Online Learning with Primary and Secondary Losses

Avrim Blum
Toyota Technological Institute at Chicago
avrim@ttic.edu

Han Shao
Toyota Technological Institute at Chicago
han@ttic.edu

Abstract

We study the problem of online learning with primary and secondary losses. For example, a recruiter making decisions of which job applicants to hire might weigh false positives and false negatives equally (the primary loss) but the applicants might weigh false negatives much higher (the secondary loss). We consider the following question: Can we combine "expert advice" to achieve low regret with respect to the primary loss, while at the same time performing *not much worse than the worst expert* with respect to the secondary loss? Unfortunately, we show that this goal is unachievable without any bounded variance assumption on the secondary loss. More generally, we consider the goal of minimizing the regret with respect to the primary loss and bounding the secondary loss by a linear threshold. On the positive side, we show that running any switching-limited algorithm can achieve this goal if all experts satisfy the assumption that the secondary loss does not exceed the linear threshold by $o(T)$ for any time interval. If not all experts satisfy this assumption, our algorithms can achieve this goal given access to some external oracles which determine when to deactivate and reactivate experts.

1 Introduction

The online learning problem has been studied extensively in the literature and used increasingly in many applications including hiring, advertising and recommender systems. One classical problem in online learning is prediction with expert advice, in which a decision maker makes a sequence of T decisions with access to K strategies (also called "experts"). At each time step, the decision maker observes a scalar-valued loss of each expert. The standard objective is to perform as well as the best expert in hindsight. For example, a recruiter (the decision maker) sequentially decides which job applicants to hire with the objective of minimizing errors (of hiring an unqualified applicant and rejecting a qualified one). However, this may give rise to some social concerns since the decision receiver has a different objective (getting a job) which does not receive any attention. This problem can be modeled as an online learning problem with the primary loss (for the decision maker) and secondary loss (for the decision receiver). Taking the social impact into consideration, we ask the following question:

Can we achieve low regret with respect to the primary loss, while performing not much worse than the worst expert with respect to the secondary loss?

Unfortunately, we answer this question negatively. More generally, we consider a bicriteria goal of minimizing the regret to the best expert with respect to the primary loss while minimizing the regret to a linear threshold cT with respect to the secondary loss for some c. When the value of c is set to the average secondary loss of the worst expert with respect to the secondary loss, the objective reduces to no-regret for the primary loss while performing no worse than the worst expert with respect to the secondary loss. Other examples, e.g., the average secondary loss of the worst expert with respect to the secondary loss among the experts with optimal primary loss, lead to different criteria of the secondary loss. Therefore, with the notion of regret to the linear threshold, we are able to study a more general goal. Based on this goal, we pose the following two questions:

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

1. If all experts have secondary losses no greater than $cT + o(T)$ for some c, can we achieve no-regret (compete comparably to the best expert) for the primary loss while achieving secondary loss no worse than $cT + o(T)$?
2. If we are given some external oracles to deactivate some "bad" experts with unsatisfactory secondary loss, can we perform as well as each expert with respect to the primary loss during the time they are active while achieving secondary loss no worse than $cT + o(T)$?

These two questions are trivial in the i.i.d. setting as we can learn the best expert with respect to the primary loss within $O(\log(T))$ rounds and then we just need to follow the best expert. In this paper, we focus on answering these two questions in the adversarial online setting.

1.1 Contributions

An impossibility result without a bounded variance assumption We show that without any constraints on the variance of the secondary loss, even if all experts have secondary loss no greater than cT, achieving no-regret with respect to the primary loss and bounding secondary loss by $cT + O(T)$ is still unachievable. This answers our motivation question that it is impossible to achieve low regret with respect to the primary loss, while performing not much worse than the worst expert with respect to the secondary loss. This result explains why minimizing one loss while bounding another is non-trivial and applying existing algorithms for scalar-valued losses after scalarizing primary and secondary losses does not work. We propose an assumption on experts that the secondary loss of the expert during any time interval does not exceed cT by $O(T^\alpha)$ for some $\alpha \in [0, 1)$.

Then we study the problem in two scenarios, a "good" one in which all experts satisfy this assumption and a "bad" one in which experts partially satisfy this assumption and we are given access to an external oracle to deactivate and reactivate experts.

Our results in the "good" scenario In the "good" scenario, we show that running an algorithm with limited switching rounds such as Follow the Lazy Leader [Kalai and Vempala, 2005] and Shrinking Dartboard (SD) [Geulen et al., 2010] can achieve both regret to the best with respect to the primary loss and regret to cT with respect to the secondary loss at $O(T^{\frac{1+\alpha}{2}})$. We also provide a lower bound of $\Omega(T^\alpha)$.

From another perspective, we relax the "good" scenario constraint by introducing adaptiveness to the secondary loss and constraining the variance of the secondary loss between any two switchings for any algorithm instead of that of any expert. We show that in this weaker version of "good" scenario, the upper bound of running switching-limited algorithms matches the lower bound at $\Theta(T^{\frac{1+\alpha}{2}})$.

Our results in the "bad" scenario In the "bad" scenario, we assume that we are given an external oracle to determine which experts to deactivate as they do not satisfy the bounded variance assumption. We study two oracles here. One oracle deactivates the experts which do not satisfy the bounded variance assumption once detecting and never reactivates them. The other one reactivates those inactive experts at fixed rounds. In this framework, we are limited to select among the active experts at each round and we adopt a more general metric, sleeping regret, to measure the performance of the primary loss. We provide algorithms for the two oracles with theoretical guarantees on the sleeping regrets with respect to the primary loss and the regret to cT with respect to the secondary loss.

1.2 Related work

One line of closely related work is online learning with multi-objective criterion. A bicriteria setting which examines not only the regret to the best expert but also the regret to a fixed mixture of all experts is investigated by Even-Dar et al. [2008], Kapralov and Panigrahy [2011], Sani et al. [2014]. The objective by Even-Dar et al. [2009] is to learn an optimal static allocation over experts with respect to a global cost function. Another multi-objective criterion called the Pareto regret frontier studied by Koolen [2013] examines the regret to each expert. Different from our work, all these criteria are studied in the setting of scalar-valued losses. The problem of multiple loss functions is studied by Chernov and Vovk [2009] under a heavy geometric restriction on loss functions. For vector losses, one fundamental concept is the Pareto front, the set of feasible points in which none can be dominated by any other point given several criteria to be optimized [Hwang and Masud, 2012, Auer et al., 2016]. However, the Pareto front contains unsatisfactory solutions such as the one minimizing the secondary loss, which implies that learning the Pareto front can not achieve our goal. Another classical concept is approachability, in which a learner aims at making the averaged vector loss converge to a pre-specified target set [Blackwell et al., 1956, Abernethy et al., 2011]. However,

we show that our fair solution is unapproachable without additional bounded variance assumptions. Approachability to an expansion target set based on the losses in hindsight is studied by Mannor et al. [2014]. However, the expansion target set is not guaranteed to be meet our criteria. Multi-objective criterion has also been studied in multi-armed bandits [Turgay et al., 2018].

2 Model

We consider the adversarial online learning setting with a set of K experts $\mathcal{H} = \{1, \dots, K\}$. At round $t = 1, 2, \dots, T$, given an active expert set $\mathcal{H}_t \subseteq \mathcal{H}$, an online learner $\mathcal{A}$ computes a probability distribution $p_t \in \Delta_K$ over $\mathcal{H}$ with support only over $\mathcal{H}_t$ and selects one expert from p_t. Simultaneously an adversary selects two loss vectors $\ell_t^{(1)}, \ell_t^{(2)} \in [0,1]^K$, where $\ell_{t,h}^{(1)}$ and $\ell_{t,h}^{(2)}$ are the primary and secondary losses of expert $h \in \mathcal{H}$ at time t. Then $\mathcal{A}$ observes the loss vector and incurs expected losses $\ell_{t,\mathcal{A}}^{(i)} = p_t^\top \ell_t^{(i)}$ for $i \in \{1, 2\}$. Let $L_{T,h}^{(i)} = \sum_{t=1}^T \ell_{t,h}^{(i)}$ denote the loss of expert h and $L_{T,\mathcal{A}}^{(i)} = \sum_{t=1}^T p_t^\top \ell_t^{(i)}$ denote the loss of algorithm $\mathcal{A}$ for $i \in \{1, 2\}$ during the first T rounds. We will begin by focusing on the case that the active expert set $\mathcal{H}_t = \mathcal{H}$.

