

# Online learning for changing environments using coin betting\*

Kwang-Sung Jun

*Wisconsin Institute for Discovery, University of Wisconsin-Madison*  
e-mail: [kjun@discovery.wisc.edu](mailto:kjun@discovery.wisc.edu)

Francesco Orabona

*Department of Computer Science, Stony Brook University*  
e-mail: [francesco@orabona.com](mailto:francesco@orabona.com)

Stephen Wright

*Department of Computer Sciences, University of Wisconsin-Madison*  
e-mail: [swright@cs.wisc.edu](mailto:swright@cs.wisc.edu)

and

Rebecca Willett

*Department of Electrical and Computer Engineering, University of Wisconsin-Madison*  
e-mail: [willett@discovery.wisc.edu](mailto:willett@discovery.wisc.edu)

**Abstract:** A key challenge in online learning is that classical algorithms can be slow to adapt to changing environments. Recent studies have proposed “meta” algorithms that convert any online learning algorithm to one that is adaptive to changing environments, where the adaptivity is analyzed in a quantity called the strongly-adaptive regret. This paper describes a new meta algorithm that has a strongly-adaptive regret bound that is a factor of  $\sqrt{\log(T)}$  better than other algorithms with the same time complexity, where  $T$  is the time horizon. We also extend our algorithm to achieve a first-order (i.e., dependent on the observed losses) strongly-adaptive regret bound for the first time, to our knowledge. At its heart is a new parameter-free algorithm for the learning with expert advice (LEA) problem in which experts sometimes do not output advice for consecutive time steps (i.e., *sleeping* experts). This algorithm is derived by a reduction from optimal algorithms for the so-called coin betting problem. Empirical results show that our algorithm outperforms state-of-the-art methods in both learning with expert advice and metric learning scenarios.

**MSC 2010 subject classifications:** Primary 68T05.

**Keywords and phrases:** Online learning, changing environments, learning with expert advice, online convex optimization.

Received June 2017.

---

\*This paper is an extended version of Jun et al. [11]

## Contents

1	Introduction . . . . .	5283
2	Meta algorithms for changing environments . . . . .	5286
3	Coin betting meets sleeping experts . . . . .	5289
4	Coping with a changing environment by sleeping CB . . . . .	5296
5	Experiments . . . . .	5302
	5.1 Learning with expert advice (LEA) . . . . .	5303
	5.2 Metric learning . . . . .	5303
6	Future work . . . . .	5305
	Appendix . . . . .	5305
	A.1 The coin betting potential . . . . .	5305
	A.2 SA-Regret is stronger than $m$ -shift regret . . . . .	5306
	A.3 The data streaming intervals can replace the geometric covering intervals . . . . .	5307
	A.4 A subtle difference between the geometric covering and data streaming intervals . . . . .	5308
	A.5 Technical results . . . . .	5308
	Acknowledgements . . . . .	5309
	References . . . . .	5309

## 1. Introduction

Online learning algorithms are typically tailored to stationary environments, but in many applications the environment is dynamic. In online portfolio management, for example, stock price trends can vary unexpectedly, and the ability to track changing trends and adapt to them are crucial in maximizing profit. In product reviews, words describing product quality may change over time as products evolve and the tastes of customers change. Keeping track of the changes in the metric describing the relationship between review text and rating is crucial for improving analysis and the quality of recommendations.

We consider the problem of adapting to changing environments in the online learning context. Let  $\mathcal{D}$  be the decision space,  $\mathcal{L}$  be a family of loss functions that map  $\mathcal{D}$  to  $\mathbb{R}$ , and  $T$  be the target time horizon. Let  $\mathcal{A}$  be an online learning algorithm. We define the online learning protocol in Figure 1.

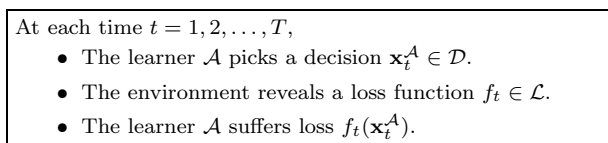


FIG 1. *Online learning protocol*

The usual goal of online learning is to find a strategy that compares favorably with the best fixed comparator in a subset  $\mathcal{W}$  of decision space  $\mathcal{D}$ , in hindsight.

Algorithm	SA-Regret order	Time factor
FLH [9]	$\sqrt{T \log T}$	$T$
AFLH [9]	$\sqrt{T \log T \log(I_2 - I_1)}$	$\log T$
SAOL [5]	$\sqrt{(I_2 - I_1) \log^2(I_2)}$	$\log T$
CBCE (ours)	$\sqrt{(I_2 - I_1) \log(I_2)}$	$\log T$

TABLE 1

SA-Regret bounds of meta algorithms on  $I = [I_1..I_2] \subseteq [T]$ . We show the part of the regret due to the meta algorithm only, not the black-box. The last column is the multiplicative factor in the time complexity introduced by the meta algorithm. CBCE (our algorithm) achieves the best SA-Regret and time complexity.

(Often,  $\mathcal{W} = \mathcal{D}$ .) Classically, one seeks a low value of the following (cumulative) static regret objective:

$$\text{Regret}_T^A := \sum_{t=1}^T f_t(\mathbf{x}_t^A) - \min_{\mathbf{w} \in \mathcal{W}} \sum_{t=1}^T f_t(\mathbf{w}).$$

When the environment is changing, static regret is not a suitable measure, since it compares the learning strategy against a decision that is fixed for all  $t$ . We need to make use of stronger notions of regret that allow comparators to change over time. We introduce the notation  $[T] := \{1, \dots, T\}$  and  $[A..B] := \{A, A+1, \dots, B\}$ . Daniely et al. [5] defined *strongly-adaptive regret* (SA-Regret), which measures the regret over *any* time interval  $I = [I_1..I_2] \subseteq [T]$ :

$$\text{SA-Regret}_T^A(I) := \left( \sum_{t \in I} f_t(\mathbf{x}_t^A) - \min_{\mathbf{w} \in \mathcal{W}} \sum_{t \in I} f_t(\mathbf{w}) \right). \quad (1)$$

Throughout,  $I_1$  ( $I_2$ ) denotes the starting (ending) time step of an interval  $I$ . We call an algorithm *strongly-adaptive* if it has a *low* value of SA-Regret, by which we mean a value  $O(\text{polylog}(T)R_{\mathcal{P}}(I))$ , where  $R_{\mathcal{P}}(I)$  is the minimax static regret of the online learning problem  $\mathcal{P}$  restricted to interval  $I$ .

Let us call  $\mathbf{w}_{1:T} := \{\mathbf{w}_1, \dots, \mathbf{w}_T\}$  an *m-shift sequence* if it changes at most  $m$  times, that is,  $\sum_{j=1}^{T-1} \mathbf{1}\{\mathbf{w}_j \neq \mathbf{w}_{j+1}\} \leq m$ . A related notion, *m-shift regret* [10], measures the regret with respect to a comparator that changes at most  $m$  times in  $T$  time steps.

$$m\text{-Shift-Regret}_T^A := \sum_{t=1}^T f_t(\mathbf{x}_t^A) - \min_{\mathbf{w}_{1:T} \in \mathcal{W}^T : m\text{-shift seq.}} \sum_{t=1}^T f_t(\mathbf{w}_t).$$

While the *m-shift regret* is more interpretable, SA-Regret is a stronger notion since it is well-known that a tight SA-Regret bound implies a tight *m-shift regret* bound [14, 5], as we discuss further in Section A.2. As noted by [21], SA-Regret has a strong connection to so-called dynamic regret (with respect to the temporal variations of the  $f_t$ 's).

Several generic online algorithms that adapt to changing environments have been proposed recently. Rather than being designed for a specific learning problem, these are “meta” algorithms that take *any* online learning algorithm as a black-box and turn it into an adaptive one. We summarize the SA-Regret of ex-

Algorithm	$m$ -shift regret	Running time	Agnostic to $m$
Fixed Share [10, 4]	$\sqrt{mT(\log N + \log T)}$	$NT$	✗
	$\sqrt{m^2T(\log N + \log T)}$	$NT$	✓
GeneralTracking(EXP) [8]  ( $\gamma \in (0, 1)$ )	$\sqrt{mT(\log N + m \log^2 T)}$	$NT \log T$	✓
	$\sqrt{mT(\log N + \log^2 T)}$	$NT \log T$	✗
	$\sqrt{\frac{1}{\gamma}mT(\log N + m \log T)}$	$NT^{1+\gamma} \log T$	✓
	$\sqrt{\frac{1}{\gamma}mT(\log N + \log T)}$	$NT^{1+\gamma} \log T$	✗
ATV [14]	$\sqrt{mT(\log N + \log T)}$	$NT^2$	✓
SAOL(MW) [5]	$\sqrt{mT(\log N + \log^2 T)}$	$NT \log T$	✓
CBCE(CB) (ours)	$\sqrt{mT(\log N + \log T)}$	$NT \log T$	✓

TABLE 2

$m$ -shift regret bounds of LEA algorithms. Our proposed algorithm (last line) achieves the best regret among those with the same time complexity and does not need to know  $m$ . Each quantity omits constant factors. “Agnostic to  $m$ ” means that an algorithm does not need to know the number  $m$  of switches in the best expert.

isting meta algorithms in Table 1. In particular, the pioneering work of Hazan & Seshadhri [9] introduced *adaptive regret*, a slightly weaker notion than the SA-Regret, and proposed two meta algorithms called Follow-the-Leading-History (FLH) and Advanced FLH (AFLH).<sup>1</sup> However, their SA-Regret depends on  $T$  rather than  $|I|$  and hence can be significantly larger. The SAOL approach of [5] improves the SA-Regret to  $O\left(\sqrt{(I_2 - I_1) \log^2(I_2)}\right)$ .

