - \frac{1}{\eta_{t-1}}\mathcal{D}_\psi(\mathbf{u}_{t-1}, \widehat{\mathbf{x}}_t)\big)$, hence reflecting the *non-stationarity* of environments;

(iii) the last one $-\sum_{t=1}^{T} \frac{1}{\eta_t}\big(\mathcal{D}_\psi(\widehat{\mathbf{x}}_{t+1}, \mathbf{x}_t) + \mathcal{D}_\psi(\mathbf{x}_t, \widehat{\mathbf{x}}_t)\big)$ is the *negative* term, which can 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 $\mathbf{u}_1, \ldots, \mathbf{u}_T$ as well as the setting of optimistic vectors $M_1, \ldots, M_T$. 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 setting the optimistic vectors as all zero vectors as well as competing with time-varying compactors, it recovers the dynamic regret bound of Online Gradient Descent (Zinkevich, 2003). The versatility of this optimistic OMD 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 general dynamic regret analysis of Optimistic OMD (Theorem 1) does not require the smoothness assumption. Nevertheless, this assumption is crucial for achieving problem-dependent dynamic regret bounds. In fact, smoothness has been demonstrated to be essential even in the static regret analysis for first-order methods to achieve gradient-variation bounds, as evidenced in 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. Note that Assumption 4 requires non-negativity outside the domain $\mathcal{X}$, which is a precondition for establishing the self-bounding property for smooth functions, see Lemma 3.1 of Srebro et al. (2010) and Lemma 13.2 of Cutkosky (2023).

Finally, we mention that following previous studies (Adamskiy et al., 2012; Luo and Schapire, 2015), we treat double logarithmic factors in T as a constant. More concretely, our usage of the $\mathcal{O}(\cdot)$-notation emphasizes the dependence on the time horizon T while hiding the $\log \log T$ factors, and also highlights the dependence on path length P_T, as well as the problem-dependent gradient-variation quantity V_T and small-loss quantity F_T.

4. Gradient-Variation Dynamic Regret

Our paper aims to develop online algorithms that can *simultaneously* achieve problem-dependent dynamic regret bounds, scaling with two quantities: the gradient-variation term

V_T and the small-loss term F_T, as defined in (5). 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 (i.e., achieving gradient-variation and small-loss bounds simultaneously) as direct implications.

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 $\widehat{\mathbf{x}}_1 \in \mathcal{X}$ and then updates by

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

where we define $f_0(\mathbf{x}_0) = \mathbf{0}$ and $\Pi_{\mathcal{X}}[\cdot]$ the Euclidean projection onto the nearest point in $\mathcal{X}$. We consider a fixed step size $\eta > 0$ for simplicity. For convex and smooth functions, Chiang et al. (2012) prove that OEGD enjoys an $\mathcal{O}(\sqrt{1 + V_T})$ gradient-variation static regret.

Actually, OEGD can be viewed as a specialization of Optimistic OMD (6) presented in Section 3.2, by choosing the regularizer $\psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$ and the optimistic vector $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$ as well as a fixed step size $\eta > 0$. Therefore, Theorem 1 directly implies the following dynamic regret upper bound for OEGD, with proof in Appendix A.

Lemma 1. *Under Assumptions 1, 2, and 3, by choosing $\eta \leq \frac{1}{4L}$, dynamic regret of OEGD (namely,* OPTIMISTIC OMD *with $\psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$ and $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$) satisfies*

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

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

Lemma 1 immediately implies a static regret bound. By choosing comparators as the best decision in hindsight $\mathbf{u}_1 = \ldots = \mathbf{u}_T \in \arg\min_{\mathbf{x} \in \mathcal{X}} \sum_{t=1}^{T} f_t(\mathbf{x})$, we have $P_T = 0$ and thereby obtain the existing result (Chiang et al., 2012, Theorem 11): $\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \min_{\mathbf{x} \in \mathcal{X}} \sum_{t=1}^{T} f_t(\mathbf{x}) \leq \eta(G^2 + 2V_T) + \frac{D^2}{2\eta} = \mathcal{O}(\sqrt{1 + V_T})$ when setting the step size $\eta = \min\{\sqrt{D^2/(G^2 + 2V_T)}, 1/(4L)\}$. Note that the requirement of V_T in tuning can be removed by 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. Lemma 1 suggests that it is crucial to tune the step size to balance non-stationarity (path length P_T) and adaptivity (gradient-variation V_T) for achieving a tight dynamic regret bound. Ideally, the optimal tuning is $\eta^* = \sqrt{(D^2 + 2DP_T)/(2G^2 + 2V_T)}$, but this requires the prior information of P_T and V_T that are generally unavailable. We note 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, $P_T = \sum_{t=2}^{T} \|\mathbf{u}_t - \mathbf{u}_{t-1}\|_2$ cannot be known or approximated during the learning process. The ideal best comparator sequence, which tightens the upper bound of cumulative loss, satisfies the condition $\sum_{t=1}^{T} f_t(\mathbf{x}_t) \leq \min_{\mathbf{u}_1, \ldots, \mathbf{u}_T} \{\sum_{t=1}^{T} f_t(\mathbf{u}_t) + \mathcal{R}_T(P_T, V_T)\}$, where $\mathcal{R}_T(P_T, V_T)$ denotes the dynamic regret upper bound. Universal dynamic regret does not specifically target this optimal

comparator sequence but aims to adapt to all feasible comparators, making the choice of $\mathbf{u}_1, \ldots, \mathbf{u}_T$ arbitrarily and entirely unknown within the feasible domain.

In summary, while the self-confident tuning can be used to remove the dependence on the unknown gradient variation V_T, it 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 in 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 design an adaptive *online ensemble* method (Zhou, 2012) 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, inspired by the recent advance in learning with multiple learning rates (van Erven and Koolen, 2016; van Erven et al., 2021), we first construct a pool of candidate step sizes to discretize possible range of the optimal step size, and then initialize multiple base-learners 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 step size η_i from the pool. Finally, those predictions from 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 from the meta-algorithm.

Due to the meta-base structure of the above procedures, we can naturally decompose dynamic regret into the following two parts:

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

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 $\{\mathbf{x}_{t,i}\}_{t=1,\dots,T}$ 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 $\mathcal{H} = \{\eta_1, \dots, \eta_N\}$ 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 $\{\nabla f_t(\mathbf{x}_{t,i})\}_{i=1,\dots,N}$ at round $t \in [T]$. Notably, the gradient query complexity is $N = \mathcal{O}(\log T)$ per round, which means the method developed in this part only suits for the multi-gradient feedback model. In Section 4.3, we will further design an improved algorithm applicable for the one-gradient feedback model.

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 $\mathcal{O}(\sqrt{V_T \log N})$. 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 $\mathcal{O}(\sqrt{T \log N})$ meta-regret, which is problem-independent and thus not suitable for us. To this end, we introduce the optimistic Hedge (Syrgkanis et al., 2015) by exploiting the optimistic online learning, and further design a 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.[3] 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 is well known that Optimistic Hedge can be regarded as an instance of Optimistic 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*

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

---

3. 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	**Algorithm 2** Sword: base-algorithm
Input: step size pool $\mathcal{H}$; learning rate ε	**Input:** step size $\eta_i \in \mathcal{H}$
1: Initialization: $\forall i \in [N], p_{0,i} = 1/N$	1: Let $\widehat{\mathbf{x}}_{1,i}, \mathbf{x}_{1,i}$ be any point in $\mathcal{X}$
2: **for** $t = 1$ **to** T **do**	2: **for** $t = 1$ **to** T **do**
3: Receive $\mathbf{x}_{t+1,i}$ from base-learner $\mathcal{B}_i$	3: $\widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_{t,i})]$
4: Update weight $p_{t+1,i}$ by (16)	4: $\mathbf{x}_{t+1,i} = \Pi_{\mathcal{X}}[\widehat{\mathbf{x}}_{t+1,i} - \eta_i \nabla f_t(\mathbf{x}_{t,i})]$
5: Predict $\mathbf{x}_{t+1} = \sum_{i=1}^{N} p_{t+1,i} \mathbf{x}_{t+1,i}$	5: Send $\mathbf{x}_{t+1,i}$ to the meta-algorithm
6: **end for**	6: **end for**

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 $\mathcal{O}(\sqrt{D_T \log N})$ 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:

$$\ell_{t,i} = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t,i} \rangle, \ \forall t \in [T] \text{ and } \forall i \in [N]. \tag{14}$$

- The optimism $\boldsymbol{m}_t \in \mathbb{R}^N$ is set with a careful design: $\boldsymbol{m}_1 = \mathbf{0}$ and

$$m_{t,i} = \langle \nabla f_{t-1}(\bar{\mathbf{x}}_t), \mathbf{x}_{t,i} \rangle, \ \forall t \in [T] \text{ and } \forall i \in [N], \text{ where } \bar{\mathbf{x}}_t = \sum_{i=1}^{N} p_{t-1,i} \mathbf{x}_{t,i}. \tag{15}$$

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

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

Algorithm 1 summarizes detailed procedures of the meta-algorithm, which in conjunction with the base-algorithm of Algorithm 2 yields the Sword algorithm. We also discuss the gradient query complexity of the overall algorithm. In each round $t \in [T]$, at the base level, the algorithm computes the gradient information $\nabla f_t(\mathbf{x}_{t,i})$ for all $i \in [N]$. At the meta level, it additionally requires the gradients $\nabla f_t(\mathbf{x}_t)$ and $\nabla f_t(\bar{\mathbf{x}}_{t+1})$. Consequently, the total gradient query complexity per round is $N + 2 = \mathcal{O}(\log T)$.