2.1 Regret notions

Traditionally, the regret (to the best) is used to measure the scalar-valued loss performance of a learner, which compares the loss of the learner and the best expert in hindsight. Similar to Even-Dar et al. [2008], we adopt the regret notion of $\mathcal{A}$ with respect to the primary loss as

$$\mathrm{Reg}^{(1)} \triangleq \max\left(L_{T,\mathcal{A}}^{(1)} - \min_{h \in \mathcal{H}} L_{T,h}^{(1)}, 1\right) .$$

We introduce another metric for the secondary loss called *regret to* cT for some $c \in [0,1]$, which compares the secondary loss of the learner with a linear term cT,

$$\mathrm{Reg}_c^{(2)} \triangleq \max\left(L_{T,\mathcal{A}}^{(2)} - cT, 1\right) .$$

Sleeping experts are developed to model the problem in which not all experts are available at all times [Blum, 1997, Freund et al., 1997]. At each round, each expert $h \in \mathcal{H}$ decides to be active or not and then a learner can only select among the active experts, i.e. have non-zero probability $p_{t,h}$ over the active experts. The goal is to perform as well as h^* in the rounds where h^* is active for all $h^* \in \mathcal{H}$. We denote by $\mathcal{H}_t$ the set of active experts at round t. The sleeping regret for the primary loss with respect to expert h^* is defined as

$$\mathrm{SleepReg}^{(1)}(h^*) \triangleq \max\left(\sum_{t: h^* \in \mathcal{H}_t} \sum_{h \in \mathcal{H}_t} p_{t,h} \ell_{t,h}^{(1)} - \sum_{t: h^* \in \mathcal{H}_t} \ell_{t,h^*}^{(1)}, 1\right) .$$

The sleeping regret notion we adopt here is different from the regret to the best ordering of experts in the sleeping expert setting of Kleinberg et al. [2010]. Since achieving the optimal regret bound in Kleinberg's setting is computationally hard [Kanade and Steinke, 2014], we focus on the sleeping regret notion defined above.

2.2 Assumptions

Following a standard terminology, we call an adversary oblivious if her selection is independent of the learner's actions. Otherwise, we call the adversary adaptive. First, we assume that the primary loss is oblivious. This is a common assumption in the online learning literature and this assumption holds throughout the paper.

Assumption 1. *The primary losses* $\{\ell_t^{(1)}\}_{t \in [T]}$ *are oblivious.*

For an expert $h \in \mathcal{H}$, we propose a bounded variance assumption on her secondary loss: the average secondary loss for any interval does not exceed c much. More formally, the assumption is described as below.

Assumption 2. *For some given* $c, \delta, \alpha \in [0,1]$ *and for all expert* $h \in \mathcal{H}$*, for any* $T_1, T_2 \in [T]$ *with* $T_1 \le T_2$*,*

$$\sum_{t=T_1}^{T_2} (\ell_{t,h}^{(2)} - c) \le \delta T^\alpha .$$

We show that such a bounded variance assumption is necessary in Section 3. We call a scenario "good" if all experts satisfy assumption 2. Otherwise, we call the scenario "bad". This "good" constraint can be relaxed by introducing adaptiveness to the secondary loss. We have a relaxed version of the "good" scenario in which the average secondary loss between any two switchings does not exceed c much for any algorithm. More formally,

Assumption 2′. *For some given $c, \delta, \alpha \in [0, 1]$, for any algorithm $\mathcal{A}$, let $\mathcal{A}_t \in \mathcal{H}$ denote the selected expert at round t. For any expert $h \in \mathcal{H}$ and $T_1 \in [T]$ such that $\mathcal{A}_{T_1} = h$ and $\mathcal{A}_{T_1-1} \neq h$ (where $\mathcal{A}_{T+1} = T\mathcal{A}_0 = 0$ for notation simplicity), we have*

$$\sum_{\tau=T_1}^{\min_{t>T_1:\mathcal{A}_t\neq h} t-1} \left(\ell_{\tau,h}^{(2)} - c\right) \leq \delta T^{\alpha} .$$

In the "good" scenario, the active expert set $\mathcal{H}_t = \mathcal{H}$ for all rounds and the goal is minimizing both $\mathrm{Reg}^{(1)}$ and $\mathrm{Reg}_c^{(2)}$. In the "bad" scenario, we consider that we are given an oracle which determines $\mathcal{H}_t$ at each round and the goal is minimizing $\mathrm{SleepReg}^{(1)}(h^*)$ for all $h^* \in \mathcal{H}$ and $\mathrm{Reg}_c^{(2)}$.

3 Impossibility result without any bounded variance assumption

In this section, we show that without any additional assumption on the secondary loss, even if all experts have secondary loss no greater than cT for some $c \in [0, 1]$, there exists an adversary such that any algorithm incurs $\mathbb{E}[\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)})] = \Omega(T)$.

Theorem 1. *Given a fixed expert set $\mathcal{H}$, there exists an adversary such that any algorithm will incur $\mathbb{E}[\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)})] = \Omega(T)$ with $c = \max_{h\in\mathcal{H}} L_{T,h}^{(2)}/T$, where the expectation is taken over the randomness of the adversary.*

Proof. To prove this theorem, we construct a binary classification example as below.

In a binary classification problem, for each sample with true label $y \in \{+, -\}$ and prediction $\widehat{y} \in \{+, -\}$, the primary loss is defined as the expected $0/1$ loss for incorrect prediction, i.e., $\mathbb{E}_{y,\widehat{y}}\left[\mathbb{1}_{\{\widehat{y}\neq y\}}\right]$ and the secondary loss is defined as the expected $0/1$ loss for false negatives, i.e., $\mathbb{E}_{y,\widehat{y}}\left[\mathbb{1}_{\{\widehat{y}\neq y, y=+\}}\right]$. We denote by $h(b)$ the expert predicting $-$ with probability b and $+$ otherwise. Then every expert can be represented by a sequence of values of b. At round t, the true label is negative with probability a. We divide T into two phases evenly, $\{1, \ldots, T/2\}$ and $T/2 + 1, \ldots, T$, in each of which the adversary generates outcomes with different values of a and two experts $\mathcal{H} = \{h_1, h_2\}$ have different values of b in different phases. We construct two worlds with different values of a and b in phase 2 and any algorithm should have the same behavior in phase 1 of both worlds. The adversary randomly chooses one world with equal probability. The specific values of a and b are given in Table 1. Let $c = 1/16$.

Table 1: The values of a and b in different phases for the binary classification example.

experts\phase	$1: a = \frac{5}{8}$	$2: a = \frac{3}{4}$ (world I)	$2: a = \frac{5}{8}$ (world II)
h_1	$b = \frac{1}{6}$	$b = 0$	$b = \frac{1}{6}$
h_2	$b = 0$	$b = \frac{1}{2}$	$b = 0$

The loss of expert $h(b)$ is $\ell_{t,h(b)}^{(1)} = (1-a)b + a(1-b)$ and $\ell_{t,h(b)}^{(2)} = (1-a)b$. In phase 1 and phase 2 of world II, $\ell_{t,h_1}^{(1)} = 7/12$, $\ell_{t,h_1}^{(2)} = 1/16$, $\ell_{t,h_2}^{(1)} = 5/8$ and $\ell_{t,h_2}^{(2)} = 0$. In phase 2 of world I, $\ell_{t,h_1}^{(1)} = 3/4$, $\ell_{t,h_1}^{(2)} = 0$, $\ell_{t,h_2}^{(1)} = 1/2$ and $\ell_{t,h_2}^{(2)} = 1/8$. For any $h \in \mathcal{H}$, we have $L_{T,h}^{(2)} \leq T/16$.

For any algorithm which selects h_1 for T_1 (in expectation) rounds in phase 1 and T_2 (in expectation) rounds in phase 2 of world I. If $T_1 \leq T/4$, then $\mathrm{Reg}^{(1)} \geq (T/2 - T_1)/24 \geq T/96$ in world II; else if $T_1 > T/4$ and $T_2 \geq T_1/4$, then $\mathrm{Reg}^{(1)} \geq T_2/4 - T_1/24 \geq T/192$ in world I; else $\mathrm{Reg}_c^{(2)} = T_1/16 + (T/2 - T_2)/8 - T/16 = (T_1 - 2T_2)/16 \geq T/128$ in world I. In any case, we have $\mathbb{E}[\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)})] = \Omega(T)$. □

The proof of Theorem 1 implies that an expert with total secondary loss no greater than cT but high secondary loss at the beginning will consume a lot of budget for secondary loss, which makes switching to other experts with low primary loss later costly in terms of secondary loss. The theorem answers our first question negatively, i.e., we are unable to achieve no-regret for primary loss while performing as well as the worst expert with respect to the secondary loss.

4 Results in the "good" scenario

In this section, we consider the problem of minimizing $\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)})$ with Assumption 2 or 2′. We first provide lower bounds of $\Omega(T^\alpha)$ under Assumption 2 and of $\Omega(T^{\frac{1+\alpha}{2}})$ under Assumption 2′. Then we show that applying any switching-limited algorithms such as Shrinking Dartboard (SD) [Geulen et al., 2010] and Follow the Lazy Leader (FLL) [Kalai and Vempala, 2005] can achieve $\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)}) = O(T^{\frac{1+\alpha}{2}})$ under Assumption 2 or 2′, which matches the lower bound under Assumption 2′.