In this paper, we propose a new meta algorithm called *Coin Betting for Changing Environments* (CBCE) that combines the idea of “sleeping experts” introduced in [2, 6] with the Coin Betting (CB) algorithm [16]. The SA-Regret of CBCE is better by a factor  $\sqrt{\log(I_2)}$  than that of SAOL, as shown in Table 1. We present our extension of CB to sleeping experts and prove its regret bound in Section 3. This result leads to the improved SA-Regret bound of CBCE in Section 4.

Our improved bound yields a number of improvements in various online learning problems. In describing these improvements, we designate by  $\mathcal{M}\langle\mathcal{B}\rangle$  a complete algorithm assembled from meta algorithm  $\mathcal{M}$  and black-box  $\mathcal{B}$ . In this notation, our algorithm is designated by CBCE(CB).

Consider the learning with expert advice (LEA) problem with  $N$  experts. We make comparisons with respect to  $m$ -shift regret bounds, as many LEA algorithms provide only bounds of this type. Our algorithm CBCE(CB) has  $m$ -shift regret

$$O\left(\sqrt{mT(\log N + \log T)}\right)$$

and time complexity  $O(NT \log T)$ . This regret is a factor  $\sqrt{\log T}$  better than existing algorithms with the same time complexity. Although AdaNormalHedge.TV (ATV) and Fixed Share achieve the same regret, the former has larger time com-

<sup>1</sup>Strongly adaptive regret is similar to the notion of adaptive regret introduced by [9], but emphasizes the dependency on the interval length  $|I|$ .

plexity, and the latter requires prior knowledge of the number of shifts  $m$ . We summarize the  $m$ -shift regret bounds of various algorithms in Table 2. We emphasize that the same regret order and time complexity as CBCE⟨CB⟩ can be achieved by combining our proposed CBCE meta algorithm with any blackbox algorithm (e.g., AdaNormalHedge [14]).

In online convex optimization with  $G$ -Lipschitz loss functions over a convex set  $D \in \mathbb{R}^d$  of diameter  $B$ , Online Gradient Descent (OGD) has regret  $O(BG\sqrt{T})$  [19]. Thus, CBCE with OGD (CBCE⟨OGD⟩) has the following SA-Regret:

$$O\left((BG + \sqrt{\log(I_2)})\sqrt{|I|}\right),$$

which improves by a factor  $\sqrt{\log(I_2)}$  over SAOL⟨OGD⟩.

We also propose an improved version of CBCE that has a so-called *first-order* regret bound. That is, the SA-Regret on an interval  $I = \{I_1, \dots, I_2\}$  scales with  $\min_{\mathbf{w} \in \mathcal{W}} \sum_{t \in I} f_t(\mathbf{w})$  rather than  $|I|$  as follows:

$$O\left(\log(I_2) \sqrt{\min_{\mathbf{w} \in \mathcal{W}} \sum_{t \in I} f_t(\mathbf{w}) + \text{polylog}(I_2)}\right),$$

where we omit the term due to the blackbox algorithm. We emphasize that, while there is an extra  $\sqrt{\log(I_2)}$  factor and additive term, the main quantity  $\min_{\mathbf{w} \in \mathcal{W}} \sum_{t \in [T]} f_t(\mathbf{w})$  can be significantly smaller than  $|I|$  if there exists the decision  $\mathbf{w}$  whose loss is very small in  $I$ . To our knowledge, this is the first first-order SA-Regret bound in online learning.<sup>2</sup>

In Section 5, we compare CBCE empirically to a number of meta algorithms for changing environments in two online learning problems: LEA and Mahalanobis metric learning. We observe that CBCE outperforms the state-of-the-art methods in both tasks, thus confirming our theoretical findings.

## 2. Meta algorithms for changing environments

Let  $\mathcal{B}$  be a black-box online learning algorithm following the protocol in Figure 1. A trick commonly used in designing a meta algorithm  $\mathcal{M}$  for changing environments is to initiate a new instance of  $\mathcal{B}$  at every time step [9, 8, 1]. That is, we run  $\mathcal{B}$  independently for each interval  $J$  in  $\{[t, \infty] \mid t = 1, 2, \dots\}$ . Denoting by  $\mathcal{B}_J$  the run of black-box  $\mathcal{B}$  on interval  $J$ , a meta algorithm at time  $t$  takes a weighted average of decisions from the runs  $\{\mathcal{B}_J : t \in J\}$ . The underlying idea is as follows. Suppose we are at time  $t$  and the environment has changed at an earlier time  $t' < t$ . We hope that the meta algorithm would assign a large weight to the black-box run  $\mathcal{B}_{J'}$  (where  $J' = [t', \infty]$ ), since other runs are either based on data prior to  $t'$  or use only a subset of the data generated since  $t'$ . Ideally, the meta algorithm would assign a large weight to  $\mathcal{B}_{J'}$  soon after time  $t'$ , by carefully examining the online performance of each black-box run.

<sup>2</sup>First-order bounds are available for specific online learning problems. For LEA, for example, AdaNormalHedge.TV [14] has a first-order regret bound.

This schema requires updating of  $t$  instances of the black-box algorithm at each time step  $t$ , leading to a  $O(t)$  multiplicative increase in complexity over a single run. This factor can be reduced to  $O(\log t)$  by restarting black-box algorithms on a carefully designed set of intervals, such as the geometric covering intervals [5] (GC) or the data streaming intervals [9, 8] (DS), which is a special case of a more general set of intervals considered in [20]. While both GC and DS achieve the same goal, as we show in Section A.3,<sup>3</sup> we use the former as our starting point for ease of exposition.

**Geometric covering intervals** We define the  $\mathcal{J}_k$  to be the collection of intervals of length  $2^k$ :

$$\mathcal{J}_k := \{[(i \cdot 2^k) .. ((i + 1) \cdot 2^k - 1)] : i \in \mathbb{N}\}, \quad \forall k \in \{0, 1, \dots\}.$$

The geometric covering intervals [5] are

$$\mathcal{J} := \bigcup_{k \in \{0, 1, \dots\}} \mathcal{J}_k .$$

That is,  $\mathcal{J}$  is the set of intervals of doubling length, with intervals of size  $2^k$  exactly partitioning the set  $\mathbb{N} \setminus \{1, \dots, 2^k - 1\}$ ; see Figure 2.

Define the set of intervals that includes time  $t$  as follows:

$$\text{Active}(t) := \{J \in \mathcal{J} : t \in J\} .$$

It can be shown that  $|\text{Active}(t)| = \lfloor \log_2(t) \rfloor + 1$ . Since at most  $O(\log(t))$  intervals contain any given time point  $t$ , the time complexity of the meta algorithm is a factor  $O(\log(t))$  larger than that of the black-box  $\mathcal{B}$ .

The following Lemma from Daniely et al. [5] shows that an arbitrary interval  $I$  can be partitioned into a sequence of smaller blocks whose lengths successively double, then successively halve. This result is key to the usefulness of the geometric covering intervals.

**Lemma 1.** [5, Lemma 5] *Any interval  $I \subseteq \mathbb{N}$  can be partitioned into two finite sequences of disjoint and consecutive intervals, denoted  $\{J^{(-a)}, J^{(-a+1)}, \dots, J^{(0)}\}$  and  $\{J^{(1)}, J^{(2)}, \dots, J^{(b)}\}$  where for all  $i \in [(-a)..b]$ , we have  $J^{(i)} \in \mathcal{J}$  and  $J^{(i)} \subset I$ , such that*

$$\begin{aligned} |J^{(-i)}|/|J^{(-i+1)}| &\leq 1/2, & i = 1, 2, \dots, a; \\ |J^{(i+1)}|/|J^{(i)}| &\leq 1/2, & i = 1, 2, \dots, b - 1 . \end{aligned}$$

**Regret decomposition** We show now how to use the geometric covering intervals to decompose the SA-Regret of a complete algorithm  $\mathcal{M}(\mathcal{B})$ . Using the notation

$$R_I^A(\mathbf{w}) := \sum_{t \in I} f_t(\mathbf{x}_t^A) - \sum_{t \in I} f_t(\mathbf{w}) ,$$

---

<sup>3</sup>Except for a subtle case, which we discuss in Section A.4.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	...		
$\mathcal{J}_0$	[	]	[	]	[	]	[	]	[	]	[	]	[	]	[	]	[	]	[	]	...
$\mathcal{J}_1$		[		]		[		]		[		]		[		]		[		]	...
$\mathcal{J}_2$				[				]					[							]	...
$\mathcal{J}_3$								[												]	...