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,

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

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. As such, the adaptivity quantity D_T can be converted to the desired gradient variation V_T plus the decision variation of the meta-algorithm, concretely,

$$
\begin{aligned}
\|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2 &\overset{(15)}{=} \max_{i \in [N]} \langle \nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\bar{\mathbf{x}}_t), \mathbf{x}_{t,i} \rangle^2 \\
&\leq D^2 \|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\bar{\mathbf{x}}_t)\|_2^2 \\
&\leq 2D^2 (\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_t)\|_2^2 + \|\nabla f_{t-1}(\mathbf{x}_t) - \nabla f_{t-1}(\bar{\mathbf{x}}_t)\|_2^2) \qquad (18) \\
&\leq 2D^2 \sup_{\mathbf{x} \in \mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2 + 2D^2 L^2 \|\mathbf{x}_t - \bar{\mathbf{x}}_t\|_2^2 \\
&\leq 2D^2 \sup_{\mathbf{x} \in \mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2 + 2D^4 L^2 \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2,
\end{aligned}
$$

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 $\mathcal{O}(\sqrt{(1 + V_T)\log N})$, 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* $\varepsilon = \min\{1/(4D^2 L), \sqrt{(\ln N)/(2D^2(G^2 + V_T))}\}$, *the meta-regret of Sword (Algorithm 1) is at most*

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

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, $p_{t+1,i} \propto \exp\left(-\varepsilon_t(\sum_{s=1}^{t} \ell_{s,i} + m_{t+1,i})\right), \forall i \in [N]$ with $\varepsilon_t = \mathcal{O}(1/\sqrt{1 + V_t})$ with an internal estimate $V_t = \sum_{s=2}^{t} \sup_{\mathbf{x} \in \mathcal{X}} \|\nabla f_s(\mathbf{x}) - \nabla f_{s-1}(\mathbf{x})\|_2^2$. 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 the 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. A caveat of this Optimistic Hedge update when implemented time-varying learning rates is that it essentially is a special case of Optimistic FTRL rather than Optimistic OMD. For a more thorough technical discussion, see Remark 9 in Appendix B.

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*

$$\mathcal{H} = \left\{ \eta_i = \min\left\{\frac{1}{4L}, 2^{i-1}\sqrt{\frac{D^2}{8G^2T}}\right\} \mid i \in [N] \right\}, \tag{19}$$

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 enjoys the following dynamic regret against any comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$,

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

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 the same dynamic regret guarantee as Sword *using only one gradient query per iteration*, making them applicable to the one-gradient feedback online learning.

We resolve the question affirmatively. The new algorithm, called Sword++, also implements an online ensemble structure. Compared 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:

$$\mathbf{x}_{t,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_{t-1}(\mathbf{x}_{t-1})\right], \quad \widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_t)\right]. \tag{20}$$

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 a 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

$$(\#) \leq \eta_i(G^2 + 2V_T) + \frac{D^2 + 2DP_T}{2\eta_i} + 2\eta_i L^2 \sum_{t=2}^T \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 - \frac{1}{4\eta_i}\sum_{t=2}^T \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2. \tag{21}$$

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 $2\eta_i L^2 \sum_{t=2}^T \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$, which can be cancelled by the negative term $-\sum_{t=2}^T \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2/(4\eta_i)$ whenever the step size is set appropriately. However, when the base-learner updates her decision over the surrogate loss, the additional positive term becomes $2\eta_i L^2 \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2$, which *cannot* be handled by the negative term of any base-learner. Thus, more advanced mechanisms are required to achieve 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. The term "collaboration" refers to the interplay between meta and base layers. Indeed, on their own, neither the base level nor the meta level can achieve a gradient-variation base/meta regret; each incurs some additional positive terms. This positive term necessitates the negative terms from the other layer to help effectively cancel it out.

In the following, we describe the details 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 as the meta-algorithm, but additionally require innovative designs in 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,

$$\ell_{t,i} = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_{t,i} \rangle + \lambda \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2. \tag{22}$$

- 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,

$$m_{t,i} = \langle M_t, \mathbf{x}_{t,i} \rangle + \lambda \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 = \langle \nabla f_{t-1}(\mathbf{x}_{t-1}), \mathbf{x}_{t,i} \rangle + \lambda \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2. \tag{23}$$

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. We will explain soon in Remark 4 on the crucial role and design motivation of this correction. Overall, the meta-algorithm of Sword++ updates the weight $\boldsymbol{p}_{t+1} \in \mathbb{R}^N$ as follows: for any $i \in [N]$,

$$p_{t+1,i} \propto \exp\left(-\varepsilon \Big(\sum_{s=1}^t \ell_{s,i} + m_{t+1,i} \Big)\right), \tag{24}$$

where $\varepsilon > 0$ is a (for simplicity) fixed learning rate. Notably, the update only requires the gradient information of $\nabla f_t(\mathbf{x}_t)$ and thus is feasible for 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 mentioned earlier in (21), 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 directly 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:

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

which concludes that

$$
\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 \leq 2\sum_{t=2}^{T}\sum_{i=1}^{N} p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 + 2D^2\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2. \tag{25}
$$

The right hand side of (25) is a weighted combination of stability of base-learners (hence called *mixed stability*), and the second one is the stability of the meta-algorithm's weights (hence called *meta stability*). We also similarly define $\sum_{t=2}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$ as the *base stability* (of the base learner $i \in [N]$). Clearly, the meta stability can be readily canceled by the negative term of meta-regret. However, it is challenging to address the first positive term, namely, the mixed stability. To overcome the difficulty, we *algorithmically* add 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 — 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 $(-\sum_{t=2}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2)$. A notable characteristic is that the positive terms of meta/base/mixed stability cannot be cancelled solely by the negative terms within their respective layer. Instead, they necessitate additional negative terms, either from regret analysis or algorithmic corrections, to help effectively cancel out. 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 general presentation of this collaborative online ensemble will be provided in Section 5. ¶

Algorithm 3 Sword++: meta-algorithm	**Algorithm 4** Sword++: base-algorithm
Input: step size pool $\mathcal{H}$; learning rate ε	**Input:** step size $\eta_i \in \mathcal{H}$
1: Initialization: $\forall i \in [N], p_{0,i} = 1/N$	1: Let $\widehat{\mathbf{x}}_{1,i}, \mathbf{x}_1$ be any point in $\mathcal{X}$
2: **for** $t = 1$ **to** T **do**	2: **for** $t = 1$ **to** T **do**
3: Receive $\mathbf{x}_{t+1,i}$ from base-learner $\mathcal{B}_i$	3: $\widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_t)]$
4: Update weight $p_{t+1,i}$ by (22)–(24)	4: $\mathbf{x}_{t+1,i} = \Pi_{\mathcal{X}}[\widehat{\mathbf{x}}_{t+1,i} - \eta_i \nabla f_t(\mathbf{x}_t)]$
5: Predict $\mathbf{x}_{t+1} = \sum_{i=1}^{N} p_{t+1,i}\mathbf{x}_{t+1,i}$	5: Send $\mathbf{x}_{t+1,i}$ to the meta-algorithm
6: **end for**	6: **end for**

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 up to constants, 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*

$$\mathcal{H} = \left\{ \eta_i = \min\left\{ \frac{1}{8L}, \sqrt{\frac{D^2}{8G^2T}} \cdot 2^{i-1} \right\} \mid i \in [N] \right\}, \tag{26}$$

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, $\bar{V}_T = \sum_{t=2}^{T}\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2$ is the variant of gradient variation V_T.

Note that the learning rate tuning of the meta-algorithm requires the knowledge of $\bar{V}_T$. Yet, 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 easily realized under the one-gradient feedback model, namely, only $\nabla f_t(\mathbf{x}_t)$ available at round t. To avoid clutter, we here stick to a fixed learning rate instead of a time-varying one, which is more convenient to demonstrate the collaboration between meta and base layers in the regret analysis (also see Remark 8 in the proof of Theorem 4). We also 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 under the challenging one-gradient feedback online learning, and meanwhile suffer no computational overhead in terms of the gradient query complexity.

5. A General Framework: Collaborative Online Ensemble

In this section, we formally introduce the proposed *collaborative online ensemble* framework, a general 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 $g_t^{\texttt{base}} : \mathcal{X} \mapsto \mathbb{R}$ be the base surrogate loss and $h_t^{\texttt{base}} : \mathcal{X} \mapsto \mathbb{R}$ be the base surrogate optimism;

- for the meta-algorithm, let $g_t^{\texttt{meta}} : \mathcal{X} \mapsto \mathbb{R}$ be the meta surrogate loss and $h_t^{\texttt{meta}} : \mathcal{X} \mapsto \mathbb{R}$ be the meta surrogate optimism, and let $\boldsymbol{c}_t \in \mathbb{R}^N$ 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,

$$\mathbf{x}_{t,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla h_t^{\texttt{base}}(\mathbf{x}_{t-1,i})\right], \quad \widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla g_t^{\texttt{base}}(\mathbf{x}_{t,i})\right], \quad (27)$$

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 $\boldsymbol{\ell}_t \in \mathbb{R}^N$ and optimism $\boldsymbol{m}_t \in \mathbb{R}^N$,

$$p_{t+1,i} \propto \exp\left(-\varepsilon\left(\sum_{s=1}^t \ell_{s,i} + m_{t+1,i}\right)\right), \quad (28)$$

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

$$\ell_{t,i} = g_t^{\texttt{meta}}(\mathbf{x}_{t,i}) + \lambda c_{t,i}, \text{ and } m_{t,i} = h_t^{\texttt{meta}}(\mathbf{x}_{t,i}) + \lambda c_{t,i}, \quad (29)$$

where $\lambda \geq 0$ is the coefficient of the correction terms and $c_{t,i}$ is the i-th entry of $\boldsymbol{c}_t$.

Remark 5. The meta-base updates in (27) and (28) are quite versatile, as there are many options for constructing the surrogate (meta/base) loss, optimism, and the correction term.