4.1 Lower bound

Theorem 2. *If Assumption 2 holds with some given c, δ, α, then there exists an adversary such that any algorithm incurs $\mathbb{E}[\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)})] = \Omega(T^\alpha)$.*

Proof. We construct a binary classification example to prove the lower bound.

The losses and the experts $\mathcal{H} = \{h_1, h_2\}$ are defined based on $h(b)$ in the same way as that in the proof of Theorem 1. We divide T into 3 phases, the first two of which have T^α rounds and the third has $T - 2T^\alpha$ rounds. Each expert has different bs in different phases as shown in Table 2. At each time t, the sample is negative with probability $3/4$. We set $c = 0$.

Since $(\ell_{t,h(0)}^{(1)}, \ell_{t,h(0)}^{(2)}) = (3/4, 0)$ and $(\ell_{t,h(1)}^{(1)}, \ell_{t,h(1)}^{(2)}) = (1/4, 1/4)$, the cumulative loss for both experts are $(L_{T,h}^{(1)}, L_{T,h}^{(2)}) = (3T/4 - T^\alpha/2, T^\alpha/4)$. Any algorithm $\mathcal{A}$ achieving $L_{T,h}^{(1)} \le 3T/4 - T^\alpha/4$ will incur $\mathrm{Reg}_c^{(2)} \ge T^\alpha/8$. □

Table 2: The values of b in different phases for the binary classification example.

experts\phase	$1 : T^\alpha$	$2 : T^\alpha$	$3 : T - 2T^\alpha$
h_1	$b = 1$	$b = 0$	$b = 0$
h_2	$b = 0$	$b = 1$	$b = 0$

Combined with the classical lower bound of $\Omega(\sqrt{T})$ in online learning [Cesa-Bianchi and Lugosi, 2006], $\mathbb{E}[\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)})] = \Omega(\max(T^\alpha, \sqrt{T}))$. In the relaxed version of the "good" scenario, we have the following theorem.

Theorem 3. *If Assumption 2′ holds with some given c, δ, α, then there exists an adversary such that any algorithm incurs $\mathbb{E}[\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)})] = \Omega(T^{\frac{1+\alpha}{2}})$.*

Sketch of the proof Inspired by the proof of the lower bound by Altschuler and Talwar [2018], we construct an adversary such that any algorithm achieving $\mathrm{Reg}^{(1)} = O(T^{\frac{1+\alpha}{2}})$ has to switch for some number of times. For the secondary loss, the adversary sets $\ell_{t,h}^{(2)} = c$ only if h has been selected for more than T^α rounds consecutively until time $t-1$; otherwise $\ell_{t,h}^{(2)} = c + \delta$. In this case, every switching will increase the secondary loss. Then we can show that either $\mathrm{Reg}^{(1)}$ or $\mathrm{Reg}_c^{(2)}$ is $\Omega(T^{\frac{1+\alpha}{2}})$. The complete proof can be found in Appendix A.

4.2 Algorithm

Under Assumption 2 or 2′, we are likely to suffer an extra δT^α secondary loss every time we switch from one expert to another. Inspired by this, we can upper bound $\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)})$ by limiting the number of switching times. Given a switching-limited learner $\mathcal{L}$ on scalar-valued losses, e.g., Shrinking Dartboard (SD) [Geulen et al., 2010] and Follow the Lazy Leader (FLL) [Kalai and Vempala, 2005], our algorithm $\mathcal{A}_{\mathrm{SL}}(\mathcal{L})$ is described as below.

We divide the time horizon into $T^{1-\alpha}$ epochs evenly and within each epoch we select the same expert. Let $e_i = \{(i-1)T^\alpha + 1, \dots, iT^\alpha\}$ denote the i-th epoch and $\ell^{(1)}_{e_i,h} = \sum_{t\in e_i} \ell^{(1)}_{t,h}/T^\alpha$ denote the average primary loss of the i-th epoch. We apply $\mathcal{L}$ over $\{\ell^{(1)}_{e_i,h}\}_{h\in\mathcal{H}}$ for $i = 1, \dots, T^{1-\alpha}$. Let $s_{\mathrm{SL}}(E)$ and $r_{\mathrm{SL}}(E)$ denote the expected number of switching times and the regret of running $\mathcal{L}$ for E rounds. Then we have the following theorem.

Theorem 4. *Under Assumption 2 or 2′, given a switching-limited learner $\mathcal{L}$, $\mathcal{A}_{\mathrm{SL}}(\mathcal{L})$ achieves $\mathrm{Reg}^{(1)} \le T^\alpha r_{\mathrm{SL}}(T^{1-\alpha})$ and $\mathrm{Reg}^{(2)}_c \le \delta T^\alpha(s_{\mathrm{SL}}(T^{1-\alpha}) + 1)$. By adopting SD or FLL as the learner $\mathcal{L}$, $\mathcal{A}_{\mathrm{SL}}(\mathrm{SD})$ and $\mathcal{A}_{\mathrm{SL}}(\mathrm{FLL})$ achieve $\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}^{(2)}_c) = O(\sqrt{\log(K)T^{1+\alpha}})$.*

Proof. It is obvious that $\mathrm{Reg}^{(1)} \le T^\alpha r_{\mathrm{SL}}(T^{1-\alpha})$. We denote by S the random variable of the total number of switching times and $\tau_1, \dots, \tau_S$ the time steps the algorithm switches. For notation simplicity, let $\tau_0 = 1$ and $\tau_{S+1} = T + 1$. Then $\mathrm{Reg}^{(2)}_c = \mathbb{E}_{\mathcal{A}}[\sum_{t=1}^T(\ell^{(2)}_{t,\mathcal{A}_t} - c)] \le \mathbb{E}_{\mathcal{A}}[\sum_{s=0}^S \sum_{t=\tau_s}^{\tau_{s+1}-1}(\ell^{(2)}_{t,\mathcal{A}_t} - c)] \le \mathbb{E}_{\mathcal{A}}[\sum_{s=0}^S \delta T^\alpha] = \delta T^\alpha(s_{\mathrm{SL}}(T^{1-\alpha}) + 1)$. Both SD and FLL have $s_{\mathrm{SL}}(T^{1-\alpha}) = O(\sqrt{\log(K)T^{1-\alpha}})$ and $r_{\mathrm{SL}}(T^{1-\alpha}) = O(\sqrt{\log(K)T^{1-\alpha}})$ [Geulen et al., 2010, Kalai and Vempala, 2005], which completes the proof. □

$\mathcal{A}_{\mathrm{SL}}(\mathrm{SD})$ and $\mathcal{A}_{\mathrm{SL}}(\mathrm{FLL})$ match the lower bound at $\Theta(T^{\frac{1+\alpha}{2}})$ under Assumption 2′. But there is a gap between the upper bound $O(T^{\frac{1+\alpha}{2}})$ and the lower bound $\Omega(T^\alpha)$ under Assumption 2, which is left as an open question. We investigate this question a little bit by answering negatively if the analysis of $\mathcal{A}_{\mathrm{SL}}(\mathcal{L})$ can be improved to achieve $O(T^\alpha)$. We define a class of algorithms which depends only on the cumulative losses of the experts, i.e., there exists a function $g : \mathbb{R}^{2K} \mapsto \Delta^K$ such that $p_t = g(L^{(1)}_{t-1}, L^{(2)}_{t-1})$. Many classical algorithms such as Exponential Weights [Littlestone et al., 1989] and Follow the Perturbed Leader [Kalai and Vempala, 2005] are examples in this class. The following theorem show that any algorithm dependent only on the cumulative losses cannot achieve a better bound than $\Omega(T^{\frac{1+\alpha}{2}})$, which provides some intuition on designing algorithms for future work. The detailed proof can be found in Appendix B.

Theorem 5. *Under Assumption 2, for any algorithm only dependent on the cumulative losses of the experts, $\mathbb{E}[\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}^{(2)}_c)] = \Omega(T^{\frac{1+\alpha}{2}})$.*

5 Results in the "bad" scenario

In the "bad" scenario, some experts may have secondary losses with high variance. To compete with the best expert in the period in which it has low variance, we assume that the learner is given some fixed external oracle determining which experts to deactivate and reactivate. In this section, we consider the goal of minimizing $\mathrm{SleepReg}^{(1)}(h^*)$ for all $h^* \in \mathcal{H}$ and $\mathrm{Reg}^{(2)}_c$. Here we study two oracles: one deactivates the "unsatisfactory" expert if detecting high variance of the secondary loss and never reactivates it again; the other one deactivates the "unsatisfactory" expert if detecting high variance of the secondary loss and reactivates it at fixed time steps.