FIG 2. Geometric covering intervals. Each interval is denoted by  $[ \ ]$ .

we can restate (1) as follows:

$$\text{SA-Regret}_T^A(I) = \max_{\mathbf{w} \in \mathcal{W}} R_I^A(\mathbf{w}).$$

Suppose we denote by  $\mathbf{x}_t^{\mathcal{B}_J}$  the decision from black-box run  $\mathcal{B}_J$  at time  $t$  and by  $\mathbf{x}_t^{\mathcal{M}(\mathcal{B})}$  the combined decision of the meta algorithm at time  $t$ . Since  $\mathcal{M}(\mathcal{B})$  is a combination of a meta  $\mathcal{M}$  and a black-box  $\mathcal{B}$ , its regret depends on both  $\mathcal{M}$  and  $\mathcal{B}$ . Perhaps surprisingly, we can decompose the two sources of regret additively through the geometric covering  $\mathcal{J}$ , as we now describe. For any  $I \subseteq [T]$ , let  $\bigcup_{i=-a}^b J^{(i)}$  be the partition of  $I$  obtained from Lemma 1. The regret of  $\mathcal{M}(\mathcal{B})$  on  $I$  can be decomposed as follows:

$$\begin{aligned} & R_I^{\mathcal{M}(\mathcal{B})}(\mathbf{w}) \\ &= \sum_{t \in I} \left( f_t(\mathbf{x}_t^{\mathcal{M}(\mathcal{B})}) - f_t(\mathbf{w}) \right) \\ &= \sum_{i=-a}^b \left( \sum_{t \in J^{(i)}} f_t(\mathbf{x}_t^{\mathcal{M}(\mathcal{B})}) - f_t(\mathbf{x}_t^{\mathcal{B}_{J^{(i)}}}) + f_t(\mathbf{x}_t^{\mathcal{B}_{J^{(i)}}}) - f_t(\mathbf{w}) \right) \\ &= \underbrace{\sum_{i=-a}^b \sum_{t \in J^{(i)}} \left( f_t(\mathbf{x}_t^{\mathcal{M}(\mathcal{B})}) - f_t(\mathbf{x}_t^{\mathcal{B}_{J^{(i)}}}) \right)}_{=:(\text{meta regret on } J^{(i)})} + \underbrace{\sum_{i=-a}^b \sum_{t \in J^{(i)}} \left( f_t(\mathbf{x}_t^{\mathcal{B}_{J^{(i)}}}) - f_t(\mathbf{w}) \right)}_{=:(\text{black-box regret on } J^{(i)})}. \quad (2) \\ & \qquad \qquad \qquad \underbrace{\hspace{10em}}_{=:(\text{meta regret on } I)} \end{aligned}$$

(We purposely use symbol  $J$  for intervals in  $\mathcal{J}$  and  $I$  for a generic interval that is not necessarily in  $\mathcal{J}$ .) The black-box regret on  $J = [J_1..J_2] \in \mathcal{J}$  is exactly the standard regret for  $T = |J|$ , since the black-box run  $\mathcal{B}_J$  was started from time  $J_1$ . Thus, in order to prove that a meta algorithm  $\mathcal{M}$  suffers low SA-Regret, it remains to show two things:

1.  $\mathcal{M}$  has low regret on interval  $J \in \mathcal{J}$ ;
2. The outer sums over  $i$  in (2) are small for both the meta algorithm and the black-box algorithm.

Daniely et al. [5] address the second issue above in their analysis. They show that if the black-box regret on  $J^{(i)}$  is  $c\sqrt{|J^{(i)}|}$  for some  $c$  then the second double summation of (2) is bounded by  $8c\sqrt{|I|}$ ,<sup>4</sup> which is perhaps the best one can hope

<sup>4</sup>The argument is essentially the “doubling trick” described in [3, Section 2.3].

for. The same holds true for the meta algorithm. Thus, it remains to focus on the first issue above. This is our main contribution.

In the next two sections, we describe the design and application of our meta algorithm. In Section 3, we propose a novel method that incorporates sleeping experts and the coin betting framework. Section 4 describes how our method can be used as a meta algorithm that has an SA-Regret guarantee.

### 3. Coin betting meets sleeping experts

Our meta algorithm CBCE extends the coin-betting framework [16] to a variant of the learning with expert advice (LEA) problem called “sleeping experts” [2, 6]. CBCE is parameter-free (there is no explicit learning rate) and has near-optimal regret. Our construction below has further interest as a near-optimal solution for the sleeping bandits problem.

**Sleeping experts** In the LEA framework, the decision set is  $\mathcal{D} = \Delta^N$ , an  $N$ -dimensional probability simplex of weights assigned to the various experts. To distinguish LEA from the general online learning problem, we use notation  $\mathbf{p}_t$  in place of  $\mathbf{x}_t$ , and  $h_t$  in place of  $f_t$ . Denoting by  $\boldsymbol{\ell}_t := (\ell_{t,1}, \dots, \ell_{t,N})^\top \in [0, 1]^N$  the vector of loss values of experts at time  $t$  provided by the environment, the learner’s loss function is  $h_t(\mathbf{p}) := \mathbf{p}^\top \boldsymbol{\ell}_t$ .

Since  $\mathbf{p} \in \mathcal{D}$  is a probability vector, the learner’s decision can be viewed as hedging between the  $N$  alternatives. Let  $\mathbf{e}_i$  be an indicator vector for dimension  $i$ ; e.g.,  $\mathbf{e}_2 = (0, 1, 0, \dots, 0)^\top$ . In this notation, the comparator set  $\mathcal{W} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ , that is, the learner competes with a strategy that commits to a single expert for the entire time interval  $[1..T]$ .<sup>5</sup>

Recall that each black-box run  $\mathcal{B}_J$  is on a different interval  $J$ . The meta algorithm’s role is to hedge bets over multiple black-box runs. Thus, it is natural to treat each run  $\mathcal{B}_J$  as an *expert* and use an LEA algorithm to combine decisions from each expert  $\mathcal{B}_J$ . The loss incurred on run  $\mathcal{B}_J$  is  $\ell_{t,\mathcal{B}_J} := f_t(\mathbf{x}_t^{\mathcal{B}_J})$ .

The challenge is that each expert  $\mathcal{B}_J$  may not output decisions at time steps outside the interval  $J$ . This problem can be reduced to the sleeping experts problem studied in [2, 6], in which experts are not required to provide decisions at every time step; see [14]. We introduce an indicator variable  $\mathcal{I}_{t,i} \in \{0, 1\}$ , which is set to 1 if expert  $i$  is awake (that is, outputting a decision) at time  $t$ , and zero otherwise. Define  $\boldsymbol{\mathcal{I}}_t := [\mathcal{I}_{t,1}, \mathcal{I}_{t,2}, \dots, \mathcal{I}_{t,N}]^\top$ , where  $N$  can be countably infinite. The algorithm is said to be “aware” of  $\boldsymbol{\mathcal{I}}_t$  and it assigns zero weight to the experts that are sleeping, that is,  $\mathcal{I}_{t,i} = 0 \implies p_{t,i} = 0$ . We would like to have a guarantee on the regret with respect to expert  $i$ , but only for the time

<sup>5</sup>The decision set may be nonconvex, or even discrete, for example,  $\mathcal{D} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$  [3, Section 3]. In this discrete case, no hedging is allowed; the learner must pick a single expert for the entire interval. To choose an element of this set, one could first choose an element  $\mathbf{p}_t$  from  $\Delta^N$ , then choose a decision  $\mathbf{e}_i \in \mathcal{D}$  with probability  $p_{t,i}$ . The regret guarantee for such a scheme is the same as for the standard LEA, but with *expected* rather than deterministic regret.



steps for which expert  $i$  is awake. Following Luo & Schapire [14], we define a regret bound with respect to  $\mathbf{u} \in \Delta^N$  as follows:

$$\text{Regret}_T(\mathbf{u}) := \sum_{t=1}^T \sum_{i=1}^N \mathcal{I}_{t,i} u_i (\langle \ell_t, \mathbf{p}_t \rangle - \ell_{t,i}). \quad (3)$$

If we set  $\mathbf{u} = \mathbf{e}_j$  for some  $j$ , the above is simply regret with respect to expert  $j$  while that expert is awake, and we aim to achieve a regret of  $O(\sqrt{\sum_t \mathcal{I}_{t,j}})$  up to logarithmic factors. If  $\mathcal{I}_{t,j} = 1$  for all  $t \in [T]$ , then it recovers the standard static regret in LEA.

**Coin betting for LEA** We consider the coin betting framework of Orabona & Pál [16], which constructs an LEA algorithm from a *coin betting potential function* (explained below). A player starts from the initial endowment 1. At each time step, the adversary chooses an outcome arbitrarily while the player decides on which side to bet (heads or tails) and how much to bet. Then the outcome is revealed. The outcome can be a head (+1), tail (-1), or any point on the continuum between these two extremes (e.g., -0.3) where the absolute value of the outcome indicates the weight of being a head or tail. We encode a coin flip at iteration  $t$  as  $g_t \in [-1, 1]$  where  $|g_t|$  indicates the weight of the outcome. Let  $\text{Wealth}_{t-1}$  be the total money the player possesses after time step  $t-1$ . (Note that  $\text{Wealth}_0 = 1$ .) We encode the player's betting decision as the signed betting fraction  $\beta_t \in (-1, 1)$ , where the positive (negative) sign indicates head (tail) and the absolute value  $|\beta_t| < 1$  indicates the fraction of his current wealth to bet. Thus, the actual amount wagered is  $w_t := \beta_t \text{Wealth}_{t-1}$ . Once the coin flip  $g_t$  is revealed, the player's wealth changes as follows:  $\text{Wealth}_t = (1 + g_t \beta_t) \text{Wealth}_{t-1}$ . The player makes (loses) money when the betted side is correct (wrong), and the amount of wealth change depends on both the flip weight  $|g_t|$  and his betting amount  $|w_t|$ .