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

Algorithm	$g_t^{\texttt{base}}(\mathbf{x})$	$h_t^{\texttt{base}}(\mathbf{x})$	$g_t^{\texttt{meta}}(\mathbf{x})$	$h_t^{\texttt{meta}}(\mathbf{x})$	$\boldsymbol{c}_t$
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++	$\langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle$	$\langle \nabla f_{t-1}(\mathbf{x}_{t-1}), \mathbf{x} \rangle$	$\langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle$	$\langle \nabla f_{t-1}(\mathbf{x}_{t-1}), \mathbf{x} \rangle$	$c_{t,i} = \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$
Sword.optimism	$\langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle$	$\langle M_t, \mathbf{x} \rangle$	$\langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle$	$\langle M_t, \mathbf{x} \rangle$	$c_{t,i} = \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$

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 general algorithmic template, including the proposed Sword and Sword++ in the earlier subsections. ¶

5.2 Instantiations

In this part, we present three instantiations of the general 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 $g_t^{\texttt{base}}(\mathbf{x}) = f_t(\mathbf{x})$ and base optimism as $h_t^{\texttt{base}}(\mathbf{x}) = f_{t-1}(\mathbf{x})$;

- meta surrogate loss as $g_t^{\texttt{meta}}(\mathbf{x}) = \langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle$ and meta optimism as $h_t^{\texttt{meta}}(\mathbf{x}) = \langle \nabla f_{t-1}(\bar{\mathbf{x}}_t), \mathbf{x} \rangle$, 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 are 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++. 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 $g_t^{\mathtt{base}}(\mathbf{x}) = \langle\nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle$ and base optimism $h_t^{\mathtt{base}}(\mathbf{x}) = \langle M_t, \mathbf{x}\rangle$;

- meta surrogate loss $g_t^{\mathtt{meta}}(\mathbf{x}) = \langle\nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle$ and meta optimism $h_t^{\mathtt{meta}}(\mathbf{x}) = \langle M_t, \mathbf{x}\rangle$, as well as correction term $\mathbf{c}_t$ 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

$$\mathbf{x}_{t,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i M_t\right], \quad \widehat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}\left[\widehat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_t)\right], \tag{30}$$

and the meta-algorithm updates by

$$p_{t+1,i} \propto \exp\left(-\varepsilon\Big(\sum_{s=1}^{t}\langle\nabla f_s(\mathbf{x}_s), \mathbf{x}_{s,i}\rangle + \lambda\sum_{s=1}^{t+1}\|\mathbf{x}_{s,i} - \mathbf{x}_{s-1,i}\|_2^2 + \langle M_{t+1}, \mathbf{x}_{t+1,i}\rangle\Big)\right). \tag{31}$$

We refer to the above meta-base updates, (30) and (31), 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 general template presented in Section 5.1, specifically, the meta-base updates in (27) and (28). However, the various variables in the general 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 sufficiently general (for instance, it can specialize Sword++).

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

$$\mathcal{H} = \left\{\eta_i = \min\left\{\bar{\eta}, \sqrt{\frac{D^2}{8G^2T}} \cdot 2^{i-1}\right\} \mid i \in [N]\right\}, \tag{32}$$

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

$$\varepsilon = \min \left\{ \bar{\varepsilon}, \sqrt{\frac{\ln N}{D^2 \sum_{t=1}^{T} \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2}} \right\}. \tag{33}$$

Then, Sword.optimism satisfies that for any comparator sequence $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$,

$$\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq 2\sqrt{D^2 (\ln N) A_T} + 2\sqrt{(D^2 + 2DP_T) A_T}$$
$$+ \frac{2 \ln N}{\bar{\varepsilon}} + \frac{2(D^2 + 2DP_T)}{\bar{\eta}} + \left(\lambda - \frac{1}{4\bar{\eta}} \right) S_{x,i} - \frac{1}{4\bar{\varepsilon}} S_p - \lambda S_{\text{mix}}. \tag{34}$$

In above, $A_T = \sum_{t=1}^{T} \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2$ is the adaptivity term, $P_T = \sum_{t=2}^{T} \|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2$ is the path length, $S_{x,i} = \sum_{t=2}^{T} \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$, $S_p = \sum_{t=2}^{T} \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2$, and $S_{\text{mix}} = \sum_{t=2}^{T} \sum_{i=1}^{N} p_{t,i} \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$ are base stability, meta stability, and mixed stability.

The proof of Theorem 5 is presented 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 (34). By further combining the analysis of (25) 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 justification.

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 $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$.

Comparing with Theorem 4, one more assumption (Assumption 4) is required. As mentioned, this non-negativity assumption is a precondition for establishing the self-bounding property for smooth functions (Srebro et al., 2010; Cutkosky, 2023), 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. This also, to some extent, demonstrates the fundamental importance of achieving gradient-variation bounds — sometimes this can directly imply a small-loss bound in the analysis. ¶

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 in Theorem 8 that no algorithm can achieve an $\mathcal{O}(\sqrt{(1 + F_T)(1 + P_T)})$ bound in general. Nevertheless, we fail to provide a similar reasoning for the gradient-variation bound. Indeed, we have the following conjectures. For the multi-gradient feedback model, we are inclined to believe that the $\mathcal{O}(\sqrt{(1 + P_T + V_T)(1 + P_T)})$ rate may not be optimal and it might be possible to achieve $\mathcal{O}(\sqrt{(1 + V_T)(1 + P_T)})$. For the one-gradient feedback model, we conjecture that our obtained rate has already been optimal. We leave this problem-dependent lower bound as an important future work to investigate. ¶

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, for the worst-case dynamic regret, there are two kinds of regularities: path length $P_T^* = \sum_{t=2}^{T} \|\mathbf{x}_{t-1}^* - \mathbf{x}_t^*\|_2$ and function variation $V_T^f = \sum_{t=2}^{T} \sup_{\mathbf{x} \in \mathcal{X}} |f_{t-1}(\mathbf{x}) - f_t(\mathbf{x})|$. The following theorem provides a reduction to both.

Theorem 7. *Let $A_T \in \mathbb{R}_+$ be a certain adaptivity term. Suppose there exists an algorithm $\mathcal{A}$ that ensures the following guarantee: for any comparator sequence $\mathbf{u}_1, \ldots \mathbf{u}_T \in \mathcal{X}$ with path length $P_T = \sum_{t=2}^{T} \|\mathbf{u}_t - \mathbf{u}_{t-1}\|_2$, it holds that*

$$\text{D-Regret}_T(\mathbf{u}_1, \ldots, \mathbf{u}_T) \leq \sqrt{A_T(D + P_T)}, \tag{35}$$

for some constant $D > 0$. Then $\mathcal{A}$ enjoys the worst-case dynamic regret bound:

$$\text{D-Regret}_T(\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*) \leq 3\sqrt{DA_T} + \min\left\{\sqrt{A_T P_T^*}, 5D^{1/3}T^{1/3}A_T^{1/3}(V_T^f)^{1/3}\right\}. \tag{36}$$

Theorem 7 demonstrates that an $\mathcal{O}(\sqrt{A_T(1 + P_T)})$ universal dynamic regret bound can *directly* imply an $\mathcal{O}(\sqrt{A_T} + \min\{\sqrt{A_T P_T^*}, (T A_T V_T^f)^{1/3}\})$ 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 carefully designed 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 (35), 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. Using a similar analysis, we can also convert the gradient-variation/small-loss universal dynamic regret bounds, attained by Sword and Sword++, into the worst-case dynamic regret bounds. Details are omitted here.

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 $\mathcal{O}(\sqrt{T(1 + P_T)})$.

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 $x \in \mathcal{X} \triangleq [-1, 1]$. 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

$$V_T = \sum_{t=2}^{T} \sup_{x \in \mathcal{X}} |(a_{t-1}^2 - a_t^2)x - (a_{t-1} - a_t)|^2 = \sum_{t=2}^{T} \sup_{x \in \mathcal{X}} \left| \left(\frac{T - 2t + 3}{T^2} \right) \cdot x - \frac{1}{T} \right|^2$$

$$= \sum_{t=2}^{K+2} \left(\frac{2T - (2t - 3)}{T^2} \right)^2 + \sum_{t=K+3}^{T} \left(\frac{2t - 3}{T^2} \right)^2 \leq \sum_{t=2}^{T} \left(\frac{2}{T} \right)^2 = \mathcal{O}(1).$$

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

We can observe that the gradient variation $V_T = \mathcal{O}(1)$ is significantly smaller than the small-loss quantity $F_T = \Theta(T)$ in this problem 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 against $u_1, \ldots, u_T$ 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 $V_T = \Theta(T)$ 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 against $u_1, \ldots, u_T$ 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, for any constant $c > 0$,*

$$\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) > c\sqrt{(1 + F_T)(1 + P_T)}. \tag{37}$$

when the time horizon T is sufficiently large.

For the static regret bound, the worst-case minimax rate $\mathcal{O}(\sqrt{T})$ can be improved to $\mathcal{O}(\sqrt{F_T})$ or $\mathcal{O}(\sqrt{V_T})$ by substituting the dependence on T with problem-dependent quantities. A natural question for universal dynamic regret is whether it is possible to also attain an $\mathcal{O}(\sqrt{(1 + \min\{V_T, F_T\})(1 + P_T)})$ bound that improves the minimax rate $\mathcal{O}(\sqrt{T(1 + P_T)})$. Theorem 8 shows that no algorithm can achieve the $\mathcal{O}(\sqrt{(1 + F_T)(1 + P_T)})$ universal dynamic regret bound. We provide the proof in Section 7.8, where the probabilistic method is applied to show the contradiction. In the constructed problem instance, the small-loss quantity is always $F_T = 0$, and there exist a certain online function sequence $\{f_t\}_{t=1}^{T}$ such that the dynamic regret is lower bound by $\Omega(T)$. Therefore, the $\mathcal{O}(\sqrt{(1 + F_T)(1 + P_T)})$ upper bound would violate this lower bound, rendering it unfeasible. 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 1 of Section 3, Theorems 2–4 of Section 4, Theorem 5 of Section 5, and Theorems 6–8 of Section 6.

7.1 Proof of Theorem 1

Proof The instantaneous dynamic regret can be upper bounded and decomposed as

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

In the following, 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).

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, which implies $\|\mathbf{x}_t - \widehat{\mathbf{x}}_{t+1}\| \leq \eta_t \|\nabla f_t(\mathbf{x}_t) - M_t\|_*$ and consequently,

$$\texttt{term (a)} \leq \|\nabla f_t(\mathbf{x}_t) - M_t\|_* \|\mathbf{x}_t - \widehat{\mathbf{x}}_{t+1}\| \leq \eta_t \|\nabla f_t(\mathbf{x}_t) - M_t\|_*^2.$$

We now analyze term (b) and term (c). By the Bregman proximal inequality (Lemma 4) and the OPTIMISTIC 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 update $\widehat{\mathbf{x}}_{t+1} = \arg\min_{\mathbf{x} \in \mathcal{X}} \{\eta_t \langle \nabla f_t(\mathbf{x}_t), \mathbf{x} \rangle + \mathcal{D}_\psi(\mathbf{x}, \widehat{\mathbf{x}}_t)\}$ 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 Substituting the definitions of feedback loss and optimism into Lemma 2 yields

$$\sum_{t=1}^{T} \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_{t,i} \rangle \leq \varepsilon D^2 \|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\bar{\mathbf{x}}_t)\|_2^2 + \frac{\ln N}{\varepsilon} - \frac{1}{4\varepsilon} \sum_{t=2}^{T} \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2.$$

Together with the derivations in (17) and (18), this implies

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

Setting the learning rate as $\varepsilon = \min\{1/(4D^2L), \sqrt{(\ln N)/(2D^2(G^2 + V_T))}\}$, by Lemma 7 we obtain an $2D\sqrt{2(G^2 + V_T)\ln N} + 8D^2L\ln N$ upper bound, which ends 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 due to the convexity of online functions. ∎

7.3 Proof of Theorem 3

Proof As stated in (10), dynamic regret can be decomposed into the meta-regret and base-regret, and the decomposition holds for any base-learner's index $i \in [N]$.