5.1 The first oracle: deactivating the "unsatisfactory" experts

The oracle is described as below. The active expert set is initialized to contain all experts $\mathcal{H}_1 = \mathcal{H}$. At time $t = 1, \dots, T$, we let $\Delta\mathcal{H}_t = \{h \in \mathcal{H}_t : \exists t' \le t, \sum_{\tau=t'}^t(\ell^{(2)}_{\tau,h} - c) > \delta T^\alpha\}$ denote the set of active experts which do not satisfy Assumption 2. Then we remove these experts from the active expert set, i.e., $\mathcal{H}_{t+1} = \mathcal{H}_t \setminus \Delta\mathcal{H}_t$. We assume that there always exist some active experts, i.e. $\mathcal{H}_T \neq \emptyset$.

One direct way is running $\mathcal{A}_{\mathrm{SL}}(\mathcal{L})$ as a subroutine and restarting $\mathcal{A}_{\mathrm{SL}}(\mathcal{L})$ at time t if there exist experts deactivated at the end of $t-1$, i.e., $\Delta H_{t-1} \neq \emptyset$. However, restarting will lead to linear dependency on K for sleeping regrets. To avoid this linear dependency, we construct pseudo primary losses for each expert such that if h is active at time t, $\widetilde{\ell}^{(1)}_{t,h} = \ell^{(1)}_{t,h}$; otherwise, $\widetilde{\ell}^{(1)}_{t,h} = 1$. The probability of selecting inactive experts degenerates due to the high pseudo losses. For those inactive experts we cannot select, we construct a mapping $f : \mathcal{H} \mapsto \mathcal{H}$, which maps each expert to an active expert. If $\mathcal{A}_{\mathrm{SL}}(\mathcal{L})$ decides to select an inactive expert h at time t, we will select $f(h)$ instead. The detailed algorithm is described in Algorithm 1. Although the algorithm takes α as an input, it is worth to mention that the algorithm only uses α to decide the length of each epoch. We can choose a different epoch length and derive different regret upper bounds.

Algorithm 1 $\mathcal{A}_1$

Input: T, $\mathcal{H}$, α and a learner $\mathcal{L}$
Initialize $f(h) = h$ for all $h \in \mathcal{H}$.
Start an instance $\mathcal{A}_{\mathrm{SL}}(\mathcal{L})$.
for $t = 1, \dots, T$ **do**
  Get expert h_t from $\mathcal{A}_{\mathrm{SL}}(\mathcal{L})$.
  Select expert $f(h_t)$.
  Feed $\widetilde{\ell}_t^{(1)}$ to $\mathcal{A}_{\mathrm{SL}}(\mathcal{L})$.
  For all h with $f(h) \in \Delta\mathcal{H}_t$, set $f(h) = h_0$, where h_0 is any expert in $\mathcal{H}_{t+1}$.
end for

Theorem 6. *Let T_{h^*} denote the number of rounds where expert h^* is active. Running Algorithm 1 with learner $\mathcal{L}$ being SD or FLL can achieve*

$$\mathrm{SleepReg}^{(1)}(h^*) = O(\sqrt{\log(K) T_{h^*} T^\alpha}) \,, \tag{1}$$

for all $h^* \in \mathcal{H}$ and

$$\mathrm{Reg}_c^{(2)} = O(\sqrt{\log(K) T^{1+\alpha}} + KT^\alpha) \,. \tag{2}$$

Proof. Since $\ell_{m,h}^{(1)} \le \widetilde{\ell}_{m,h}^{(1)}$, we have

$$\begin{aligned}\mathrm{SleepReg}^{(1)}(h^*) = \left(\sum_{t=1}^{T_{h^*}} \mathbb{E}_{\mathcal{A}} \left[\ell_{t,\mathcal{A}_t}^{(1)}\right] - \sum_{t=1}^{T_{h^*}} \ell_{t,h^*}^{(1)}\right) \le \left(\sum_{t=1}^{T_{h^*}} \mathbb{E}_{\mathcal{A}} \left[\widetilde{\ell}_{t,\mathcal{A}_t}^{(1)}\right] - \sum_{t=1}^{T_{h^*}} \widetilde{\ell}_{t,h^*}^{(1)}\right) \\ = O(\sqrt{\log(K) T_{h^*} T^\alpha}) \,,\end{aligned}$$

where the last step uses the results in Theorem 4. It is quite direct to have $\mathrm{Reg}_c^{(2)} = O(\delta T^\alpha(\sqrt{\log(K) T^{1-\alpha}} + K)) = O(\sqrt{\log(K) T^{1+\alpha}} + KT^\alpha)$, where the first term comes from the number of switching times for running $\mathcal{A}_{\mathrm{SL}}$ and the second term comes from an extra switching caused by deactivating one expert. □

For the sleeping regret for expert h^*, the right hand side in Eq. (1) is $o(T_{h^*})$ if $T_{h^*} = \omega(T^\alpha)$, which is consistent with the impossibility result without bounded variance in Section 3. When $\alpha \ge 1/2$, the right hand side of Eq. (2) is dominated by KT^α. This linear dependency on K is inevitable if we want to have $\mathrm{SleepReg}_{h^*}^{(1)} = o(T_{h^*})$ for all $h^* \in \mathcal{H}$. The proof is given in Appendix C.

Theorem 7. *Let $T_{h^*} = \omega(T^\alpha)$ for all $h^* \in \mathcal{H}$. There exists an adversary such that any algorithm achieving $\mathrm{SleepReg}_{h^*}^{(1)} = o(T_{h^*})$ for all $h^* \in \mathcal{H}$ will incur $\mathrm{Reg}_c^{(2)} = \Omega(KT^\alpha)$ for $K = O(\log(T))$.*

5.2 The second oracle: reactivating at fixed times

Now we consider the oracle which deactivates the unsatisfactory experts once detecting and reactivate them at fixed times. The oracle is described as follows. At given $N+1$ fixed time steps $t_0 = 1, t_1, \dots, t_N$ with $t_{n+1} - t_n = \Omega(T^\beta)$ for some $\beta > \alpha$ (where $t_{N+1} = T+1$ for notation simplicity), the active expert set $\mathcal{H}_t$ is reset to $\mathcal{H}$. At time $t = t_n, \dots, t_{n+1} - 2$ for any $n = 0, \dots, N$, the experts $\Delta\mathcal{H}_t = \{h \in \mathcal{H}_t : \exists t' \text{ such that } t_n \le t' \le t, \sum_{\tau=t'}^{t} (\ell_{\tau,h}^{(2)} - c) > \delta T^\alpha\}$ will be deactivated, i.e. $\mathcal{H}_{t+1} = \mathcal{H}_t \setminus \Delta\mathcal{H}_t$. We assume that there always exists some satisfactory experts, i.e. $\mathcal{H}_{t_n - 1} \neq \emptyset$ for all $n = 1, \dots, N+1$.

Restarting Algorithm 1 at $t = t_0, \dots, t_N$ is one of the most direct methods. Let $T_{h^*}^{(n)}$ denote the number of rounds h^* is active during $t = t_n, \dots, t_{n+1} - 1$ and $T_{h^*} = \sum_{n=0}^{N} T_{h^*}^{(n)}$ denote the total number of rounds h^* is active. Then we have $\mathrm{SleepReg}_{h^*}^{(1)} = O(\sum_{n=0}^{N} \sqrt{\log(K) T_{h^*}^{(n)} T^\alpha}) = O(\sqrt{\log(K) T_{h^*} T^\alpha N})$ and $\mathrm{Reg}_c^{(2)} = O(\sum_{n=0}^{N} (\sqrt{\log(K) T^\alpha (t_{n+1} - t_n)} + K\delta T^\alpha)) = O(\sqrt{\log(K) T^{1+\alpha} N} + NKT^\alpha)$.

However, if all experts are active all times, then the upper bound of $\mathrm{SleepReg}^{(1)}(h^*)$ for the algorithm of restarting is $O(\sqrt{\log(K) T^{1+\alpha} N}) = O(\sqrt{\log(K) T^{2+\alpha-\beta}})$, which is quite large. We consider

a smarter algorithm with better sleeping regrets when T_{h^*} is large. The algorithm combines the methods of constructing meta experts for time-selection functions by Blum and Mansour [2007] to bound the sleeping regrets and inside each interval, we select experts based on SD [Geulen et al., 2010] to bound the number of switching times. We run the algorithm in epochs with length T^α and within each epoch we play the same expert. For simplicity, we assume that the active expert set will be updated only at the beginning of each epoch, which can be easily generalized. Let $e_i = \{(i-1)T^\alpha + 1, \dots, iT^\alpha\}$ denote the i-th epoch and $E = \{e_i\}_{i \in [T^{1-\alpha}]}$ denote the set of epochs. We let $\ell_{e,h}^{(1)} = \sum_{t \in e} \ell_{t,h}^{(1)} / T^\alpha$ and $\ell_{e,\mathcal{A}}^{(1)} = \sum_{t \in e} \ell_{t,\mathcal{A}_t}^{(1)} / T^\alpha$ denote the average primary loss of expert h and the algorithm. And we let $\mathcal{H}_e$ and $\Delta\mathcal{H}_e$ denote the active expert set at the beginning of epoch e and the deactivated expert set at the end of epoch e. Then we define the time selection function for epoch e as $I_{h^*}(e) = \mathbb{1}(h^* \text{ is active in epoch } e)$ for each $h^* \in \mathcal{H}$. Then we construct K meta experts for each time selection function. Similar to Algorithm 1, we adopt the same expert mapping function f and using pseudo losses $\widetilde{\ell}_{e,h}^{(1)} = \ell_{e,h}^{(1)}$ if h is active and $\widetilde{\ell}_{e,h}^{(1)} = 1$ if not. The detailed algorithm is shown as Algorithm 2. Then we have the following theorem, the detailed proof of which is provided in Appendix D.