In the coin betting framework, a potential function denoted by  $F_t(g_1, \dots, g_t)$  has an important role. Given this function, and denoting  $g_{1:t} := g_1, g_2, \dots, g_t$ , the betting fraction  $\beta_t$  and the amount wagered  $w_t$  are determined as follows:

$$\beta_t(g_{1:t-1}) := \frac{F_t(g_{1:t-1}, 1) - F_t(g_{1:t-1}, -1)}{F_t(g_{1:t-1}, 1) + F_t(g_{1:t-1}, -1)}, \quad (4a)$$

$$w_t = \beta_t(g_{1:t-1}) \cdot \left( 1 + \sum_{s=1}^{t-1} g_s w_s \right). \quad (4b)$$

(We use  $\beta_t$  in place of  $\beta_t(g_{1:t-1})$  when it is clear from the context.) A precise definition of  $F_t$  appears in Section A.1; it suffices for now to say that the sequence  $F_1, F_2, \dots$  must satisfy the following key condition by (4a):

$$\forall t, F_t(g_{1:t}) \leq 1 + \sum_{s=1}^t g_s w_s. \quad (5)$$

That is,  $F_t$  is a lower bound on the wealth of a player who bets by (4a). We emphasize that the term  $w_t$  is decided before  $g_t$  is revealed, yet the in-

---

**Algorithm 1** Sleeping CB
 

---

**Input:** Number of experts  $N$ , prior distribution  $\pi \in \Delta^N$

**for**  $t = 1, 2, \dots$  **do**

    For each  $i \in \text{Active}(t)$ , set

$$w_{t,i} \leftarrow \beta_{t,i}(z_{1:t-1,i}) \cdot \left(1 + \sum_{s=1}^{t-1} z_{s,i} w_{s,i}\right).$$

    For all  $i \in [N]$ , set  $\hat{p}_{t,i} \leftarrow \pi_i \mathcal{I}_{t,i}[w_{t,i}]_+$ .

    Predict with  $\mathbf{p}_t \leftarrow \begin{cases} \hat{\mathbf{p}}_t / \|\hat{\mathbf{p}}_t\|_1 & \text{if } \|\hat{\mathbf{p}}_t\|_1 > 0 \\ \pi_{\mathcal{I}_t} & \text{if } \|\hat{\mathbf{p}}_t\|_1 = 0. \end{cases}$

    Receive loss vector  $\ell_t \in [0, 1]^N$ .

    The learner suffers loss  $h_t(\mathbf{p}_t) = \langle \ell_t, \mathbf{p}_t \rangle_{\mathcal{I}_t}$ .

    For each  $i \in \text{Active}(t)$ , set

$$g_{t,i} \leftarrow \begin{cases} h_t(\mathbf{p}_t) - \ell_{t,i} & \text{if } w_{t,i} > 0 \\ [h_t(\mathbf{p}_t) - \ell_{t,i}]_+ & \text{if } w_{t,i} \leq 0. \end{cases}$$

**end for**

---

equality (5) holds for any  $g_t \in [-1, 1]$ . Property (5) is key to analyzing the wealth arising from the strategy (4a); see Section A.1. In the restricted setting in which  $g_s \in \{\pm 1\}$ , a betting strategy  $\beta_t$  based on a potential function proposed by Krichevsky & Trofimov [12] achieves the optimal wealth up to constant factors [3].

Orabona & Pál [16] have devised a reduction of LEA to the simple coin betting problem described above. The idea is to instantiate a coin betting problem for each expert  $i$  where the signed coin flip  $g_{t,i}$  is set as a conditionally truncated regret with respect to expert  $i$ , rather than being set by an adversary. We denote by  $\beta_{t,i}$  the betting fraction for expert  $i$  and by  $w_{t,i}$  the amount of betting for expert  $i$ ,  $\forall i \in [N]$ .

We apply this treatment to the sleeping experts setting and propose a new algorithm **Sleeping CB**. Modifications are required because some experts may not output a decision for some time steps. Defining  $z_{t,i} := \mathcal{I}_{t,i} g_{t,i}$ , we modify (4) as follows:

$$\beta_{t,i}(z_{1:t-1,i}) := \frac{F_{t,i}(z_{1:t-1,i}, 1) - F_{t,i}(z_{1:t-1,i}, -1)}{F_{t,i}(z_{1:t-1,i}, 1) + F_{t,i}(z_{1:t-1,i}, -1)}, \quad (6a)$$

$$w_{t,i} = \beta_t(z_{1:t-1,i}) \cdot \left(1 + \sum_{s=1}^{t-1} z_{s,i} w_{s,i}\right).$$

Condition (5) on the potential functions is modified accordingly to

$$\forall t, F_{t,i}(z_{1:t,i}) \leq 1 + \sum_{s=1}^t z_{s,i} w_{s,i}. \quad (7)$$

We denote by  $\pi_{\mathcal{I}_t}$  the prior  $\pi$  restricted to experts that are awake (for which  $\mathcal{I}_{t,i} = 1$ ). The Sleeping CB algorithm is specified in Algorithm 1. (Here and subsequently, we use notation  $[x]_+ := \max(x, 0)$ .)

The regret of Sleeping CB is bounded in Theorem 2. Unlike the standard CB, in which all the experts use  $F_t$  at time  $t$ , expert  $i$  in Sleeping CB uses  $F_{t,i}$ , which is different for each expert. For this reason, the proof of the CB regret in [16] does not transfer easily to the regret (3) of Sleeping CB. However, this result is crucial to an improved strongly-adaptive regret bound.

**Theorem 2.** (Regret of Sleeping CB) *Define  $S_{t,i} := \sum_{s=1}^t \mathcal{I}_{s,i}$  and for every  $i \in [N]$  let  $\{F_{t,i}\}_{t \geq 1}$  be a sequence of potential functions that satisfies (7). Suppose that*

$$\log(F_{T,i}(z_{1:T})) \geq H_{T,i}(z_{1:T,i}) := c_1 \frac{(\sum_{s=1}^T z_{s,i})^2}{S_{T,i}} + c_{2,i}, \quad \text{for all } i \in [N],$$

for some  $c_1 > 0$  and  $c_{2,i} \in \mathbb{R}$ . Then for the regret defined in (3), Algorithm 1 satisfies

$$\text{Regret}_T(\mathbf{u}) \leq \sqrt{\frac{1}{c_1} \cdot \left( \sum_{i=1}^N u_i S_{T,i} \right) \cdot \left( \text{KL}(\mathbf{u} \parallel \boldsymbol{\pi}) - \sum_{i=1}^N u_i c_{2,i} \right)}.$$

*Proof.* We show first that  $\sum_{i=1}^N \pi_i \mathcal{I}_{t,i} g_{t,i} w_{t,i} \leq 0$ . Define  $r_{t,i} := \langle \boldsymbol{\ell}_t, \mathbf{p}_t \rangle_{\mathcal{I}_t} - \ell_{t,i}$ . Using the fact that

$$\sum_{i: \pi_i \mathcal{I}_{t,i} w_{t,i} > 0} p_{t,i} r_{t,i} = \sum_{i: \mathcal{I}_{t,i} = 1} p_{t,i} r_{t,i} = \sum_{i: \mathcal{I}_{t,i} = 1} p_{t,i} \langle \boldsymbol{\ell}_t, \mathbf{p}_t \rangle_{\mathcal{I}_t} - \sum_{i: \mathcal{I}_{t,i} = 1} p_{t,i} \ell_{t,i} = 0,$$

we have

$$\begin{aligned} \sum_{i=1}^N \pi_i \mathcal{I}_{t,i} g_{t,i} w_{t,i} &= \sum_{i: \pi_i \mathcal{I}_{t,i} w_{t,i} > 0} \pi_i [w_{t,i}]_+ r_{t,i} + \sum_{i: \pi_i \mathcal{I}_{t,i} w_{t,i} \leq 0} \pi_i \mathcal{I}_{t,i} w_{t,i} [r_{t,i}]_+ \\ &= \|\widehat{\mathbf{p}}_t\|_1 \sum_{i: \pi_i \mathcal{I}_{t,i} w_{t,i} > 0} p_{t,i} r_{t,i} + \sum_{i: \pi_i \mathcal{I}_{t,i} w_{t,i} \leq 0} \pi_i \mathcal{I}_{t,i} w_{t,i} [r_{t,i}]_+ \\ &= 0 + \sum_{i: \pi_i \mathcal{I}_{t,i} w_{t,i} \leq 0} \pi_i \mathcal{I}_{t,i} w_{t,i} [r_{t,i}]_+ \leq 0. \end{aligned}$$

Then, because of the property (7) of the coin betting potentials, we have

$$\sum_{i=1}^N \pi_i F_{T,i}(z_{1:T,i}) \leq 1 + \sum_{i=1}^N \pi_i \sum_{t=1}^T \mathcal{I}_{t,i} g_{t,i} w_{t,i} \leq 1, \tag{8}$$

and

$$\begin{aligned} \sum_{i=1}^N u_i \log(F_{T,i}(z_{1:T,i})) &= \sum_{i=1}^N u_i \left( \log \left( \frac{u_i}{\pi_i} \right) + \log \left( \frac{\pi_i}{u_i} \cdot F_{T,i}(z_{1:T,i}) \right) \right) \\ &\stackrel{\text{(Jensen's)}}{\leq} \text{KL}(\mathbf{u} \parallel \boldsymbol{\pi}) + \log \left( \sum_{i=1}^N u_i \cdot \frac{\pi_i}{u_i} \cdot F_{T,i}(z_{1:T,i}) \right) \end{aligned}$$

$$\stackrel{(8)}{\leq} \text{KL}(\mathbf{u} \parallel \boldsymbol{\pi}) .$$

Define  $Z_{T,i} := \sum_{t=1}^T z_{t,i}$  and  $H'_{T,i}(x) := c_1 \frac{x^2}{S_{T,i}} + c_{2,i}$ . We see from the definition of  $H_{T,i}$  in Theorem 2 that  $H'_{T,i}(Z_{T,i}) = H_{T,i}(z_{1:T,i})$ . Since  $H'$  is symmetric around 0, its inverse  $H'^{-1}$  usually maps to two distinct values with opposite sign. To resolve this ambiguity, we define it to map to the nonnegative real value. Then, for any comparator  $\mathbf{u} \in \Delta^N$ , we have