Upper bound of meta-regret. Theorem 2 shows that for any $i \in [N]$,

$$\texttt{meta-regret} = \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{x}_{t,i}) \leq 2D\sqrt{2(4G^2 + V_T)\ln N} + 8D^2L\ln N. \quad (38)$$

Upper bound of base-regret. Lemma 1 indicates that for any index $i \in [N]$,

$$\texttt{base-regret} = \sum_{t=1}^{T} f_t(\mathbf{x}_{t,i}) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \eta_i(G^2 + 2V_T) + \frac{1}{2\eta_i}(D^2 + 2DP_T), \quad (39)$$

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 $\eta_1 \leq \eta^\dagger \leq \eta_N$, due to path length $P_T \in [0, DT]$ and gradient variation $V_T \leq 4G^2(T-1)$ by Assumption 1 and Assumption 2. More importantly, owing to the construction of the step size pool $\mathcal{H}$ in (19), we can assure that there exists an index $i^* \in [N]$ such that $\eta_{i^*} \leq \eta^\dagger \leq \eta_{i^*+1} = 2\eta_{i^*}$. As a result, we pick $i = i^*$ in (39) and get

$$\texttt{base-regret} \leq \eta_{i^*}(G^2 + 2V_T) + \frac{D^2 + 2DP_T}{2\eta_{i^*}}$$

$$\leq \eta^\dagger(G^2 + 2V_T) + \frac{D^2 + 2DP_T}{\eta^\dagger} \quad (40)$$

$$\leq 2\sqrt{(G^2 + 2V_T)(D^2 + 2DP_T)} + 8L(D^2 + 2DP_T) \quad (41)$$

$$\leq \mathcal{O}\left(\sqrt{(1 + P_T + V_T)(1 + P_T)}\right). \quad (42)$$

In above, (41) holds because $\eta^\dagger$ is either η^* or $1/(4L)$ and

- when $\eta^\dagger = \eta^*$, the right hand side of (40) $= 2\sqrt{(G^2 + 2V_T)(D^2 + 2DP_T)}$;

- when $\eta^\dagger = 1/(4L)$, we have $\eta^* = \sqrt{(D^2 + 2DP_T)/(G^2 + 2V_T)} \geq \frac{1}{4L}$, which implies that $\frac{1}{4L}(G^2 + 2V_T) \leq 4L(D^2 + 2DP_T)$. Under such a case, the right hand side of (40) $= 4L(D^2 + 2DP_T) + \frac{1}{4L}(G^2 + 2V_T) \leq 8L(D^2 + 2DP_T)$.

Combining two upper bounds yields (41) and further obtains (42).

Upper bound of overall dynamic regret. Note that the meta-base regret decomposition (10) and meta-regret upper bound (38) hold for any index $i \in [N]$. Hence, we can choose the index as i^* as specified above and further combine the base-regret upper bound (39) to achieve the final desired result. Hence, we complete the proof of Theorem 3. ∎

7.4 Proof of Theorem 4

Proof Since Sword++ is essentially an instantiation of the collaborative online ensemble framework, we prove its dynamic regret building upon the general result of Theorem 5.

We substitute $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$ into (34) of Theorem 5 and notice that

$$
\begin{aligned}
A_T &\leq G^2 + 2\sum_{t=2}^{T}\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_t)\|_2^2 + 2\sum_{t=2}^{T}\|\nabla f_{t-1}(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2 \\
&\leq G^2 + 2\sup_{\mathbf{x}\in\mathcal{X}}\sum_{t=2}^{T}\|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2 + 2L^2\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 \\
&\leq G^2 + 2V_T + 4L^2 S_{\text{mix}} + 4D^2 L^2 S_p.
\end{aligned} \tag{43}
$$

As a result, the first term of (34) of Theorem 5 can be further bounded by

$$
\begin{aligned}
&2\sqrt{D^2(\ln N)A_T} \\
&\leq 2\sqrt{D^2(\ln N)\left(G^2 + 2V_T + 4L^2 S_{\text{mix}} + 4D^2 L^2 S_p\right)} \\
&\leq 2\sqrt{D^2(\ln N)\left(G^2 + 2V_T\right)} + 2\sqrt{D^2(\ln N)(4L^2 S_{\text{mix}} + 4D^2 L^2 S_p)} \\
&\leq 2\sqrt{D^2(\ln N)\left(G^2 + 2V_T\right)} + \frac{2\ln N}{\bar{\varepsilon}} + 8\bar{\varepsilon}D^2 L^2 S_{\text{mix}} + 8\bar{\varepsilon}D^4 L^2 S_p,
\end{aligned} \tag{44}
$$

where the last inequality is a consequence of the AM-GM inequality. Using a similar argument, we can bound the second term of (34) of Theorem 5 by

$$
\begin{aligned}
&2\sqrt{(D^2 + 2DP_T)A_T} \\
&\leq 2\sqrt{(D^2 + 2DP_T)(G^2 + 2V_T + 4L^2 S_{\text{mix}} + 4D^2 L^2 S_p)} \\
&\leq 2\sqrt{(D^2 + 2DP_T)(G^2 + 2V_T)} + \frac{2D^2 + 4DP_T}{\bar{\eta}} + 8\bar{\eta}L^2 S_{\text{mix}} + 8\bar{\eta}D^2 L^2 S_p.
\end{aligned} \tag{45}
$$

31

Plugging (44) and (45) into (34), we get the following dynamic regret bound,

$$
\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t)
$$

$$
\leq 2\sqrt{\ln N \left(G^2 D^2 + 2D^2 V_T\right)} + 2\sqrt{(D^2 + 2DP_T)(G^2 + 2V_T)} + \frac{4\ln N}{\bar{\varepsilon}} + \frac{4(D^2 + 2DP_T)}{\bar{\eta}}
$$

$$
+ \left(\lambda - \frac{1}{4\bar{\eta}}\right) S_{x,i} + \left(8\bar{\eta}D^2 L^2 + 8\bar{\varepsilon}D^4 L^2 - \frac{1}{4\bar{\varepsilon}}\right) S_p + \left(8\bar{\eta}L^2 + 8\bar{\varepsilon}D^2 L^2 - \lambda\right) S_{\mathrm{mix}}. \quad (46)
$$

We complete the proof by dropping the last three non-positive terms, which is ensured by the parameter configurations $\lambda = 2L$, $\bar{\eta} = 1/(8L)$ and $\bar{\varepsilon} = 1/(8D^2 L)$. We finally mention that the $\ln N = \mathcal{O}(\log \log T)$ term is treated as a constant throughout the paper. Actually, this term can be improved to $\log \log P_T$ by imposing a non-uniform prior over base-learners (Zhang et al., 2018a, Proof of Theorem 3). Details are omitted here. ∎

Remark 8 (Collaboration in Regret Analysis)**.** The derivation uses a fixed learning rate for the meta-algorithm, which not only simplifies the proof but also more effectively illustrates the collaboration of meta-base two layers in the analysis. The analysis in (46) highlights the crucial role of *collaboration* between meta and base layers. The positive terms — base stability $S_{x,i}$, meta stability S_p, and mixed stability S_{mix} — *cannot* be cancelled solely by negative terms within their respective layer. Instead, they necessitate additional negative terms, either from regret analysis or algorithmic corrections, to help cancel out. ¶

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,

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

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 (28), we can rewrite the meta-regret as

$$
\texttt{meta-regret} = \sum_{t=1}^{T} \langle \boldsymbol{p}_t, \boldsymbol{\ell}_t\rangle - \sum_{t=1}^{T} \ell_{t,i} - \lambda \sum_{t=1}^{T} \sum_{i=1}^{N} p_{t,i} \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 + \lambda \sum_{t=1}^{T} \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2.
$$

We use Lemma 2 and the setting of step size $\varepsilon = \min\{\bar{\varepsilon}, \sqrt{(\ln N)/(D^2 A_T)}\}$ to get

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

$$
\leq 2\sqrt{D^2(\ln N)A_T} + \frac{2\ln N}{\bar{\varepsilon}} - \frac{1}{4\bar{\varepsilon}}\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2.
$$

Combining above two inequalities, we obtain

$$
\texttt{meta-regret} \leq 2\sqrt{D^2(\ln N)A_T} + \frac{2\ln N}{\bar{\varepsilon}} - \frac{1}{4\bar{\varepsilon}}\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2
$$

$$
- \lambda \sum_{t=1}^{T}\sum_{i=1}^{N}p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 + \lambda \sum_{t=1}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2. \tag{48}
$$

Upper bound of base-regret. By Lemma 1, we obtain the base-regret for any $i \in [N]$,

$$
\texttt{base-regret} \leq \eta_i A_T + \frac{D^2 + 2DP_T}{2\eta_i} - \frac{1}{4\bar{\eta}}\sum_{t=2}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2. \tag{49}
$$

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

$$
\sum_{t=1}^{T}f_t(\mathbf{x}_t) - \sum_{t=1}^{T}f_t(\mathbf{u}_t)
$$

$$
\leq 2\sqrt{D^2(\ln N)A_T} + \eta_i A_T + \frac{D^2 + 2DP_T}{2\eta_i} + \frac{2\ln N}{\bar{\varepsilon}}
$$

$$
+ \left(\lambda - \frac{1}{4\bar{\eta}}\right)\sum_{t=2}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 - \frac{1}{4\bar{\varepsilon}}\sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2 - \lambda \sum_{t=1}^{T}\sum_{i=1}^{N}p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2. \tag{50}
$$

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 $\eta^\dagger = \min\{\eta^*, \bar{\eta}\}$ due to the threshold in the step size pool (32). It can be verified that candidate step sizes range from $\eta_1 = \sqrt{\frac{D^2}{8G^2 T}}$ to $\eta_N = \bar{\eta}$.