Theorem 8. *Running Algorithm 2 can achieve*

$$\mathrm{SleepReg}^{(1)}(h^*) = O(\sqrt{\log(K) T^{1+\alpha}} + T_{h^*} \sqrt{\log(K) T^{\alpha-1}}) \,,$$

for all $h^* \in \mathcal{H}$ and

$$\mathrm{Reg}_c^{(2)} = O(\sqrt{\log(K) T^{1+\alpha}} + \log(K) T^\alpha N + NKT^\alpha) \,.$$

Algorithm 2 achieves $o(T_{h^*})$ sleeping regrets for h^* with $T_{h^*} = \omega(T^{\frac{1+\alpha}{2}})$ and outperforms restarting Algorithm 1 when $NT_{h^*} = \omega(T)$. $\mathrm{SleepReg}^{(1)}(h^*)$ of Algorithm 2 is $O(\sqrt{\log(K) T_{h^*}^{1+\alpha}})$ when $T_{h^*} = \Theta(T)$, which matches the results in Theorem 4.

Algorithm 2 $\mathcal{A}_2$

Input: T, $\mathcal{H}$, α and η
Initialize $f(h) = h$ for all $h \in \mathcal{H}$.
$w_{1,h}^{h^*} = \frac{1}{K}$ for all $h \in \mathcal{H}$, for all $h^* \in \mathcal{H}$.
for $m = 1, \dots, T^{1-\alpha}$ **do**
  $w_{m,h} = \sum_{h^*} I_{h^*}(e_m) w_{m,h}^{h^*}$, $W_m = \sum_h w_{m,h}$ and $p_{m,h} = \frac{w_{m,h}}{W_m}$.
  if $m \in \{(t_n - 1)/T^{1-\alpha} + 1\}_{n=0}^N$ **then** get h_m from p_m. **else**
  With prob. $\frac{w_{m,h_{m-1}}}{w_{m-1,h_{m-1}}}$, get $h_m = h_{m-1}$; with prob. $1 - \frac{w_{m,h_{m-1}}}{w_{m-1,h_{m-1}}}$, get h_m from p_m.
  end if
  Select expert $f(h_m)$.
  Update $w_{m+1,h}^{h^*} = w_{m,h}^{h^*} \eta^{I_{h^*}(e_m)(\widetilde{\ell}_{e_m,h}^{(1)} - \eta \widetilde{\ell}_{e_m,\mathcal{A}}^{(1)}) + 1}$ for all $h, h^* \in \mathcal{H}$.
  For all h with $f(h) \in \Delta\mathcal{H}_{e_m}$, set $f(h) = h_0$, where h_0 is any expert in $\mathcal{H}_{e_{m+1}}$.
end for

6 Discussion

We introduce the study of online learning with primary and secondary losses. We find that achieving no-regret with respect to the primary loss while performing no worse than the worst expert with respect to the secondary loss is impossible in general. We propose a bounded variance assumption over experts such that we can control secondary losses by limiting the number of switching times. Therefore, we are able to bound the regret with respect to the primary loss and the regret to cT with respect to the secondary loss. Our work is only a first step in this problem and there are several open questions.

One is the optimality under Assumption 2. As aforementioned, our bounds of $\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)})$ in the "good" scenario are not tight and we show that any algorithm only dependent on the cumulative losses will have $\mathrm{Reg}^{(1)} = \Omega(T^{\frac{1+\alpha}{2}})$, which indicates that the optimal algorithm cannot only depends on the cumulative losses if the optimal bound is $o(T^{\frac{1+\alpha}{2}})$. Under Assumption 2′, the upper bound of the algorithm of limiting switching matches the lower bound. This possibly implies that limiting switching may not be the best way to make use of the information provided by Assumption 2.

In the "bad" scenario with access to the oracle which reactivates experts at fixed times, our sleeping regret bounds depend not only on T_{h^*} but also on T, which makes the bounds meaningless when T_{h^*} is small. It is unclear if we can obtain optimal sleeping regrets dependent only on T_{h^*} for all $h^* \in \mathcal{H}$. The algorithm of *Adanormalhedge* by Luo and Schapire [2015] can achieve sleeping regret of $O(\sqrt{T_{h^*}})$ without bound on the number of switching actions. However, how to achieve sleeping regret of $o(T_{h^*})$ with limited switching cost is of independent research interest.

In the "bad" scenario where Assumption 2 does not hold, we assume that c is pre-specified and known to the oracle. Theorem 1 show that achieving $\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)}) = o(T)$ with $c = \max_h L_{T,h}^{(2)}$ is impossible without any external oracle. How to define a setting an unknown c and design a reasonable oracle in this setting is an open question.

Broader Impact

This research studies a society-constrained online decision making problem, where we take the decision receiver's objective into consideration. Therefore, in a decision making process (e.g. deciding whether to hire a job applicant, whether to approve a loan, or whether to admit a student to an honors class), the decision receiver (e.g., job applicants, loan applicants, students) could benefit from our study at the cost of increasing the loss of the decision maker (e.g., recruiters, banks, universities) a little. The consequences of failure of the system and biases in the data are not applicable.

Acknowledgments and Disclosure of Funding

This work was supported in part by the National Science Foundation under grant CCF-1815011.

References

Jacob Abernethy, Peter L Bartlett, and Elad Hazan. Blackwell approachability and no-regret learning are equivalent. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 27–46, 2011.

Jason Altschuler and Kunal Talwar. Online learning over a finite action set with limited switching. In *Conference On Learning Theory*, pages 1569–1573, 2018.

Peter Auer, Chao-Kai Chiang, Ronald Ortner, and Madalina Drugan. Pareto front identification from stochastic bandit feedback. In *Artificial intelligence and statistics*, pages 939–947, 2016.

David Blackwell et al. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956.

Avrim Blum. Empirical support for winnow and weighted-majority algorithms: Results on a calendar scheduling domain. *Machine Learning*, 26(1):5–23, 1997.

Avrim Blum and Yishay Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8(Jun):1307–1324, 2007.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

Alexey Chernov and Vladimir Vovk. Prediction with expert evaluators' advice. *arXiv preprint arXiv:0902.4127*, 2009.

Eyal Even-Dar, Michael Kearns, Yishay Mansour, and Jennifer Wortman. Regret to the best vs. regret to the average. *Machine Learning*, 72(1-2):21–37, 2008.

Eyal Even-Dar, Robert Kleinberg, Shie Mannor, and Yishay Mansour. Online learning for global cost functions. In *COLT*, 2009.

Yoav Freund, Robert E Schapire, Yoram Singer, and Manfred K Warmuth. Using and combining predictors that specialize. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 334–343, 1997.

Sascha Geulen, Berthold Vöcking, and Melanie Winkler. Regret minimization for online buffering problems using the weighted majority algorithm. In *COLT*, pages 132–143, 2010.

C-L Hwang and Abu Syed Md Masud. *Multiple objective decision making—methods and applications: a state-of-the-art survey*, volume 164. Springer Science & Business Media, 2012.

Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

Varun Kanade and Thomas Steinke. Learning hurdles for sleeping experts. *ACM Transactions on Computation Theory (TOCT)*, 6(3):1–16, 2014.

Michael Kapralov and Rina Panigrahy. Prediction strategies without loss. In *Advances in Neural Information Processing Systems*, pages 828–836, 2011.

Robert Kleinberg, Alexandru Niculescu-Mizil, and Yogeshwer Sharma. Regret bounds for sleeping experts and bandits. *Machine learning*, 80(2-3):245–272, 2010.

Wouter M Koolen. The pareto regret frontier. In *Advances in Neural Information Processing Systems*, pages 863–871, 2013.

Nick Littlestone, Manfred K Warmuth, et al. *The weighted majority algorithm*. University of California, Santa Cruz, Computer Research Laboratory, 1989.

Haipeng Luo and Robert E Schapire. Achieving all with no parameters: Adanormalhedge. In *Conference on Learning Theory*, pages 1286–1304, 2015.

Shie Mannor, Vianney Perchet, and Gilles Stoltz. Approachability in unknown games: Online learning meets multi-objective optimization. In *Conference on Learning Theory*, pages 339–355, 2014.

Amir Sani, Gergely Neu, and Alessandro Lazaric. Exploiting easy data in online optimization. In *Advances in Neural Information Processing Systems*, pages 810–818, 2014.