$$\begin{aligned} \text{Regret}_T(\mathbf{u}) &= \sum_{t=1}^T \sum_{i=1}^N \mathcal{I}_{t,i} u_i (\langle \boldsymbol{\ell}_t, \mathbf{p}_t \rangle - \ell_{t,i}) \leq \sum_{t=1}^T \sum_{i=1}^N \mathcal{I}_{t,i} u_i g_{t,i} = \sum_{i=1}^N u_i Z_{T,i} \\ &= \sum_{i=1}^N u_i H'^{-1}_{T,i}(H'_{T,i}(Z_{T,i})) = \sum_{i=1}^N u_i H'^{-1}_{T,i}(H_{T,i}(z_{1:T,i})) \\ &\leq \sum_{i=1}^N u_i H'^{-1}_{T,i}(\log(F_{T,i}(z_{1:T,i}))) \\ &= \sum_{i=1}^N u_i \sqrt{c_1^{-1} S_{T,i} \cdot (\log(F_{T,i}(z_{1:T,i})) - c_{2,i})} \\ &= \sum_{i=1}^N \sqrt{u_i c_1^{-1} S_{T,i}} \cdot \sqrt{u_i (\log(F_{T,i}(z_{1:T,i})) - c_{2,i})} \\ &\stackrel{(a)}{\leq} \sqrt{c_1^{-1} \left( \sum_{i=1}^N u_i S_{T,i} \right) \cdot \left( \sum_{i=1}^N u_i (\log(F_{T,i}(z_{1:T,i})) - c_{2,i}) \right)} \\ &\leq \sqrt{c_1^{-1} \left( \sum_{i=1}^N u_i S_{T,i} \right) \cdot \left( \text{KL}(\mathbf{u} \parallel \boldsymbol{\pi}) - \sum_{i=1}^N u_i c_{2,i} \right)}, \end{aligned}$$

where (a) is due to the Cauchy-Schwartz inequality (noting that the factors under the square root are all nonnegative since  $\log F_{T,i}(x) \geq H_{T,i}(x)$ ).  $\square$

Note that if  $\mathbf{u} = \mathbf{e}_j$ , then the regret scales with  $S_{T,j}$ , which is essentially the number of time steps at which expert  $j$  is awake.

While any potential function satisfying the condition (7) and symmetricity around 0 can be used, we present two interesting choices: the Krichevsky-Trofimov potential and the AdaptiveNormal potential.

**Krichevsky-Trofimov potential** The Krichevsky-Trofimov (KT) potential [16] is defined as follows:

$$F_t(g_{1:t}) = \frac{2^t \cdot \Gamma(\delta + 1) \cdot \Gamma\left(\frac{t+\delta+1}{2} + \frac{\sum_{s=1}^t g_s}{2}\right) \cdot \Gamma\left(\frac{t+\delta+1}{2} - \frac{\sum_{s=1}^t g_s}{2}\right)}{\Gamma\left(\frac{\delta+1}{2}\right)^2 \cdot \Gamma(t + \delta + 1)}, \quad (9)$$

where  $\delta \geq 0$  is a time shift parameter set to 0 in this work. Orabona & Pál [16] show that the KT potential satisfies (5). We modify the KT potential as follows to handle sleeping experts by replacing  $t$  in several places by  $S_{t,i}$ :

$$F_{t,i}(z_{1:t,i}) = \frac{2^{S_{t,i}} \cdot \Gamma(\delta + 1) \cdot \Gamma\left(\frac{S_{t,i} + \delta + 1}{2} + \frac{\sum_{s=1}^t z_{s,i}}{2}\right) \cdot \Gamma\left(\frac{S_{t,i} + \delta + 1}{2} - \frac{\sum_{s=1}^t z_{s,i}}{2}\right)}{\Gamma\left(\frac{\delta + 1}{2}\right)^2 \cdot \Gamma(S_{t,i} + \delta + 1)}, \quad (10)$$

which satisfies (7).<sup>6</sup> The betting fraction  $\beta_t$  defined in (4a) with KT potential exhibits the simple form  $\beta_t = \frac{\sum_{s=1}^{t-1} g_s}{t + \delta}$  and, for sleeping experts, we have  $\beta_{t,i} = \frac{\sum_{s=1}^{t-1} z_{s,i}}{S_{t,i} + \delta}$ . We present the regret of Algorithm 1 with the KT potential in Corollary 3.

**Corollary 3.** *Let  $\delta = 0$ . Suppose we run Algorithm 1 with the KT potential. Then,*

$$\text{Regret}_T(\mathbf{u}) \leq \sqrt{2 \left( \sum_{i=1}^N u_i S_{T,i} \right) \cdot \left( \text{KL}(\mathbf{u} \parallel \boldsymbol{\pi}) + \frac{1}{2} \ln(T) + 2 \right)}.$$

*Proof.* Define  $H'_{T,i}(x) := \frac{x^2}{2S_{T,i}} + \frac{1}{2} \ln\left(\frac{1}{S_{T,i}}\right) - \ln(e\sqrt{\pi})$ . Note that this definition is identical with the one used in Theorem 2 if we set  $c_1 = \frac{1}{2}$  and  $c_{2,i} = \frac{1}{2} \ln\left(\frac{1}{S_{T,i}}\right) - \ln(e\sqrt{\pi})$ . According to Orabona & Pál [16, Lemma 15] with  $\delta = 0$ , we have  $H'_{T,i}(Z_{T,i}) = H_{T,i}(z_{1:T,i}) \leq \ln F_{T,i}(z_{1:T,i})$ . It follows that

$$-\sum_{i=1}^N u_i c_{2,i} = \sum_{i=1}^N u_i \left( (1/2) \ln(S_{T,i}) + \ln(e\sqrt{\pi}) \right) \leq \frac{1}{2} \ln(T) + 2.$$

By plugging  $c_1$  and  $c_{2,i}$  into Theorem 2, we obtain the result.  $\square$

**AdaptiveNormal potential** Let  $\bar{G}_t := \sum_{s=1}^t |g_s|$ . The AdaptiveNormal (AN) potential proposed by Orabona & Tommasi [17] is defined by

$$F_t(g_{1:t}) = \exp \left( \frac{(\sum_{s=1}^t g_s)^2}{2(\xi + \bar{G}_t)} - \sum_{s=1}^t \frac{|g_s|}{2(\xi + \bar{G}_{s-1} + 1)} \right), \quad (11)$$

where  $\xi > 0$  is a parameter of minor importance in our context that we set to 1. Orabona & Tommasi [17, Lemma 2] show that the AN potential satisfies the condition (5). Let  $\bar{Z}_{t,i} := \sum_{s=1}^t |z_{s,i}|$ . For sleeping experts, we use the following potential that satisfies (7) due to a trivial consequence of (11) satisfying (5):

$$F_{t,i}(z_{1:t,i}) = \exp \left( \frac{(\sum_{s=1}^t z_{s,i})^2}{2(\xi + \bar{Z}_{t,i})} - \sum_{s=1}^t \frac{|z_{s,i}|}{2(\xi + \bar{Z}_{s-1,i} + 1)} \right). \quad (12)$$

<sup>6</sup>This is trivially implied by the fact that (9) satisfies (5) since the only modification in (10) compared to (9) is to allow “individual clock”  $S_{t,i}$  that counts the time steps expert  $i$  was awake up to  $t$ .

The betting fraction (6a) using the AN potential can be simplified to

$$\beta_t = 2\sigma \left( \frac{2 \sum_{s=1}^{t-1} g_s}{\xi + \bar{G}_{t-1} + 1} \right) - 1, \quad \text{where } \sigma(z) = \frac{1}{1 + \exp(-z)}.$$

For sleeping experts, we have

$$\beta_{t,i} = 2\sigma \left( \frac{2 \sum_{s=1}^{t-1} z_{s,i}}{\xi + \bar{Z}_{t-1,i} + 1} \right) - 1.$$

Define  $r_{t,i} := \mathcal{I}_{t,i}(h_t(\mathbf{p}_t) - \ell_{t,i})$  and  $\tilde{L}_{T,i} := \sum_{t=1}^T [-r_{t,i}]_+$ . We present the regret bound of Sleeping CB with the AN potential in Corollary 4.