- when $\eta^\dagger = \eta^*$, there must be an index i^* satisfying $\eta_{i^*} \leq \eta^* \leq \eta_{i^*+1} = 2\eta_{i^*}$. We choose $i = i^*$ and obtain $\eta_{i^*} A_T + \frac{D^2 + 2DP_T}{2\eta_{i^*}} \leq \eta^* A_T + \frac{D^2 + 2DP_T}{\eta^*} = 2\sqrt{(D^2 + 2DP_T)A_T}$;

- 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 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 $F_T^X = \sum_{t=1}^T f_t(\mathbf{x}_t)$.

$$A_T \le \|\nabla f_1(\mathbf{x}_1)\|_2^2 + 2\sum_{t=2}^T \|\nabla f_t(\mathbf{x}_t)\|_2^2 + 2\sum_{t=2}^T \|\nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2$$

$$\le 8L\sum_{t=1}^T f_t(\mathbf{x}_t) + 8L\sum_{t=2}^T f_{t-1}(\mathbf{x}_{t-1}) \le 16L\sum_{t=1}^T f_t(\mathbf{x}_t) = 16LF_T^X,$$

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

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

$$+ 16D^2 L \ln N + 16L(D^2 + 2DP_T).$$

According to the definition of F_T and F_T^X, the above inequality implies that

$$F_T^X - F_T \le 2\sqrt{16L(D^2 \ln N + D^2 + 2DP_T)F_T^X} + 16L(D^2 \ln N + D^2 + 2DP_T)$$

$$\le 2\sqrt{16L(D^2 \ln N + D^2 + 2DP_T)(F_T + 16L(D^2 \ln N + D^2 + 2DP_T))}$$

$$+ 80L(D^2 \ln N + D^2 + 2DP_T)$$

$$= \mathcal{O}(\sqrt{(1 + P_T + F_T)(1 + P_T)}) + \mathcal{O}(1 + P_T)$$

$$= \mathcal{O}(\sqrt{(1 + P_T + F_T)(1 + P_T)}), \tag{51}$$

where the second inequality is by the converting trick in Lemma 9. This ends the proof. ∎

7.7 Proof of Theorem 7

Proof By the universal dynamic regret bound D-Regret$_T(\mathbf{u}_1, \dots, \mathbf{u}_T) \le \sqrt{A_T(D + P_T)}$ and choosing $\mathbf{u}_t = \mathbf{x}_t^*$, we directly obtain the path-length worst-case dynamic regret bound:

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

In the following, we focus on the function-variation type bound. This is achieved following the argument 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 $\mathbf{x}_{\mathcal{I}_m}^* \in \arg\min_{\mathbf{x} \in \mathcal{X}} \sum_{t \in \mathcal{I}_m} f_t(\mathbf{x})$. There are in total $M = \lceil T/\Delta \rceil$

intervals. Then, we can decompose the worst-case dynamic regret as

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

For term (a), since the piece-wise stationary comparator sequence only changes $M-1$ times, its path length is at most $D(M-1)$. Thus, the universal dynamic regret (35) ensures

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

Moreover, the argument in Besbes et al. (2015, Proposition 2) shows that $\texttt{term (b)} \leq 2\Delta V_T^f$. Combining the upper bounds for term (a) and term (b), we obtain

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

The optimal interval length is $\Delta_* \triangleq (DTA_T)^{1/3}(V_T^f)^{-2/3}$, 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, so we use the clipped version $\Delta_\dagger \triangleq \min\{\lceil \Delta_* \rceil, T\}$. We show that (36) 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 thus $\Delta_* \leq \Delta_\dagger \leq 2\Delta_*$,

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

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

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

- **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 (36).

The proof is completed by combining above three cases and the path-length bound (52). ∎

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 $\mathcal{P}$.

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 $\mathbf{u}_t = \mathbf{x}_t^* = \arg\min_{\mathbf{x} \in \mathcal{X}} f_t(\mathbf{x}) = \sigma \boldsymbol{\varepsilon}_t$. Denote by $\mathbf{x}_t$ the decision returned by any deterministic online algorithm $\mathcal{A}$. Then the expected dynamic regret is defined as $\mathbb{E}[\text{D-Regret}_T] = \mathbb{E}\left[\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t)\right]$ with expectation taken over the randomness of the online function f_t. Then, we show that $\mathbb{E}[\text{D-Regret}_T] \geq \sqrt{T/2} \cdot \mathbb{E}[\sqrt{(1 + F_T)(1 + P_T)}]$.

On one hand, noticing that $\mathbb{E}[\sigma\langle \mathbf{x}_t, \varepsilon_t\rangle] = 0$ and $\mathbb{E}[\sigma^2\|\boldsymbol{\varepsilon}_t\|_2^2] \geq d\sigma^2$ for any $t \geq 1$, we have

$$\mathbb{E}[\text{D-Regret}_T] = \sum_{t=1}^{T} \mathbb{E}[\|\mathbf{x}_t - \sigma\boldsymbol{\varepsilon}_t\|_2^2] = \sum_{t=1}^{T} \mathbb{E}[\|\mathbf{x}_t\|_2^2 + 2\sigma\langle\mathbf{x}_t, \varepsilon_t\rangle + \sigma^2\|\boldsymbol{\varepsilon}_t\|_2^2] \geq dT\sigma^2,$$

On the other hand, let $\boldsymbol{\delta}_t(i) = \varepsilon_t(i) - \varepsilon_{t-1}(i)$. We have

$$\mathbb{E}[P_T(\mathbf{u}_1, \ldots, \mathbf{u}_T)] = \sigma\sum_{t=2}^{T}\mathbb{E}\left[\sqrt{\sum_{i=1}^{d}\boldsymbol{\delta}_t^2(i)}\right] \leq \sigma\sum_{t=2}^{T}\sqrt{\sum_{i=1}^{d}\mathbb{E}\left[\boldsymbol{\delta}_t^2(i)\right]} \leq \sqrt{2d}T\sigma, \qquad (53)$$

where the first inequality is due to the Jensen's inequality and the second inequality is by the fact that $\mathbb{E}\left[\boldsymbol{\delta}_t^2(i)\right] = 2$ for any $t \in [T]$ and $i \in [d]$. The above equation leads to

$$\mathbb{E}\left[\sqrt{(1 + F_T)(1 + P_T)}\right] = \mathbb{E}[\sqrt{1 + P_T}] \leq \sqrt{1 + \mathbb{E}[P_T]} \leq (2d)^{\frac{1}{4}}\sqrt{T\sigma}.$$

By choosing $\sigma = 1$, we can ensure that $\mathbb{E}[\text{D-Regret}] \geq \sqrt{T/2} \cdot \mathbb{E}[\sqrt{(1 + F_T)(1 + P_T)}]$. We note that the choice of σ might lead to a violation of the bounded domain assumption, which can be easily fixed by rescaling. Then, the probabilistic argument implies that for any algorithm $\mathcal{A}$ there exists a sequence of online functions $f_1, \ldots, f_T$ and comparators $\{\mathbf{u}_t = \mathbf{x}_t^*\}_{t=1}^{T}$ such that $\text{D-Regret}_T \geq \sqrt{T/2} \cdot \sqrt{(1 + F_T)(1 + P_T)}$. This ends the proof. ∎

8. Experiments

This section provides empirical studies to validate the effectiveness of our algorithms.

Settings. We simulate the online environments as follows. The player *sequentially* receives the feature of an instance and is then required to make the prediction. We focus on the online regression problem, where at each round an instance $(\boldsymbol{\psi}_t, y_t)$ is received with $\boldsymbol{\psi}_t \in \boldsymbol{\Psi} \subseteq \mathbb{R}^d$ 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 $\boldsymbol{\psi}_t$ and is required to make the prediction by $\widehat{y}_t = \boldsymbol{\psi}_t^\top \mathbf{x}_t$ based on the learned model $\mathbf{x}_t \in \mathcal{X} \subseteq \mathbb{R}^d$; 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, $f_t : \mathcal{X} \mapsto \mathbb{R}$ with $f_t(\mathbf{x}) = \ell(y_t, \boldsymbol{\psi}_t^\top \mathbf{x})$. It can be verified that the online functions are convex, and satisfy the condition of non-negativity and smoothness.

(a) synthetic data
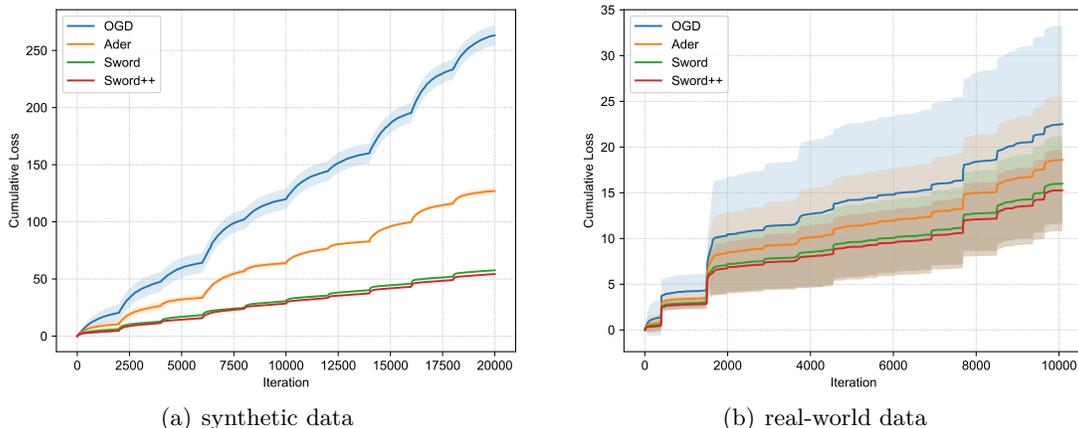
(b) real-world data

Figure 1: Performance comparisons of all algorithms in terms of cumulative loss.