Eralp Turgay, Doruk Oner, and Cem Tekin. Multi-objective contextual bandit problem with similarity information. In *International Conference on Artificial Intelligence and Statistics*, pages 1673–1681, 2018.

A Proof of Theorem 3

Theorem 3. *If Assumption 2′ holds with some given c, δ, α, then there exists an adversary such that any algorithm incurs $\mathbb{E}[\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)})] = \Omega(T^{\frac{1+\alpha}{2}})$.*

Proof. We construct an adversary with oblivious primary losses and adaptive secondary losses to prove the theorem. The adversary is inspired by the proof of the lower bound by Altschuler and Talwar [2018]. We divide T into $T^{1-\alpha}$ epochs evenly and the primary losses do not change within each epoch. Let $\lceil t \rceil_e = \min_{m: mT^\alpha \geq t} mT^\alpha$ denote the last time step of the epoch containing time step t. For each expert $h \in \mathcal{H}$, at the beginning of each epoch, we toss a fair coin and let $\ell_{t,h}^{(1)} = 0$ if it is head and $\ell_{t,h}^{(1)} = 1$ if it is tail. It is well-known that there exists a universal constant a such that $\mathbb{E}\left[\min_{h\in\mathcal{H}} Z_h\right] = E/2 - a\sqrt{E\log(K)}$ where $Z_h \sim \mathrm{Bin}(E, 1/2)$. Then we have

$$\mathbb{E}\left[\min_{h\in\mathcal{H}} \sum_{t=1}^{T} \ell_{t,h}^{(1)}\right] \leq \frac{T}{2} - aT^{\frac{1+\alpha}{2}}\sqrt{\log(K)}\,.$$

For algorithm $\mathcal{A}$, let $\mathcal{A}_t$ denote the selected expert at time t. Then we construct adaptive secondary losses as follows. First, for the first T^α rounds, $\ell_{t,h}^{(2)} = c + \delta$ for all $h \in \mathcal{H}$. For $t \geq T^\alpha + 1$,

$$\ell_{t,h}^{(2)} = \begin{cases} c & \text{if } h = \mathcal{A}_{t-1} = \ldots = \mathcal{A}_{t-T^\alpha} \\ c+\delta & \text{otherwise} \end{cases}.$$

This indicates that the algorithm can obtain $\ell_{t,\mathcal{A}_t}^{(2)} = c$ only by selecting the expert she has consecutively selected in the last T^α rounds and that each switching leads to $\ell_{t,\mathcal{A}_t}^{(2)} = c + \delta$. Let S denote the total number of switchings and $\tau_1, \ldots, \tau_S$ denote the time steps $\mathcal{A}$ switches. For notation simplicity, let $\tau_{S+1} = T+1$. If $\mathbb{E}\left[L_{T,\mathcal{A}}^{(1)}\right] \geq T/2 - aT^{\frac{1+\alpha}{2}}\sqrt{\log(K)}/2$, then $\mathbb{E}\left[\mathrm{Reg}^{(1)}\right] \geq aT^{\frac{1+\alpha}{2}}\sqrt{\log(K)}/2$; otherwise,

$$\frac{T}{2} - \frac{1}{2}\mathbb{E}\left[\sum_{s=1}^{S} \min\left(\tau_{s+1} - \tau_s, \lceil\tau_s\rceil_e + 1 - \tau_s\right)\right] \overset{\text{(a)}}{\leq} \mathbb{E}\left[L_{T,\mathcal{A}}^{(1)}\right] < \frac{T}{2} - aT^{\frac{1+\alpha}{2}}\sqrt{\log(K)}/2\,,$$

where Eq. (a) holds due to that the s-th switching helps to decrease the expected primary loss by at most $\min\left(\tau_{s+1} - \tau_s, \lceil\tau_s\rceil_e + 1 - \tau_s\right)/2$. Since the s-th switching increases the secondary loss to $c + \delta$ for at least $\min(\tau_{s+1} - 1 - \tau_s, T^\alpha)$ rounds, then we have

$$\begin{aligned}
\mathbb{E}\left[L_{T,\mathcal{A}}^{(2)}\right] \geq& cT + \delta\mathbb{E}\left[\sum_{s=1}^{S} \min(\tau_{s+1} - \tau_s, T^\alpha)\right] \\
\geq& cT + \delta\mathbb{E}\left[\sum_{s=1}^{S} \min\left(\tau_{s+1} - \tau_s, \lceil\tau_s\rceil_e + 1 - \tau_s\right)\right] \\
>& cT + \delta a T^{\frac{1+\alpha}{2}}\sqrt{\log(K)},
\end{aligned}$$

which indicates that $\mathbb{E}\left[\mathrm{Reg}_c^{(2)}\right] = \Omega(T^{\frac{1+\alpha}{2}})$. Therefore, $\mathbb{E}\left[\max\left(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)}\right)\right] \geq \max\left(\mathbb{E}\left[\mathrm{Reg}^{(1)}\right], \mathbb{E}\left[\mathrm{Reg}_c^{(2)}\right]\right) = \Omega(T^{\frac{1+\alpha}{2}})$. □

B Proof of Theorem 5

Theorem 5. *Under Assumption 2, for any algorithm only dependent on the cumulative losses of the experts, $\mathbb{E}[\max(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)})] = \Omega(T^{\frac{1+\alpha}{2}})$.*

Proof. We divide T into $T^{1-\beta}$ intervals evenly with $\beta = \frac{1+\alpha}{2}$ and construct $T^{1-\beta} + 1$ worlds with 2 experts. For computation simplicity, we let $\delta = 1/2$. The adversary selects a random world W at

the beginning. She selects world 0 with probability $1/2$ and world w with probability $1/2T^{1-\beta}$ for all $w \in [T^{1-\beta}]$.

In world 0, we design the losses of experts as shown in Table 3. During the w-th interval with $w \in [T^{1-\beta}]$ being odd, we set $(\ell_{t,h_1}^{(1)}, \ell_{t,h_1}^{(2)}, \ell_{t,h_2}^{(1)}, \ell_{t,h_2}^{(2)}) = (0, c+\delta T^{\alpha-\beta}, 1, c-\delta T^{\alpha-\beta})$ for the first $T^\beta/2$ rounds and $(\ell_{t,h_1}^{(1)}, \ell_{t,h_1}^{(2)}, \ell_{t,h_2}^{(1)}, \ell_{t,h_2}^{(1)}) = (1, c, 0, c)$ for the second $T^\beta/2$ rounds. For w being even, we swap the losses of the two experts, i.e., $(\ell_{t,h_1}^{(1)}, \ell_{t,h_2}^{(2)}, \ell_{t,h_2}^{(1)}, \ell_{t,h_2}^{(2)}) = (1, c-\delta T^{\alpha-\beta}, 0, c+\delta T^{\alpha-\beta})$ for the first $T^\beta/2$ rounds and $(\ell_{t,h_1}^{(2)}, \ell_{t,h_1}^{(2)}, \ell_{t,h_2}^{(1)}, \ell_{t,h_2}^{(2)}) = (0, c, 1, c)$ for the second $T^\beta/2$ rounds.

The intuition of constructing world $w \in [T^{1-\beta}]$ is described as below. In world w, the secondary loss is the same as that in world 0. The primary losses of each expert $h \in \mathcal{H}$ in the first $w-1$ intervals are an approximately random permutation of that in world 0. Therefore, any algorithm will attain almost the same expected primary loss (around $(w-1)T^\beta/2$) in the first $w-1$ intervals of world w. The primary losses during the first $T^\beta/2$ rounds in the w-th interval are the same as those in world 0. Therefore, the cumulative losses from the beginning to any time t in the first half of the w-th interval are almost the same in world 0 and world w, which makes the algorithm only dependent on the cumulative losses behave nearly the same during the first half of the w-th interval in two worlds. For $t = (w-1/2)T^\beta + 1, \dots, T$, we set $\ell_{t,h}^{(1)} = 1$ for all $h \in \mathcal{H}$, which indicates that any algorithms are unable to improve their primary loss after $t = (w-1/2)T^\beta + 1$. To prove the theorem, we show that if the algorithm selects expert h with loss $(1, c-\delta T^{\alpha-\beta})$ during the first half of the w-th interval with large fraction, then $\mathrm{Reg}^{(1)}$ will be large in world w; otherwise, $\mathrm{Reg}_c^{(2)}$ will be large in world 0.