**Corollary 4.** *Let  $\xi = 1$ . Suppose we run Algorithm 1 with the AN potential. Let  $W_{\mathbf{u}} := 1 + \sum_{i=1}^N u_i \bar{Z}_{T,i}$  and  $L_{T,i} := \sum_{t=1}^T \mathcal{I}_{t,i} \ell_{t,i}$ . Then we have*

$$\text{Regret}_T(\mathbf{u}) = \sqrt{2W_{\mathbf{u}} (\text{KL}(\mathbf{u}||\boldsymbol{\pi}) + \frac{1}{2} \ln(W_{\mathbf{u}}))},$$

and moreover

$$\begin{aligned} & \text{Regret}_T(\mathbf{u}) \\ &= O \left( \sqrt{\left( \sum_{i=1}^N u_i L_{T,i} \right) \cdot (\text{KL}(\mathbf{u}||\boldsymbol{\pi}) + \ln(W_{\mathbf{u}})) + \text{KL}(\mathbf{u}||\boldsymbol{\pi}) + \ln(W_{\mathbf{u}})} \right). \end{aligned}$$

*Proof.* Define

$$H'_{T,i}(x) := \frac{x^2}{2(1 + \bar{Z}_{T,i})} - \sum_{s=1}^T \frac{|z_{s,i}|}{2(1 + \bar{Z}_{s-1,i} + 1)}.$$

Note that  $H'_{T,i}(Z_{T,i}) = H_{T,i}(z_{1:T,i}) = \ln(F_{T,i}(z_{1:T,i}))$ . To match this definition with the setup of Theorem 2, we redefine  $S_{T,i} := 1 + \bar{Z}_{T,i}$ , for which the theorem still holds, since  $S_{T,i}$  is used only through the function  $H_{T,i}$ . We also set

$$c_1 = \frac{1}{2}, \quad c_{2,i} = - \sum_{s=1}^T \frac{|z_{s,i}|}{2(1 + \bar{Z}_{s-1,i} + 1)}.$$

Using  $|z_{s,i}| \leq 1$  and  $\frac{a-b}{a} \leq \ln(a) - \ln(b)$ , we have

$$\begin{aligned} -c_{2,i} &\leq \frac{1}{2} \sum_{s=1}^T \frac{|z_{s,i}|}{(1 + \bar{Z}_{s-1,i} + |z_{s,i}|)} \\ &= \frac{1}{2} \sum_{s=1}^T (\ln(1 + \bar{Z}_{s,i}) - \ln(1 + \bar{Z}_{s-1,i})) = \frac{1}{2} \ln(1 + \bar{Z}_{T,i}), \end{aligned}$$

so it follows from Jensen's inequality that

$$\sum_{i=1}^N -c_{2,i} \leq \frac{1}{2} \ln \left( 1 + \sum_{i=1}^N u_i \bar{Z}_{T,i} \right).$$

Then, by Theorem 2, we have

$$\text{Regret}_T(\mathbf{u}) = \sqrt{2 \left( 1 + \sum_{i=1}^N u_i \bar{Z}_{T,i} \right) \cdot \left( \text{KL}(\mathbf{u} \parallel \boldsymbol{\pi}) + \frac{1}{2} \log \left( 1 + \sum_{i=1}^N u_i \bar{Z}_{T,i} \right) \right)},$$

proving the first statement of the theorem.

For the second statement, we use Lemma 13 in Section A.5, which says if  $\text{Regret}_T(\mathbf{u}) \leq \sqrt{(1 + \sum_{i=1}^N u_i \bar{Z}_{T,i}) A(\mathbf{u})}$  for some function  $A(\mathbf{u})$  then

$$\text{Regret}_T(\mathbf{u}) = \sqrt{(1 + 2 \sum_{i=1}^N u_i \tilde{L}_{T,i}) A(\mathbf{u}) + A(\mathbf{u})}.$$

Setting  $A(\mathbf{u}) = \text{KL}(\mathbf{u} \parallel \boldsymbol{\pi}) + \frac{1}{2} \log \left( 1 + \sum_{i=1}^N u_i \bar{Z}_{T,i} \right)$  and the fact that  $[-r_{t,i}]_+ \leq \ell_{t,i} \implies \tilde{L}_{T,i} \leq L_{T,i}$  we verify the second claim.  $\square$

When we set  $\mathbf{u} = \mathbf{e}_i$  with this AN potential, we obtain a regret bound that scales with  $L_{T,i}$ , which is always smaller than  $S_{T,i}$ . The difference becomes significant when the expert  $i$  suffers a loss  $\ell_{t,i}$  that is close to 0 for all  $t \in \text{Active}(i)$ . Note that Sleeping CB with the AN potential is quite similar to AdaNormalHedge [14], which has the same regret order. The key difference is that the truncation operates in the potential function for AdaNormalHedge whereas for ours it operates in the reduction to LEA (see the definition of  $g_{t, \mathcal{B}_J}$ ).

According to our results, the regret bound of the KT potential can be much larger than that of the AN potential. Thus, one might wonder if we should always use the AN potential. Our empirical study in Section 5 shows a case where KT has a benefit over AN.

#### 4. Coping with a changing environment by sleeping CB

In this section, we synthesize the results in Sections 2 and 3 to specify and analyze our meta algorithm. Recall that a meta algorithm must efficiently aggregate decisions from multiple black-box runs that are active at time  $t$ . By treating each black-box run as an expert, we use Sleeping CB (Algorithm 1) as the meta algorithm, with geometric covering intervals. An important motivation for the use of Sleeping CB is that it is parameter-free. Other sleeping bandits techniques require the number of black-box runs (experts) to be specified in advance, which results in a theoretical guarantee only up to some finite time horizon  $T$ . By contrast, our approach provides an “anytime” guarantee. The complete algorithm, which we call **Coin Betting for Changing Environments (CBCE)**, is shown in Algorithm 2.

We first present the results with the KT potential and then discuss applying the AN potential in the same manner. We make use of the following assumption.

---

**Algorithm 2** Coin Betting for Changing Environment (CBCE)

---

**Input:** A black-box algorithm  $\mathcal{B}$  and a prior distribution  $\pi \in \Delta^{|\mathcal{J}|}$  over  $\{\mathcal{B}_J \mid J \in \mathcal{J}\}$ .

**for**  $t = 1$  **to**  $T$  **do**

For each  $J \in \text{Active}(t)$ , set

$w_{t,\mathcal{B}_J} \leftarrow \beta_{t,\mathcal{B}_J}(z_{t,\mathcal{B}_J}) \cdot (1 + \sum_{s=1}^{t-1} z_{s,\mathcal{B}_J} w_{s,\mathcal{B}_J})$

Set  $\hat{p}_{t,\mathcal{B}_J} \leftarrow \pi_{\mathcal{B}_J} \mathcal{I}_{t,\mathcal{B}_J}[w_{t,\mathcal{B}_J}]_+$  for  $J \in \text{Active}(t)$  and 0 for  $J \notin \text{Active}(t)$ .

Compute  $\mathbf{p}_t \leftarrow \begin{cases} \hat{\mathbf{p}}_t / \|\hat{\mathbf{p}}_t\|_1 & \text{if } \|\hat{\mathbf{p}}_t\|_1 > 0 \\ [\pi_{\mathcal{B}_J}]_{J \in \text{Active}(t)} & \text{if } \|\hat{\mathbf{p}}_t\|_1 = 0. \end{cases}$

The black-box run  $\mathcal{B}_J$  picks a decision  $\mathbf{x}_t^{\mathcal{B}_J} \in \mathcal{D}$ ,  $\forall J \in \text{Active}(t)$ .

The learner picks a decision  $\mathbf{x}_t = \sum_{J \in \mathcal{J}} p_{t,\mathcal{B}_J} \mathbf{x}_t^{\mathcal{B}_J}$ .

Each black-box run  $\mathcal{B}_J$  that is awake ( $J \in \text{Active}(t)$ ) suffers loss  $\ell_{t,\mathcal{B}_J} := f_t(\mathbf{x}_t^{\mathcal{B}_J})$ .

The learner suffers loss  $f_t(\mathbf{x}_t)$ .

For each  $J \in \text{Active}(t)$ , set

$g_{t,\mathcal{B}_J} \leftarrow \begin{cases} f_t(\mathbf{x}_t) - \ell_{t,\mathcal{B}_J} & \text{if } w_{t,\mathcal{B}_J} > 0 \\ [f_t(\mathbf{x}_t) - \ell_{t,\mathcal{B}_J}]_+ & \text{if } w_{t,\mathcal{B}_J} \leq 0. \end{cases}$

**end for**

---

**Assumption A1.** The loss function  $f_t$  is convex and maps to  $[0, 1]$ ,  $\forall t \in \mathbb{N}$ .

Nonconvex loss functions can be accommodated by randomized decisions: We choose the decision  $\mathbf{x}_t^{\mathcal{B}_J}$  from black-box  $\mathcal{B}_J$  with probability  $p_{t,\mathcal{B}_J}$ . It is not difficult to show that the same regret bound holds, but now in expectation. When loss functions are unbounded, they can be scaled and truncated to  $[0, 1]$ . Any nonconvexity that results can be handled in the manner just described.

We define our choice of prior  $\bar{\pi} \in \Delta^{|\mathcal{J}|}$  as follows:

$$\bar{\pi}_{\mathcal{B}_J} := Z'^{-1} \left( \frac{1}{J_1^2 (1 + \lceil \log_2 J_1 \rceil)} \right), \quad \forall J \in \mathcal{J}, \tag{13}$$

where  $Z'$  is a normalization factor. Since there exist at most  $1 + \lceil \log_2 J_1 \rceil$  distinct intervals starting at time  $J_1$ , we have that  $Z' < \sum_{t=1}^{\infty} t^{-2} = \pi^2/6$ .

We bound the meta regret with respect to a black-box run  $\mathcal{B}_J$  as follows.

**Lemma 5.** (Meta regret of CBCE with the KT potential) *Assume A1. Suppose we run CBCE (Algorithm 2) with a black-box algorithm  $\mathcal{B}$ , prior  $\bar{\pi}$ , and the KT potential (10) with  $\delta = 0$ . The meta regret of CBCE( $\mathcal{B}$ ) on interval  $J = [J_1 \dots J_2] \in \mathcal{J}$  is*

$$\sum_{t \in J} f_t(\mathbf{x}_t^{\text{CBCE}(\mathcal{B})}) - f_t(\mathbf{x}_t^{\mathcal{B}_J}) \leq \sqrt{|J| (7 \ln(J_2) + 5)} = O(\sqrt{|J| \log J_2}).$$

*Proof.* Note that our regret definition for meta algorithms

$$\sum_{t \in J} f_t(\mathbf{x}_t^{\text{CBCE}(\mathcal{B})}) - f_t(\mathbf{x}_t^{\mathcal{B}_J}), \tag{14}$$



is slightly different from that of Theorem 2 for  $\mathbf{u} = \mathbf{e}_i$ :  $\sum_{t \in J: \mathcal{I}_{t,i}=1} \langle \ell_t, \mathbf{p}_t \rangle - \ell_{t,i}$ . This translates to, in the language of meta algorithms,  $\sum_{t \in J: \mathcal{I}_{t, \mathcal{B}_J}=1} \langle \ell_t, \mathbf{p}_t \rangle_{\mathcal{I}_t} - \ell_{t, \mathcal{B}_J}$  for  $\mathbf{u} = \mathbf{e}_{\mathcal{B}_J}$  (recall  $\ell_{t, \mathcal{B}_J} = f_t(\mathbf{x}_t^{\mathcal{B}_J})$ ).