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., $\mathfrak{B} = \{\boldsymbol{\psi} \in \mathbb{R}^d \mid \|\boldsymbol{\psi}\|_2 \leq \Gamma\}$; the associated label is set as $y_t = \boldsymbol{\psi}_t^\top \mathbf{x}_t^* + \varepsilon_t$, where ε_t is the random noise drawn from $[0, 0.1]$ and $\mathbf{x}_t^* \in \mathbb{R}^d$ is the underlying model. The underlying model $\mathbf{x}_t^*$ 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 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) (Gama et al., 2014; 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 performance of the following algorithms: (i) OGD (Zinkevich, 2003), online gradient descent, which is an OCO algorithm designed for static regret minimization; (ii) Ader (Zhang et al., 2018a), an OCO algorithm designed for optimizing dynamic regret yet with only problem-independent guarantee; (iii) Sword, the algorithm proposed in Section 4.2, which achieves problem-dependent dynamic regret guarantees requiring multiple gradients per iteration; and (iv) Sword++, the algorithm proposed in Section 4.3, which achieves the same dynamic regret with only one gradient query per round. The implementations of all algorithms are based on PyNOL package (Li et al., 2023a).

Results. We repeat the experiments 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 the presence of distribution changes. Moreover, Sword and Sword++ exploit the adaptivity of the problem instance and thus achieve more encouraging empirical behavior than Ader, which demonstrates the empirical effectiveness. Figure 2 reports the running time

(a) synthetic data

(b) real-world data
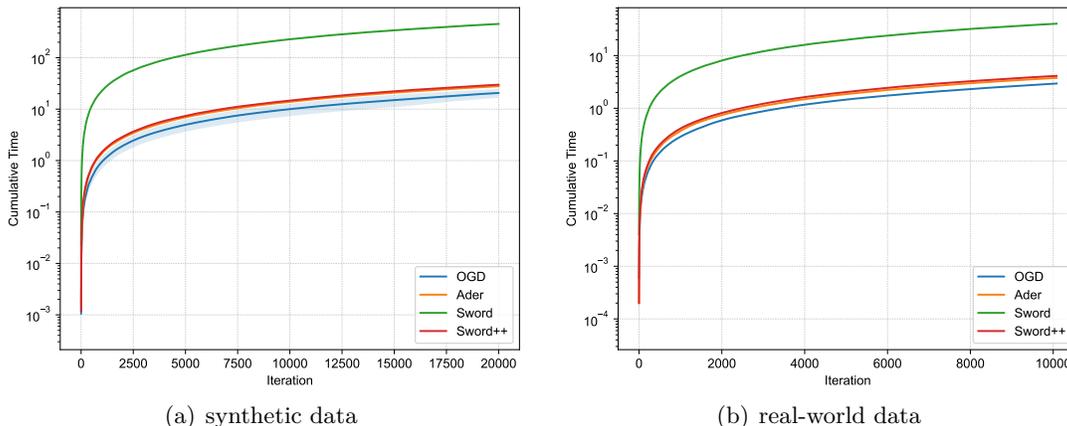
Figure 2: Performance comparisons of all algorithms in terms of running time (in seconds).

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 computationally inefficient. In summary, the empirical results validate the advantages of our proposed algorithms, notably showing that Sword++ behaves well and is computationally light.

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 online 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*. This framework emphasizes the importance of jointly using negative terms in the regret analysis and integrating additional correction terms in the algorithmic design, which facilitates effective collaboration between the meta and base 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 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 conjectures regarding the gradient-variation dynamic regret in Remark 7). However, a complete understanding requires refined problem-dependent lower bounds.

Acknowledgment

Appendix A. Proofs of Lemma 1 and Lemma 2

The versatility of OPTIMISTIC OMD makes it very general to derive many existing results in a unified view. We elucidate two implications of Theorem 1, which serve as proofs of Lemma 1 (dynamic regret of OEGD) and Lemma 2 (static regret of Optimistic Hedge).

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

It is well-known that $\psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$ is 1-strongly convex with respect to the Euclidean norm $\|\cdot\|_2$ and $\mathcal{D}_\psi(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2$. Thus, Theorem 1 implies

$$\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \eta \sum_{t=1}^{T} \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2 + \frac{1}{2\eta} \sum_{t=1}^{T} \left(\|\mathbf{u}_t - \widehat{\mathbf{x}}_t\|_2^2 - \|\mathbf{u}_t - \widehat{\mathbf{x}}_{t+1}\|_2^2 \right)$$
$$- \frac{1}{2\eta} \sum_{t=1}^{T} \left(\|\widehat{\mathbf{x}}_{t+1} - \mathbf{x}_t\|_2^2 + \|\widehat{\mathbf{x}}_t - \mathbf{x}_t\|_2^2 \right).$$

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

$$\sum_{t=1}^{T} \left(\|\mathbf{u}_t - \widehat{\mathbf{x}}_t\|_2^2 - \|\mathbf{u}_t - \widehat{\mathbf{x}}_{t+1}\|_2^2 \right) \leq \|\mathbf{u}_1 - \widehat{\mathbf{x}}_1\|_2^2 + \sum_{t=2}^{T} \left(\|\mathbf{u}_t - \widehat{\mathbf{x}}_t\|_2^2 - \|\mathbf{u}_{t-1} - \widehat{\mathbf{x}}_t\|_2^2 \right)$$

$$\leq \|\mathbf{u}_1 - \widehat{\mathbf{x}}_1\|_2^2 + \sum_{t=2}^{T} \|\mathbf{u}_t - \mathbf{u}_{t-1}\|_2 \|\mathbf{u}_t - \widehat{\mathbf{x}}_t + \mathbf{u}_{t-1} - \widehat{\mathbf{x}}_t\|_2 \leq D^2 + 2D \sum_{t=2}^{T} \|\mathbf{u}_t - \mathbf{u}_{t-1}\|_2.$$

We further evaluate the last term using the inequality of $a^2 + b^2 \geq (a+b)^2/2$ and obtain

$$\sum_{t=1}^{T} \left(\|\widehat{\mathbf{x}}_{t+1} - \mathbf{x}_t\|_2^2 + \|\widehat{\mathbf{x}}_t - \mathbf{x}_t\|_2^2 \right) \geq \sum_{t=2}^{T} \left(\|\widehat{\mathbf{x}}_t - \mathbf{x}_{t-1}\|_2^2 + \|\widehat{\mathbf{x}}_t - \mathbf{x}_t\|_2^2 \right) \geq \frac{1}{2} \sum_{t=2}^{T} \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2.$$

Hence, combining all the three inequalities, we get the following result:

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

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

$$\sum_{t=1}^{T} \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2 \leq G^2 + 2 \sum_{t=2}^{T} \left(\|\nabla f_t(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_t)\|_2^2 + \|\nabla f_{t-1}(\mathbf{x}_t) - \nabla f_{t-1}(\mathbf{x}_{t-1})\|_2^2 \right)$$

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

where the last step exploits L-smoothness of online functions. Therefore,

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

$$\leq \eta(G^2 + 2V_T) + \frac{1}{2\eta}(D^2 + 2DP_T),$$

where $\eta \leq 1/(4L)$ ensures the last term to be non-positive. This ends the proof. ∎

Next, by choosing the regularizer $\psi(\boldsymbol{p}) = \sum_{i=1}^{N} p_i \ln p_i$, the online function $f_t(\boldsymbol{p}) = \langle \boldsymbol{p}, \boldsymbol{\ell}_t \rangle$ and optimism $M_t = \boldsymbol{m}_t$, Optimistic 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 in prediction with expert advice. Theorem 1 implies the static regret for Optimistic Hedge algorithm by choosing comparators as a fixed one in the simplex.

Proof [of Lemma 2] It is well-known that the negative entropy $\psi(\boldsymbol{p}) = \sum_{i=1}^{N} p_i \ln p_i$ is 1-strongly convex with respect to $\|\cdot\|_1$ and $\mathcal{D}_{\psi}(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{N} p_i \ln(p_i/q_i)$. By choosing comparators as $\mathbf{e}_i$, Theorem 1 indicates that the regret $\sum_{t=1}^{T} (\langle \boldsymbol{p}_t, \boldsymbol{\ell}_t \rangle - \ell_{t,i})$ is bounded by

$$\varepsilon \sum_{t=1}^{T} \|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2 + \frac{1}{\varepsilon} \sum_{t=1}^{T} \left(\mathcal{D}_{\psi}(\mathbf{e}_i, \widehat{\boldsymbol{p}}_t) - \mathcal{D}_{\psi}(\mathbf{e}_i, \widehat{\boldsymbol{p}}_{t+1}) \right) - \frac{1}{\varepsilon} \sum_{t=1}^{T} \left(\mathcal{D}_{\psi}(\widehat{\boldsymbol{p}}_{t+1}, \boldsymbol{p}_t) + \mathcal{D}_{\psi}(\boldsymbol{p}_t, \widehat{\boldsymbol{p}}_t) \right).$$

The second term on the right hand side exhibits a telescoping structure, and thus

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

By Pinsker's inequality $\mathcal{D}_{\psi}(\boldsymbol{p}, \boldsymbol{q}) = \mathrm{KL}(\boldsymbol{p}, \boldsymbol{q}) \geq \frac{1}{2} \|\boldsymbol{p} - \boldsymbol{q}\|_1^2$, we have

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

where the last inequality is got by regrouping the sum and applying triangle inequality. Combining all above inequalities ends the proof. ∎

The negative term in the regret bound (13) of Lemma 2 is very essential, which is quite useful in a variety of problems requiring adaptive bounds. Our analysis is based on the unified view of Optimistic OMD (Theorem 1), and is much simpler than the original proof of Syrgkanis et al. (2015) using the mathematical induction from the lens of FTRL.

Appendix B. Adaptive Learning Rate Version

In the main text, our proposed algorithms, Sword and Sword++, deliberately employ a *fixed* learning rate for the meta-algorithm. This choice is made to simplify the presentation and the regret analysis, thereby helping reader to better understand the key idea of facilitating collaborations in meta and base levels, as highlighted in Remark 8 of the analysis.