More specifically, for the first $w-1$ intervals in world w, we need to make the cumulative primary losses to be $(w-1)T^\beta/2$ with high probability. Let $t' = (w-1)T^\beta - 2\sqrt{(w-1)T^\beta \log(T)}$. For $t = 1, \dots, t'$, $\ell_{t,h}^{(1)}$ are i.i.d. samples from $\mathrm{Ber}(1/2)$ for all $h \in \mathcal{H}$. We denote by $E_h^{(w)}$ the event of $\left|\sum_{t=1}^{t'} (\ell_{t,h}^{(1)} - 1/2)\right| \le \sqrt{(w-1)T^\beta \log(T)}$ and denote by E the event of $\cap_{h\in\mathcal{H}, w\in[T^{1-\beta}]} E_h^{(w)}$. If $E_{h_1}^{(w)} \cap E_{h_2}^{(w)}$ holds, we compensate the cumulative primary losses by assigning $\ell_{t,h}^{(1)} = 1$ for $(w-1)T^\beta/2 - \sum_{t=1}^{t'} \ell_{t,h}^{(1)}$ rounds and $\ell_{t,h}^{(1)} = 0$ for the remaining rounds during $t = t'+1, \dots, (w-1)T^\beta$ for all $h \in \mathcal{H}$ such that the cumulative primary losses in the first $w-1$ intervals for both experts are $(w-1)T^\beta/2$; otherwise, we set $\ell_{t,h}^{(1)} = 1$ for all $h \in \mathcal{H}$ during $t = t'+1, \dots, (w-1)T^\beta$. Hence, if $E_{h_1}^{(w)} \cap E_{h_2}^{(w)}$, the cumulative losses $L_{(w-1)T^\beta,h}^{(1)} = (w-1)T^\beta/2$ for all $h \in \mathcal{H}$. To make it clearer, the values of the secondary losses in world w for an even w if $E_{h_1}^{(w)} \cap E_{h_2}^{(w)}$ holds are illustrated in Table 4.

Let $q_w = 2\sum_{t=(w-1)T^\beta+1}^{(w-1/2)T^\beta} \mathbb{E}\left[\mathbb{1}(\ell_{t,\mathcal{A}_t}^{(1)} = 0)\right]/T^\beta$ denote the expected fraction of selecting the expert with losses $(0, c+\delta T^{\alpha-\beta})$ in w-th interval in world 0 as well as that in world w when E holds. We denote by $\mathrm{Reg}^{(1,w)} = L_{T,\mathcal{A}}^{(1,w)} - L_{T,h_0}^{(1,w)}$ and $\mathrm{Reg}_{t'}^{(1,w)} = L_{t',\mathcal{A}}^{(1,w)} - L_{t',h_0}^{(1,w)}$ with $h_0 = \arg\min_{h\in\mathcal{H}} L_{T,h}^{(1,w)}$ being the best expert in hindsight the regret with respect to the primary loss for all times and the regret incurred during $t = 1, \dots, t'$ in world w. We denote by $\mathrm{Reg}_c^{(2,w)}$ the regret to cT with respect to the secondary loss in world w. Then we have

$$\mathbb{E}\left[\mathrm{Reg}_c^{(2,W)} \,|W=0\right] = \sum_{w\in[T^{1-\beta}]} \frac{\delta(2q_w-1)T^\alpha}{2},$$

and for all $w \in [T^{1-\beta}]$,

$$\begin{aligned}\mathbb{E}\left[\mathrm{Reg}^{(1,W)} \,|W=w, E\right] \ge& (1-q_w)\frac{T^\beta}{2} + \mathbb{E}\left[\mathrm{Reg}_{t'}^{(1,w)} \,|W=w, E\right] - \left((w-1)T^\beta - t'\right)\\ \ge& (1-q_w)\frac{T^\beta}{2} - 2\sqrt{(w-1)T^\beta \log(T)}\,.\end{aligned}$$

Due to Hoeffding's inequality and union bound, we have $\mathbb{P}\left[\neg E_h^{(w)}\right] \le \frac{2}{T^2}$ for all $h \in \mathcal{H}$ and $w \in [T^{1-\beta}]$ and $\mathbb{P}[\neg E] \le \frac{4}{T^{1+\beta}}$. Let $Q = \frac{\sum_{w=1}^{T^{1-\beta}} q_w}{T^{1-\beta}}$ denote the average of q_w over all $w \in [T^{1-\beta}]$. By taking expectation over the adversary, we have

$$
\begin{aligned}
&\mathbb{E}\left[\max\left(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)}\right)\right] \\
\geq &\mathbb{P}\left[E\right] \cdot \mathbb{E}\left[\max\left(\mathrm{Reg}^{(1)}, \mathrm{Reg}_c^{(2)}\right) | E\right] \\
\geq &\left(1 - \frac{4}{T^{1+\beta}}\right)\left(\frac{1}{2T^{1-\beta}} \sum_{w=1}^{T^{1-\beta}} \mathbb{E}\left[\mathrm{Reg}^{(1,W)} | W = w, E\right] + \frac{1}{2}\mathbb{E}\left[\mathrm{Reg}_c^{(2,W)} | W = 0, E\right]\right) \\
\geq &\frac{1}{2}\left(\frac{1}{2T^{1-\beta}}\left(\sum_w (1-q_w)\frac{T^\beta}{2} - 2\sum_{w=1}^{T^{1-\beta}} \sqrt{(w-1)T^\beta \log(T)}\right) + \frac{\delta}{4}\sum_{w=1}^{T^{1-\beta}} (2q_w - 1)T^\alpha\right) \\
\geq &\frac{1}{8}(1-Q)T^\beta - \sqrt{T\log(T)} + \frac{\delta}{8}(2Q-1)T^{1-\beta+\alpha} \\
\geq &\frac{1}{16}T^{\frac{1+\alpha}{2}} - \sqrt{T\log(T)}, \qquad (3)
\end{aligned}
$$

where Eq. (3) holds by setting $\beta = \frac{1+\alpha}{2}$ and $\delta = 1/2$. □

Table 3: The losses in world 0.

experts\time		$T^\beta/2$	$T^\beta/2$	$T^\beta/2$	$T^\beta/2$	$T^\beta/2$	...
h_1	$\ell^{(1)}$	0	1	1	0	0	...
	$\ell^{(2)}$	$c + \delta T^{\alpha-\beta}$	c	$c - \delta T^{\alpha-\beta}$	c	$c + \delta T^{\alpha-\beta}$	...
h_2	$\ell^{(1)}$	1	0	0	1	1	...
	$\ell^{(2)}$	$c - \delta T^{\alpha-\beta}$	c	$c + \delta T^{\alpha-\beta}$	c	$c - \delta T^{\alpha-\beta}$	...

Table 4: The primary losses in world w (which is even) if $E_{h_1}^{(w)} \cap E_{h_2}^{(w)}$ holds.

experts\time		t'	$(w-1)T^\beta - t'$	$T^\beta/2$	$T^\beta/2$	$T - T^\beta$
h_1	$\ell^{(1)}$	i.i.d. from $\mathrm{Ber}(1/2)$	compensate	1	1	1
h_2	$\ell^{(1)}$	i.i.d. from $\mathrm{Ber}(1/2)$	compensate	0	1	1

C Proof of Theorem 7

Theorem 7. *Let $T_{h^*} = \omega(T^\alpha)$ for all $h^* \in \mathcal{H}$. There exists an adversary such that any algorithm achieving $\mathrm{SleepReg}_{h^*}^{(1)} = o(T_{h^*})$ for all $h^* \in \mathcal{H}$ will incur $\mathrm{Reg}_c^{(2)} = \Omega(KT^\alpha)$ for $K = O(\log(T))$.*

Proof. The idea is to construct an example in which the best expert with respect to the primary loss is deactivated sequentially while incurring an extra $\Theta(T^\alpha)$ secondary loss. In the example, we set $\mathcal{H} = [K]$. Let $T_k = T^{\alpha + \frac{(k-1)(1-\alpha)}{K-1}}$ for $k \in [K]$ and $T_0 = 0$. For each expert $k \in \mathcal{H}$, we set $(\ell_{t,k}^{(1)}, \ell_{t,k}^{(2)}) = (1, c)$ for $t \leq T_{k-1}$ and $(\ell_{t,k}^{(1)}, \ell_{t,k}^{(2)}) = (0, c + \frac{\delta T^\alpha}{T_k - T_{k-1}})$ for $t \geq T_{k-1} + 1$. Then expert k will be deactivate at time $t = T_k$. For any algorithm with $\mathrm{SleepReg}_k^{(1)} = o(T_k)$ for all $k \in \mathcal{H}$, expert k should be selected for $T_k - 2T_{k-1} - o(T_k)$ rounds during $t = T_{k-1} + 1, \ldots, T_k$. Therefore, we have $\mathrm{Reg}_c^{(2)} \geq \sum_{k\in[K]} \frac{\delta T^\alpha}{T_k - T_{k-1}}(T_k - 2T_{k-1} - o(T_k)) = \Omega(KT^\alpha)$. □

D Proof of Theorem 8

Theorem 8. *Running Algorithm 2 can achieve*

$$
\mathrm{SleepReg}^{(1)}(h^*) = O(\sqrt{\log(K)T^{1+\alpha}} + T_{h^*}\sqrt{\log(K)T^{\alpha-1}}),
$$

for all $h^* \in \mathcal{H}$ and

$$\mathrm{Reg}_c^{(2)} = O(\sqrt{\log(K)T^{1+\alpha}} + \log(K)T^\alpha N + NKT^\alpha)\,.$$