We claim that Theorem 2 and Corollary 3 for hold true for the regret (14). Note that, using Jensen’s inequality,  $f_t(\mathbf{x}_t^{\text{CBCE}(\mathcal{B})}) \leq \langle \ell_t, \mathbf{p}_t \rangle_{\mathcal{I}_t}$ . Then, in the proof of Theorem 2

$$\begin{aligned} & \sum_{J \in \mathcal{J}} \pi_{\mathcal{B}_J} \mathcal{I}_{t, \mathcal{B}_J} g_{t, \mathcal{B}_J} w_{t, \mathcal{B}_J} \\ &= \sum_{J \in \mathcal{J}: \pi_{\mathcal{B}_J} \mathcal{I}_{t, \mathcal{B}_J} w_{t, \mathcal{B}_J} > 0} \pi_{\mathcal{B}_J} [w_{t, \mathcal{B}_J}]_+ (f_t(\mathbf{x}_t^{\text{CBCE}(\mathcal{B})}) - \ell_{t, \mathcal{B}_J}) \\ & \quad + \sum_{J \in \mathcal{J}: \pi_{\mathcal{B}_J} \mathcal{I}_{t, \mathcal{B}_J} w_{t, \mathcal{B}_J} \leq 0} \pi_{\mathcal{B}_J} \mathcal{I}_{t, \mathcal{B}_J} w_{t, \mathcal{B}_J} [f_t(\mathbf{x}_t^{\text{CBCE}(\mathcal{B})}) - \ell_{t, \mathcal{B}_J}]_+ \\ & \leq \sum_{J \in \mathcal{J}: \pi_{\mathcal{B}_J} \mathcal{I}_{t, \mathcal{B}_J} w_{t, \mathcal{B}_J} > 0} \pi_{\mathcal{B}_J} [w_{t, \mathcal{B}_J}]_+ (\langle \ell_t, \mathbf{p}_t \rangle_{\mathcal{I}_t} - \ell_{t, \mathcal{B}_J}) \\ & \quad + \sum_{J \in \mathcal{J}: \pi_{\mathcal{B}_J} \mathcal{I}_{t, \mathcal{B}_J} w_{t, \mathcal{B}_J} \leq 0} \pi_{\mathcal{B}_J} \mathcal{I}_{t, \mathcal{B}_J} w_{t, \mathcal{B}_J} [\langle \ell_t, \mathbf{p}_t \rangle_{\mathcal{I}_t} - \ell_{t, \mathcal{B}_J}]_+ . \end{aligned}$$

Then, one can see that the proof of Theorem 2 goes through, so does Corollary 3.

Since  $\text{KL}(\mathbf{e}_{\mathcal{B}_J} || \bar{\boldsymbol{\pi}}) = \ln \frac{1}{\bar{\pi}_{\mathcal{B}_J}} \leq \ln \left( \frac{\pi^2}{6} J_1^2 (1 + \lfloor \log_2 J_1 \rfloor) \right) \leq 3 \ln(J_2) + \frac{1}{2}$ , it follows that

$$\begin{aligned} \sum_{t \in J} f_t(\mathbf{x}_t^{\text{CBCE}(\mathcal{B})}) - f_t(\mathbf{x}_t^{\mathcal{B}_J}) & \stackrel{(\text{Cor. 3})}{\leq} \sqrt{2S_{T, \mathcal{B}_J} \cdot \left( \text{KL}(\mathbf{e}_{\mathcal{B}_J} || \boldsymbol{\pi}) + \frac{1}{2} \ln(J_2) + 2 \right)} \\ & \leq \sqrt{2|J| \left( \frac{7}{2} \ln(J_2) + \frac{5}{2} \right)} \\ & = \sqrt{|J| (7 \ln(J_2) + 5)} . \quad \square \end{aligned}$$

We now present the bound on the SA-Regret  $R_I^{\text{CBCE}(\mathcal{B})}(\mathbf{w})$  with respect to  $\mathbf{w} \in \mathcal{W}$  on intervals  $I \subseteq [T]$  that are not necessarily in  $\mathcal{J}$ .

**Theorem 6.** (SA-Regret of CBCE( $\mathcal{B}$ ) with the KT potential) *Take the same assumption as Lemma 5. Suppose that the black-box algorithm  $\mathcal{B}$  has regret  $R_T^{\mathcal{B}}$  bounded by  $A_1 T^\alpha$ , where  $\alpha \in (0, 1)$ . Let  $I = [I_1..I_2]$ . The SA-Regret of CBCE with black-box  $\mathcal{B}$  on the interval  $I$  with respect to any  $\mathbf{w} \in \mathcal{W}$  is bounded as follows:*

$$R_I^{\text{CBCE}(\mathcal{B})}(\mathbf{w}) \leq \frac{4}{2^\alpha - 1} A_1 |I|^\alpha + 8 \sqrt{|I| (7 \ln(I_2) + 5)} = O(A_1 |I|^\alpha + \sqrt{|I| \ln I_2}) .$$

*Proof.* By Lemma 1, we know that  $J$  can be decomposed into two sequences of intervals  $\{J^{(-a)}, \dots, J^{(0)}\}$  and  $\{J^{(1)}, J^{(2)}, \dots, J^{(b)}\}$ . Continuing from (2),

$$R_I^{\text{CBCE}(\mathcal{B})}(\mathbf{w})$$

$$= \underbrace{\sum_{i=-a}^b \sum_{t \in J^{(i)}} \left( f_t(\mathbf{x}_t^{\text{CBCE}(\mathcal{B})}) - f_t(\mathbf{x}_t^{\mathcal{B}_{J^{(i)}}}) \right)}_{=:E_1} + \underbrace{\sum_{i=-a}^b \sum_{t \in J^{(i)}} \left( f_t(\mathbf{x}_t^{\mathcal{B}_{J^{(i)}}}) - f_t(\mathbf{w}) \right)}_{=:E_2} .$$

Then,

$$E_1 = \sum_{i \in [(-a)..0]} \sum_{t \in J^{(i)}} \left( f_t(\mathbf{x}_t^{\text{CBCE}(\mathcal{B})}) - f_t(\mathbf{x}_t^{\mathcal{B}_{J^{(i)}}}) \right) + \sum_{i \in [1..b]} \sum_{t \in J^{(i)}} \left( f_t(\mathbf{x}_t^{\text{CBCE}(\mathcal{B})}) - f_t(\mathbf{x}_t^{\mathcal{B}_{J^{(i)}}}) \right) .$$

The first summation is upper-bounded by, due to Lemma 5 and Lemma 1,  $\sum_{i \in [(-a)..0]} \sqrt{|J^{(i)}|(7 \ln(I_2) + 5)} \leq \sqrt{7 \ln(I_2) + 5} \cdot \sum_{i=0}^{\infty} (2^{-i}|I|)^{1/2} \leq \sqrt{7 \ln(I_2) + 5} \cdot (4\sqrt{|I|})$ . The second summation is bounded by the same quantity due to symmetry. Thus,  $E_1 \leq 8\sqrt{|I|(7 \ln(I_2) + 5)}$ .

In the same manner, one can show that  $E_2 \leq 2 \cdot \frac{2^\alpha}{2^\alpha - 1} |I|^\alpha \leq \frac{4}{2^\alpha - 1} |I|^\alpha$ , which concludes the proof.  $\square$

For the standard LEA problem, one can run the algorithm CB with KT potential (equivalent to Sleeping CB with  $\mathcal{I}_{t,i} = 1, \forall t, i$ ), which achieves static regret  $O(\sqrt{T \log(NT)})$  [16]. Using CB as the black-box algorithm, the regret of  $\text{CBCE}(\mathcal{B})$  on  $I$  is  $R_T^{\text{CBCE}(\text{CB})}(\mathbf{w}) = O(\sqrt{|I| \log(NI_2)})$ , and so  $\text{SA-Regret}_T^{\text{CBCE}(\text{CB})}(|I|) = O(\sqrt{|I| \log(NT)})$ . It follows that the  $m$ -shift regret of  $\text{CBCE}(\text{CB})$  is  $O(\sqrt{mT \log(NT)})$  using the technique presented in Section A.2.

As said above, our bound improves over the best known result with the same time complexity in [5]. The key ingredient that allows us to get a better bound is the Sleeping CB Algorithm 1, that achieves a better SA-Regret than the one of [5]. In the next section, we will show that the empirical results also confirm the theoretical gap between these two algorithms.

**The AdaptiveNormal potential** We present the meta regret bound of CBCE with the AN potential on intervals  $J \in \mathcal{J}$  in Lemma 7 and on any interval  $I$  in Lemma 8.