Nevertheless, the meta-algorithm learning rate tuning requires the knowledge of gradient variation $V_T = \sum_{t=2}^{T} \sup_{\mathbf{x} \in \mathcal{X}} \|\nabla f_t(\mathbf{x}) - \nabla f_{t-1}(\mathbf{x})\|_2^2$ (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++), which is not desired. This section demonstrates an adaptive version using the self-confident tuning framework (Auer et al., 2002) such that the meta-algorithm does not require such information ahead of time.[4]

We first extend the collaborative online ensemble framework in Section 5 to an adaptive version, and subsequently 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,

$$p_{t+1,i} \propto \exp\left(-\varepsilon_t \Big(\sum_{s=1}^{t} \ell_{s,i} + m_{t+1,i}\Big)\right), \tag{54}$$

where the loss vector $\boldsymbol{\ell}_t$ and $\boldsymbol{m}_t$ share the same configurations as (28). For any i-th base-algorithm, we use the same update rule as the fixed learning rate version

$$\mathbf{x}_{t,i} = \Pi_{\mathcal{X}}\left[\hat{\mathbf{x}}_{t,i} - \eta_i M_t\right], \quad \hat{\mathbf{x}}_{t+1,i} = \Pi_{\mathcal{X}}\left[\hat{\mathbf{x}}_{t,i} - \eta_i \nabla f_t(\mathbf{x}_t)\right], \tag{55}$$

Then, we can generate the prediction 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*

$$\varepsilon_t = \min\left\{\bar{\varepsilon}, \sqrt{\frac{\ln N}{D^2 \sum_{s=1}^{t} \|\nabla f_s(\mathbf{x}_s) - M_s\|_2^2}}\right\}, \tag{56}$$

---

4. 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.

we have the following dynamic regret bound for the decisions specified by (54) *and* (55) *against any comparators* $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$,

$$\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq 4\sqrt{D^2 \ln N A_T} + 2\sqrt{(D^2 + 2DP_T)A_T} + \frac{\ln N}{\bar{\varepsilon}} + 2\bar{\varepsilon}D^2\widetilde{G}^2$$

$$+ \frac{2(D^2 + 2DP_T)}{\bar{\eta}} + \left(\lambda - \frac{1}{4\bar{\eta}}\right)S_{x,i} - \frac{1}{4\bar{\varepsilon}}S_p - \lambda S_{\text{mix}}. \tag{57}$$

In above, $\widetilde{G} = \max_{t\in[T]}\|\nabla f_t(\mathbf{x}) - M_t\|_2$, $A_T = \sum_{t=1}^{T}\|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2$, $P_T = \sum_{t=2}^{T}\|\mathbf{u}_{t-1} - \mathbf{u}_t\|_2$, $S_{x,i} = \sum_{t=2}^{T}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$, $S_p = \sum_{t=2}^{T}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2$, *and* $S_{\text{mix}} = \sum_{t=2}^{T}\sum_{i=1}^{N} p_{t,i}\|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2$ *are base stability, meta stability, and mixed stability.*

Remark 9 (Optimistic Hedge with Time-varying Learning Rates)**.** We remark that, in the fixed learning rate case, one can show that the Optimistic Hedge (24) is identical to Optimistic OMD with the negative-entropy regularizer. However, the adaptive learning rate version (54) can only be interpreted as an 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. While this can be remedied by the dual stabilization (Fang et al., 2020), we just use FTRL for simplicity. ¶

Proof [of Theorem 9] The proof is almost the same 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 (54). Specifically, since (54) is identical to Optimistic FTRL $\boldsymbol{p}_{t+1} = \arg\min_{\boldsymbol{p}\in\Delta_N}\langle\boldsymbol{p}, \sum_{s=1}^{t}\boldsymbol{\ell}_s + \boldsymbol{m}_{t+1}\rangle + \psi_{t+1}(\boldsymbol{p})$ with regularizer $\psi_{t+1}(\boldsymbol{p}) = \frac{1}{\varepsilon_t}(\sum_{i=1}^{N} p_i \ln p_i + \ln N)$,[5] a direct application of Orabona (2019, Theorem 7.35) leads to the following lemma.

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* (54)*) to any expert* $i \in [N]$ *satisfies*

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

where $\psi_{t+1}(\boldsymbol{p}) = \frac{1}{\varepsilon_t}(\sum_{i=1}^{N} p_i \ln p_i + \ln N)$ *is the regularizer.*

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

$$\sum_{t=1}^{T}\langle\boldsymbol{p}_t, \boldsymbol{\ell}_t\rangle - \sum_{t=1}^{T}\ell_{t,i}$$

$$\leq \frac{\ln N}{\varepsilon_T} + \sum_{t=1}^{T}\varepsilon_{t-1}\|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty^2 + \sum_{t=1}^{T}\frac{1}{4\varepsilon_{t-1}}\|\boldsymbol{p}_t - \boldsymbol{p}_{t+1}\|_1^2 - \sum_{t=1}^{T}\frac{1}{2\varepsilon_{t-1}}\|\boldsymbol{p}_t - \boldsymbol{p}_{t+1}\|_1^2$$

$$\leq \frac{\ln N}{\varepsilon_T} + D^2\sum_{t=1}^{T}\varepsilon_{t-1}\|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2 - \sum_{t=2}^{T}\frac{1}{4\varepsilon_{t-1}}\|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2$$

---

5. 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.

$$\leq 2\bar{\varepsilon}D^2G^2 + \frac{\ln N}{\bar{\varepsilon}} + 4\sqrt{D^2 \ln N \sum_{t=1}^{T} \|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2} - \sum_{t=2}^{T} \frac{1}{4\varepsilon_{t-1}} \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2,$$

where the second inequality is due to the Hölder's inequality $\langle \boldsymbol{\ell}_t - \boldsymbol{m}_t, \boldsymbol{p}_t - \boldsymbol{p}_{t+1} \rangle \leq \|\boldsymbol{\ell}_t - \boldsymbol{m}_t\|_\infty \|\boldsymbol{p}_t - \boldsymbol{p}_{t+1}\|_1$ and the fact that $ab \leq \varepsilon_{t-1}a^2 + \frac{b^2}{4\varepsilon_{t-1}}$ 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 $\ln N/\varepsilon_T \leq \ln N/\bar{\varepsilon} + \sqrt{D^2 \ln N \sum_{t=1}^{T}\|\nabla f_t(\mathbf{x}) - M_t\|_2^2}$ by learning rate configuration (56) and Lemma 11, which provides a clipped version of the self-confident tuning.

By the same meta-regret analysis in the proof of Theorem 5, see (48), we have

$$\texttt{meta-regret} \leq 4\sqrt{D^2(\ln N)A_T} + \frac{\ln N}{\bar{\varepsilon}} + 2\bar{\varepsilon}D^2G^2 - \frac{1}{4\bar{\varepsilon}} \sum_{t=2}^{T} \|\boldsymbol{p}_t - \boldsymbol{p}_{t-1}\|_1^2$$

$$- \lambda \sum_{t=1}^{T} \sum_{i=1}^{N} p_{t,i} \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2 + \lambda \sum_{t=1}^{T} \|\mathbf{x}_{t,i} - \mathbf{x}_{t-1,i}\|_2^2, \tag{58}$$

which holds from any base-algorithm $i \in [N]$. 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 as $\texttt{base-regret} \leq 2\sqrt{(D^2 + 2DP_T)A_T} + \frac{2(D^2 + 2DP_T)}{\bar{\eta}}$. Combining the meta-regret and the base-regret of the i^*-th base-learner yields the result in (57). ∎

B.2 Adaptive Version of Sword++

We show that the adaptive learning rate version of the framework (54) and (55) 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 (up to constants) 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*

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

with $M_t = \nabla f_{t-1}(\mathbf{x}_{t-1})$ for $t \geq 2$ ($M_1 = \mathbf{0}$). Then, decisions specified by (54) and (55) satisfy that for any comparators $\mathbf{u}_1, \ldots, \mathbf{u}_T \in \mathcal{X}$,

$$\sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}_t) \leq \mathcal{O}\left(\sqrt{(1 + P_T + \min\{V_T, F_T\})(1 + P_T)}\right). \tag{59}$$

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) up to constant factors. Thus, the same arguments in the proof of Theorem 6 lead to the desired result. ∎

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

$$f(\mathbf{x}_k) - f(\mathbf{u}) \leq \mathcal{D}_\psi(\mathbf{u}, \mathbf{x}_{k-1}) - \mathcal{D}_\psi(\mathbf{u}, \mathbf{x}_k) - \mathcal{D}_\psi(\mathbf{x}_k, \mathbf{x}_{k-1}). \tag{60}$$

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 $\|\mathbf{x}_* - \mathbf{x}'_*\| \leq \|\mathbf{a} - \mathbf{a}'\|_*$.*

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 $\|\nabla f(\mathbf{x})\|_2 \leq \sqrt{4Lf(\mathbf{x})}, \ \forall \mathbf{x} \in \mathcal{X}$.*

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

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 ends the proof. ∎

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

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

Lemma 10 (Lemma 3.5 of Auer et al. (2002))**.** *Let $a_1, a_2, \ldots, a_T$ and δ be non-negative real numbers. Then, it holds that $\sum_{t=1}^{T} \frac{a_t}{\sqrt{\delta + \sum_{s=1}^{t} a_s}} \leq 2\sqrt{\delta + \sum_{t=1}^{T} a_t}$, 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 $c_t = \min\left\{\bar{c}, \sqrt{\frac{b}{\sum_{s=1}^{t} a_s}}\right\}$ and $c_0 = \bar{c}$. Then, we have*

$$\sum_{t=1}^{T} c_{t-1} a_t \leq 2\bar{c}B + 4\sqrt{b\sum_{t=1}^{T} a_t}. \tag{61}$$

Proof This proof shares the same spirit with that of Pogodin and Lattimore (2019, Lemma 4.8). We assume $\sum_{t=1}^{T} a_t > B$, 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 $T' = \min\{t \in [T] \mid \sum_{s=1}^{t-1} a_s \geq B\}$. 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, $\sum_{t=1}^{T'-1} c_{t-1} a_t = \sum_{t=1}^{T'-2} c_{t-1} a_t + c_{T'-2} a_{T'-1} \leq 2\bar{c}B$. For the second term,

$$\sum_{t=T'}^{T} c_{t-1} a_t \leq \sum_{t=T'}^{T} \frac{a_t \sqrt{b}}{\sqrt{\sum_{s=1}^{t-1} a_s}} \leq \sum_{t=T'}^{T} \frac{a_t \sqrt{b}}{\sqrt{\frac{1}{2}\sum_{s=1}^{t} a_s}} \leq \sum_{t=1}^{T} \frac{a_t \sqrt{b}}{\sqrt{\frac{1}{2}\sum_{s=1}^{t} a_s}} \leq 3\sqrt{b\sum_{t=1}^{T} a_t},$$

where the first inequality is by the definition of c_t and the second inequality is due to $\sum_{s=1}^{t} a_s \leq B + \sum_{s=1}^{t-1} a_s \leq 2\sum_{s=1}^{t-1} a_s$ for all $t \geq T'$. The last inequality comes from Lemma 10. We complete the proof by combining the two terms. ■

References

Jacob D. Abernethy, Peter L. Bartlett, Alexander Rakhlin, and Ambuj 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.