Proof. Let $\widetilde{L}_h^{(1,h^*)} = \sum_{m=1}^{T^{1-\alpha}} I_{h^*}(e_m)\widetilde{\ell}_{e_m,h}^{(1)}$ and $\widetilde{L}_{\mathcal{A}}^{(1,h^*)} = \sum_{m=1}^{T^{1-\alpha}} I_{h^*}(e_m)\widetilde{\ell}_{e_m,\mathcal{A}}^{(1)}$ denote the cumulative pseudo primary losses of expert h and algorithm $\mathcal{A}$ during the time when h^* is active. First, since we update $w_{m+1,h}^{h^*} = w_{m,h}^{h^*}\eta^{I_{h^*}(e_m)(\widetilde{\ell}_{e_m,h}^{(1)} - \eta\widetilde{\ell}_{e_m,\mathcal{A}}^{(1)})+1} \le w_{m,h}^{h^*}$ with $\eta \in [1/\sqrt{2}, 1]$ and experts will not be reactivated between (not including) t_n and t_{n+1}, the probability of following the first rule on Line 7 in Algorithm 2, which is $\frac{w_{m+1,h_m}}{w_{m,h_m}}$, is legal. Then we show that at each epoch m, the probability of getting $h_m = h$ is $\mathbb{P}[h_m = h] = p_{m,h}$. The proof follows Lemma 1 by [Geulen et al., 2010]. For an reactivating epoch $m \in \{(t_n-1)/T^\alpha + 1\}_{n=0}^N$, h_m is drawn from p_m and thus, $\mathbb{P}[h_m = h] = p_{m,h}$ holds. For other epochs $m \notin \{(t_n-1)/T^\alpha + 1\}_{n=0}^N$, we prove it by induction. Assume that $\mathbb{P}[h_{m-1} = h] = p_{m-1,h}$, then

$$\begin{aligned}
\mathbb{P}[h_m = h] &= \mathbb{P}[h_{m-1} = h]\frac{w_{m,h}}{w_{m-1,h}} + p_{m,h}\sum_{h'\in\mathcal{H}}\mathbb{P}[h_{m-1} = h']\left(1 - \frac{w_{m,h'}}{w_{m-1,h'}}\right) \\
&= \frac{w_{m-1,h}}{W_{m-1}}\cdot\frac{w_{m,h}}{w_{m-1,h}} + \frac{w_{m,h}}{W_m}\left(1 - \sum_{h'\in\mathcal{H}}\frac{w_{m-1,h'}}{W_{m-1}}\cdot\frac{w_{m,h'}}{w_{m-1,h'}}\right) \\
&= p_{m,h}\,.
\end{aligned}$$

To prove the upper bound on sleeping regrets, we follow Claim 12 by Blum and Mansour [2007] to show that $\sum_{h,h^*} w_{m,h}^{h^*} \le K\eta^{m-1}$ for all $m \in [T^{1-\alpha}]$.

First, we have

$$W_m\widetilde{\ell}_{e_m,\mathcal{A}}^{(1)} = W_m\sum_{h\in\mathcal{H}} p_{m,h}\widetilde{\ell}_{e_m,h}^{(1)} = \sum_{h\in\mathcal{H}} w_{m,h}\widetilde{\ell}_{e_m,h}^{(1)} = \sum_{h\in\mathcal{H}}\sum_{h^*\in\mathcal{H}} I_{h^*}(e_m)w_{m,h}^{h^*}\widetilde{\ell}_{e_m,h}^{(1)}\,. \tag{4}$$

Then according to the definition of $w_{m,h}^{h^*}$, we have

$$\begin{aligned}
&\sum_{h\in\mathcal{H},h^*\in\mathcal{H}} w_{m+1,h}^{h^*} \\
=&\sum_{h\in\mathcal{H},h^*\in\mathcal{H}} w_{m,h}^{h^*}\eta^{I_{h^*}(e_m)(\widetilde{\ell}_{e_m,h}^{(1)} - \eta\widetilde{\ell}_{e_m,\mathcal{A}}^{(1)})+1} \\
\le&\eta\left(\sum_{h\in\mathcal{H},h^*\in\mathcal{H}} w_{m,h}^{h^*}\left(1 - (1-\eta)I_{h^*}(e_m)\widetilde{\ell}_{e_m,h}^{(1)}\right)\left(1 + (1-\eta)I_{h^*}(e_m)\widetilde{\ell}_{e_m,\mathcal{A}}^{(1)}\right)\right) \\
\le&\eta\left(\sum_{h\in\mathcal{H},h^*\in\mathcal{H}} w_{m,h}^{h^*} - (1-\eta)\left(\sum_{h\in\mathcal{H},h^*\in\mathcal{H}} w_{m,h}^{h^*}I_{h^*}(e_m)\widetilde{\ell}_{e_m,h}^{(1)} - W_m\widetilde{\ell}_{e_m,\mathcal{A}}^{(1)}\right)\right) \\
=&\eta\sum_{h\in\mathcal{H},h^*\in\mathcal{H}} w_{m,h}^{h^*}\,,
\end{aligned}$$

where the last inequality adopts Eq. (4). Combined with $w_{1,h}^{h^*} = \frac{1}{K}$ for all $h \in \mathcal{H}, h^* \in \mathcal{H}$, we have $\sum_{h,h^*} w_{m+1,h}^{h^*} \le K\eta^m$. Since $w_{m+1,h}^{h^*} = w_{1,h}^{h^*}\eta^{\sum_{i=1}^m I_{h^*}(e_i)\widetilde{\ell}_{e_i,h}^{(1)} - \eta\sum_{i=1}^m I_{h^*}(e_i)\widetilde{\ell}_{e_i,\mathcal{A}}^{(1)} + m} \le K\eta^m$, we have

$$\widetilde{L}_{\mathcal{A}}^{(1,h^*)} - \widetilde{L}_h^{(1,h^*)} \le \frac{(1-\eta)\widetilde{L}_h^{(1,h^*)} + \frac{2\log(K)}{\log(1/\eta)}}{\eta}\,.$$

By setting $\eta = 1 - \sqrt{2\log(K)/T^{1-\alpha}}$, we have $\mathrm{SleepReg}^{(1)}(h^*) \le 2\sqrt{\log(K)T^{1+\alpha}} + 2T_{h^*}\sqrt{\log(K)T^{\alpha-1}}$.

To derive $\mathrm{Reg}_c^{(2)}$, we bound the number of switching times. We denote by S_n the number of epochs in which some experts are deactivated during $(t_n-1)/T^\alpha + 1 < m < (t_{n+1}-1)/T^\alpha + 1$ and by

$\tau_1, \ldots, \tau_{S_n}$ the deactivating epochs, i.e., $\Delta\mathcal{H}_{\tau_i} \neq \emptyset$ for $i \in [S_n]$. We denote by α_m the probability of following the second rule at line 7 in Algorithm 2, which is getting h_m from p_m. Then we have

$$\alpha_m = \sum_{h\in\mathcal{H}} \mathbb{P}\left[h_{m-1} = h\right]\left(1 - \frac{w_{m,h}}{w_{m-1,h}}\right) = \sum_{h\in\mathcal{H}} \frac{w_{m-1,h}}{W_{m-1}}\left(1 - \frac{w_{m,h}}{w_{m-1,h}}\right) = \frac{W_{m-1} - W_m}{W_{m-1}}\,.$$

Since $W_{\tau_{i+1}}/W_{\tau_i+1} \geq \eta^{2(\tau_{i+1}-\tau_i-1)}$, we have

$$\begin{aligned}\sum_{m=\tau_i+1}^{\tau_{i+1}} \alpha_m &\leq 1 - \sum_{m=\tau_i+2}^{\tau_{i+1}} \log(1-\alpha_m) = 1 - \sum_{m=\tau_i+2}^{\tau_{i+1}} \log\left(\frac{W_m}{W_{m-1}}\right) = 1 + \log\left(\frac{W_{\tau_{i+1}}}{W_{\tau_i+1}}\right)\\ &\leq 1 + 2\sqrt{2}(\tau_{i+1} - \tau_i - 1)(1-\eta) = 1 + 4(\tau_{i+1} - \tau_i - 1)\sqrt{\log(K)/T^{1-\alpha}}\,.\end{aligned}$$

Therefore, during time $(t_n - 1)/T^\alpha \leq m < (t_{n+1} - 1)/T^\alpha$, the algorithm will switch at most $K + 4(t_{n+1} - t_n)\sqrt{\log(K)/T^{1-\alpha}} + 1$ times in expectation, which results in $\mathrm{Reg}_c^{(2)} \leq 4\delta\sqrt{\log(K)T^{1+\alpha}} + \delta N(K+1)T^\alpha = O(\sqrt{\log(K)T^{1+\alpha}} + NKT^\alpha)$ □