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

Zeyuan Allen-Zhu, Sébastien Bubeck, and Yuanzhi 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.

Chamy Allenberg, Peter Auer, László Györfi, and György 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.

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

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

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

Dheeraj Baby and Yu-Xiang Wang. Optimal dynamic regret in proper online learning with strongly convex losses and beyond. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1805–1845, 2022a.

Dheeraj Baby and Yu-Xiang Wang. Optimal dynamic regret in LQR control. In *Advances in Neural Information Processing Systems 35 (NeurIPS)*, pages 24879–24892, 2022b.

Yong Bai, Yu-Jie Zhang, Peng Zhao, Masashi Sugiyama, and Zhi-Hua Zhou. Adapting to online label shift with provable guarantees. In *Advances in Neural Information Processing Systems 35 (NeurIPS)*, pages 29960–29974, 2022.

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

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

Nicolò Cesa-Bianchi, Yoav Freund, David Haussler, David P. Helmbold, Robert E. Schapire, and Manfred K. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.

Nicolò Cesa-Bianchi, Yishay Mansour, and Gilles 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.

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

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

Sijia Chen, Wei-Wei Tu, Peng Zhao, and Lijun Zhang. Optimistic online mirror descent for bridging stochastic and adversarial online convex optimization. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, pages 5002–5035, 2023a.

Sijia Chen, Yu-Jie Zhang, Wei-Wei Tu, Peng Zhao, and Lijun Zhang. Optimistic online mirror descent for bridging stochastic and adversarial online convex optimization. *ArXiv preprint*, arXiv: 2010.10473, 2023b.

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

Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Proceedings of the 25th Conference On Learning Theory (COLT)*, pages 6.1–6.20, 2012.

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

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

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

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

Ashok Cutkosky. Better full-matrix regret via parameter-free online learning. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, 2020b.

Ashok Cutkosky. Lecture notes for ec525: Optimization for machine learning. 2023. URL https://optmlclass.github.io/notes/optforml_notes.pdf.

Ashok Cutkosky and Francesco 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.

John C. Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. In *Proceedings of the 23rd Conference on Learning Theory (COLT)*, pages 257–269, 2010.

Huang Fang, Nick Harvey, Victor S. Portella, and Michael 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.

Yingjie Fei, Zhuoran Yang, Zhaoran Wang, and Qiaomin Xie. Dynamic regret of policy optimization in non-stationary environments. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, pages 6743–6754, 2020.

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

Yoav Freund and Robert 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.

João Gama, Indre Zliobaite, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4): 44:1–44:37, 2014.

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

Keegan Harris, Ioannis Anagnostides, Gabriele Farina, Mikhail Khodak, Zhiwei Steven Wu, and Tuomas Sandholm. Meta-learning in games. In *Proceedings of the 10th International Conference on Learning Representations (ICLR)*, 2022.

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

Elad Hazan and Satyen 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.

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

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

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

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

Pooria Joulani, András György, and Csaba Szepesvári. A modular analysis of adaptive (non-)convex optimization: Optimism, composite objectives, variance reduction, and variational bounds. *Theoretical Computer Science*, 808:108–138, 2020.

Chung-Wei Lee, Haipeng Luo, Chen-Yu Wei, and Mengxiao 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, 2020a.

Chung-Wei Lee, Haipeng Luo, and Mengxiao 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, 2020b.

Long-Fei Li, Peng Zhao, Yan-Feng Xie, Lijun Zhang, and Zhi-Hua Zhou. PyNOL: : A Python Package for Non-stationary Online Learning, 2023a. URL `https://github.com/li-lf/PyNOL`.

Long-Fei Li, Peng Zhao, and Zhi-Hua Zhou. Dynamic regret of adversarial linear mixture mdps. In *Advances in Neural Information Processing Systems 36 (NeurIPS)*, pages 60685–60711, 2023b.

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

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

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

H. Brendan McMahan. A survey of algorithms and analysis for adaptive online learning. *Journal of Machine Learning Research*, 18:90:1–90:50, 2017.

H. Brendan McMahan and Matthew J. Streeter. Adaptive bound optimization for online convex optimization. In *Proceedings of the 23rd Conference on Learning Theory (COLT)*, pages 244–256, 2010.

Aryan Mokhtari, Shahin Shahrampour, Ali Jadbabaie, and Alejandro 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.

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

Francesco Orabona. A Modern Introduction to Online Learning. *ArXiv preprint*, arXiv: 1912.13213, 2019.

Roman Pogodin and Tor 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.

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

Tim Roughgarden. *Beyond the Worst-Case Analysis of Algorithms*. Cambridge University Press, 2021.

Sarah Sachs, Hedi Hadiji, Tim van Erven, and Cristóbal A Guzmán. Between stochastic and adversarial online convex optimization: Improved regret bounds via smoothness. In *Advances in Neural Information Processing Systems 35 (NeurIPS)*, pages 691–702, 2022.

Shai Shalev-Shwartz. *Online Learning: Theory, Algorithms and Applications*. PhD thesis, The Hebrew University of Jerusalem, 2007.

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

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

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

Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E. Schapire. Fast convergence of regularized learning in games. In *Advances in Neural Information Processing Systems 28 (NIPS)*, pages 2989–2997, 2015.

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

Tim van Erven, Wouter M. Koolen, and Dirk van der Hoeven. Metagrad: Adaptation using multiple learning rates in online learning. *Journal of Machine Learning Research*, 22:161:1–161:61, 2021.

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

Chen-Yu Wei, Yi-Te Hong, and Chi-Jen Lu. Tracking the best expert in non-stationary stochastic environments. In *Advances in Neural Information Processing Systems 29 (NIPS)*, pages 3972–3980, 2016.

Yu-Hu Yan, Peng Zhao, and Zhi-Hua Zhou. Fast rates in time-varying strongly monotone games. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, pages 39138–39164, 2023.

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

Tianbao Yang, Lijun Zhang, Rong Jin, and Jinfeng 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.

Jianjun Yuan and Andrew 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.

Lijun Zhang and Zhi-Hua 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.

Lijun Zhang, Tianbao Yang, Rong Jin, and Xiaofei 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)*, pages 1121–1129, 2013.

Lijun Zhang, Tianbao Yang, Jinfeng Yi, Rong Jin, and Zhi-Hua Zhou. Improved dynamic regret for non-degenerate functions. In *Advances in Neural Information Processing Systems 30 (NIPS)*, pages 732–741, 2017.

Lijun Zhang, Shiyin Lu, and Zhi-Hua Zhou. Adaptive online learning in dynamic environments. In *Advances in Neural Information Processing Systems 31 (NeurIPS)*, pages 1330–1340, 2018a.

Lijun Zhang, Tianbao Yang, Rong Jin, and Zhi-Hua Zhou. Dynamic regret of strongly adaptive methods. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pages 5877–5886, 2018b.

Lijun Zhang, Tie-Yan Liu, and Zhi-Hua Zhou. Adaptive regret of convex and smooth functions. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 7414–7423, 2019.

Lijun Zhang, Shiyin Lu, and Tianbao 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.

Lijun Zhang, Wei Jiang, Shiyin Lu, and Tianbao Yang. Revisiting smoothed online learning. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, pages 13599–13612, 2021.

Mengxiao Zhang, Peng Zhao, Haipeng Luo, and Zhi-Hua Zhou. No-regret learning in time-varying zero-sum games. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, pages 26772–26808, 2022a.

Yu-Jie Zhang, Peng Zhao, and Zhi-Hua 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.

Yu-Jie Zhang, Zhen-Yu Zhang, Peng Zhao, and Masashi Sugiyama. Adapting to continuous covariate shift via online density ratio estimation. In *Advances in Neural Information Processing Systems 36 (NeurIPS)*, pages 29074–29113, 2023.

Zhiyu Zhang, Ashok Cutkosky, and Ioannis Ch. Paschalidis. Adversarial tracking control via strongly adaptive online learning with memory. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 8458–8492, 2022b.

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

Peng Zhao and Lijun 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.

Peng Zhao, Lijun Zhang, Yuan Jiang, and Zhi-Hua 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.

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

Peng Zhao, Guanghui Wang, Lijun Zhang, and Zhi-Hua Zhou. Bandit convex optimization in non-stationary environments. *Journal of Machine Learning Research*, 22(125):1 – 45, 2021a.

Peng Zhao, Xinqiang Wang, Siyu Xie, Lei Guo, and Zhi-Hua Zhou. Distribution-free one-pass learning. *IEEE Transaction on Knowledge and Data Engineering*, 33:951–963, 2021b.

Peng Zhao, Yu-Jie Zhang, Lijun Zhang, and Zhi-Hua Zhou. Adaptivity and non-stationarity: Problem-dependent dynamic regret for online convex optimization. *ArXiv preprint*, arXiv:2112.14368, 29 Dec 2021c.

Peng Zhao, Long-Fei Li, and Zhi-Hua Zhou. Dynamic regret of online Markov decision processes. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, pages 26865–26894, 2022a.

Peng Zhao, Yu-Xiang Wang, and Zhi-Hua 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, 2022b.

Peng Zhao, Yan-Feng Xie, Lijun Zhang, and Zhi-Hua Zhou. Efficient methods for non-stationary online learning. In *Advances in Neural Information Processing Systems 35 (NeurIPS)*, pages 11573–11585, 2022c.

Peng Zhao, Yu-Hu Yan, Yu-Xiang Wang, and Zhi-Hua Zhou. Non-stationary online learning with memory and non-stochastic control. *Journal of Machine Learning Research*, 24(206): 1–70, 2023.

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

Zhi-Hua Zhou. Open-environment machine learning. *National Science Review*, 9(8): nwac123, 2022.

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