**Lemma 7.** (Meta regret of CBCE with the AN potential) *Assume A1. Suppose we run CBCE (Algorithm 2) with a black-box algorithm  $\mathcal{B}$ , prior  $\bar{\pi}$ , and the AN potential (12) with  $\xi = 1$ . The meta regret of  $\text{CBCE}(\mathcal{B})$  on interval  $J = [J_1..J_2] \in \mathcal{J}$  is*

$$\sum_{t \in J} f_t(\mathbf{x}_t^{\text{CBCE}(\mathcal{B})}) - f_t(\mathbf{x}_t^{\mathcal{B}_J}) = O \left( \sqrt{\sum_{t \in J} f_t(\mathbf{x}_t^{\mathcal{B}_J}) \log(J_2) + \log(J_2)} \right) .$$

*Proof.* The proof deviates from the proof of Lemma 5. Since  $W_{\mathbf{e}_{\mathcal{B}_J}} = O(J_2)$ .

$$\begin{aligned} \sum_{t \in J} f_t(\mathbf{x}_t^{\text{CBCE}(\mathcal{B})}) - f_t(\mathbf{x}_t^{\mathcal{B}_J}) &\stackrel{(\text{Cor. 4})}{=} O\left(\sqrt{L_{T, \mathcal{B}_J} \log(J_2 W_{\mathbf{e}_{\mathcal{B}_J}})} + \log(J_2 W_{\mathbf{e}_{\mathcal{B}_J}})\right) \\ &= O\left(\sqrt{L_{T, \mathcal{B}_J} \log(J_2)} + \log(J_2)\right). \quad \square \end{aligned}$$

**Theorem 8.** (SA-Regret of CBCE( $\mathcal{B}$ ) with the AN potential) *Make the same same assumption as Lemma 5. Suppose that the black-box algorithm  $\mathcal{B}$  has regret  $R_T^{\mathcal{B}}(\mathbf{w})$  bounded by  $A_1(\sum_{t=1}^T q_t)^\alpha + A_2$  for some  $\{q_t\}$  where  $\alpha \in (0, 1)$  where  $A_2$  grows poly-logarithmically in  $T$ . For any  $I = [I_1..I_2]$  and  $\mathbf{w} \in \mathcal{W}$ ,  $R_I^{\text{CBCE}(\mathcal{B})}(\mathbf{w}) = O\left(\min\left\{A_1|I|^\alpha + \sqrt{|I|\log I_2}, U(\mathbf{w})\right\}\right)$  where  $U(\mathbf{w}) =$*

$$A_1(\log |I|)^{1-\alpha} \left(\sum_{t \in I} q_t\right)^\alpha + \log(I_2) \sqrt{\sum_{t \in I} f_t(\mathbf{w})},$$

and we ignore additive terms scaling at most poly-logarithmically in  $I_2$ .

*Proof.* The first equation in the minimum operator of the stated regret bound is trivial as  $\sum_{t \in J} f_t(\mathbf{x}_t^{\mathcal{B}_J}) \leq |J|$ . Thus, we focus on  $U(\mathbf{w})$ . By Lemma 1, we know that  $J$  can be decomposed into two sequences of intervals  $\{J^{(-a)}, \dots, J^{(0)}\}$  and  $\{J^{(1)}, J^{(2)}, \dots, J^{(b)}\}$ . Continuing from (2),

$$\begin{aligned} R_I^{\text{CBCE}(\mathcal{B})}(\mathbf{w}) &= \underbrace{\sum_{i=-a}^b \sum_{t \in J^{(i)}} \left(f_t(\mathbf{x}_t^{\text{CBCE}(\mathcal{B})}) - f_t(\mathbf{x}_t^{\mathcal{B}_{J^{(i)}}})\right)}_{=: E_1} + \underbrace{\sum_{i=-a}^b \sum_{t \in J^{(i)}} \left(f_t(\mathbf{x}_t^{\mathcal{B}_{J^{(i)}}}) - f_t(\mathbf{w})\right)}_{=: E_2}. \end{aligned}$$

Define  $L_{J^{(i)}}^{\mathcal{B}} := \sum_{t \in J^{(i)}} f_t(\mathbf{x}_t^{\mathcal{B}_{J^{(i)}}})$ . For simplicity, we denote by  $C_{1,i} \sqrt{L_{J^{(i)}}^{\mathcal{B}}} + C_{2,i}$  the meta regret bound of CBCE for the interval  $J^{(i)}$  (see Lemma 7). Define  $\bar{C}_1 = \max_i C_{1,i}$ . Then, due to Lemma 7 and the fact  $\sum_{i=1}^k x_i^\alpha \leq k^{1-\alpha} (\sum_{i=1}^k x_i)^\alpha$  for  $\alpha \in (0, 1)$ ,

$$\begin{aligned} E_1 &\leq \sum_i C_{1,i} \sqrt{L_{J^{(i)}}^{\mathcal{B}}} + C_{2,i} \\ &= O(\bar{C}_1 \sqrt{\log |I|} \sqrt{\sum_i L_{J^{(i)}}^{\mathcal{B}}} + \sum_i C_{2,i}) \\ &= O(\bar{C}_1 \sqrt{\log |I|} \sqrt{\sum_{t \in I} f_t(\mathbf{w}) + \sum_i \sum_{t \in J^{(i)}} f_t(\mathbf{x}_t^{\mathcal{B}_{J^{(i)}}}) - f_t(\mathbf{w})} + \sum_i C_{2,i}) \\ &= O(\bar{C}_1 \sqrt{\log |I|} \sqrt{\sum_{t \in I} f_t(\mathbf{w}) + E_2} + \sum_i C_{2,i}). \end{aligned}$$

Note that

$$E_2 = O(A_1(\log |I|)^{1-\alpha} (\sum_{t \in I} q_t)^\alpha + A_2 \log |I|).$$

Note that  $\bar{C}_{1,i} = O(\sqrt{\log I_2})$  and  $C_{2,i} = O(\log I_2)$ . For brevity, we ignore the term  $\sqrt{E_2}$  that is smaller than  $E_2$  unless  $E_2 < 1$ . Ignoring terms that cannot grow faster than poly-logarithmically in  $I_2$ , the regret of  $\text{CBCE}\langle\mathcal{B}\rangle$  for interval  $I$  can be simplified to

$$O(A_1(\log |I|)^{1-\alpha}(\sum_{t \in I} q_t)^\alpha + \log(I_2)\sqrt{\sum_{t \in I} f_t(\mathbf{w})}). \quad \square$$

We emphasize that the order of regret stated in Theorem 8 is always no larger than  $\text{CBCE}$  with the KT potential. Furthermore, the regret bound of the AN potential scales roughly with  $\min_{w \in \mathcal{W}} \sum_{t \in I} f_t(\mathbf{w}) + (\sum_{t \in I} q_t)^\alpha$ . In some cases, this form of regret can be much smaller when there exists a decision  $\mathbf{w}$  that has a very small loss in the interval  $I$ . We instantiate the result above for LEA in Corollary 9 and for online convex optimization (OCO) in Corollary 10.

**Corollary 9.** (SA-Regret of  $\text{CBCE}\langle\text{CB}\rangle$  with the AN potential) *Suppose we run  $\text{CBCE}$  with the AN potential equipped with CB with the AN potential for LEA. Then, for any  $I = [I_1..I_2]$ ,*

$$\text{SA-Regret}_T^{\text{CBCE}\langle\text{CB}\rangle}(I) = O\left(\log(I_2)\sqrt{\left(\min_{i \in [N]} \sum_{t \in I} \ell_{t,i}\right) \log N + \text{polylog}(I_2)}\right).$$

*Proof.* Consider the standard LEA problem where all experts are awake all the time. Verify that the regret of CB with the AN potential with respect to expert  $i$  is  $O(\sqrt{L_{T,i} \log(NT)})$  by Corollary 4, ignoring additive terms that are  $O(\text{polylog}(T))$ . Plugging in  $A_1 = \sqrt{\log(N|I|)}$  and  $\alpha = 1/2$  in Theorem 8 concludes the proof.  $\square$

For LEA,  $\text{AdaNormalHedge.TV}$  of Luo & Schapire [14] achieves the SA-Regret bound  $O\left(\sqrt{(\min_{i \in [N]} \sum_{t \in I} \ell_{t,i}) \log(NI_1) + \text{polylog}(I_2)}\right)$ , which is at least  $\sqrt{\log(I_2)}$  factor smaller than  $\text{CBCE}\langle\text{CB}\rangle$ . However, the time complexity of  $\text{AdaNormalHedge.TV}$  is  $O(NT)$  per time step rather than  $O(N \log(T))$ .

**Corollary 10.** (SA-Regret of  $\text{CBCE}\langle\text{FTRL}\rangle$  with the AN potential) *Consider the OCO problem where the functions  $\{f_t\}$  are nonnegative, smooth (gradients are Lipschitz continuous), and  $L$ -Lipschitz. Suppose we run  $\text{CBCE}$  with the AN potential and use Follow The Regularized Leader (FTRL) on the linearized losses as the blackbox [18, 15], setting the regularizer at time  $t$  to  $\frac{\sqrt{L^2 + \sum_{s=1}^{t-1} \|\nabla f_s(\mathbf{x}_s)\|_2^2}}{2}$ . Then, for any  $I = [I_1..I_2]$ ,*

$$\text{SA-Regret}_T^{\text{CBCE}\langle\text{OGD}\rangle}(I) = O\left(\log(I_2)\sqrt{\min_{\mathbf{w} \in \mathcal{W}} \sum_{t \in I} f_t(\mathbf{w}) + \text{polylog}(I_2)}\right).$$

*Proof.* One can show that the regret bound of FTRL with the assumed regularizer achieves the regret bound of order  $O(\sqrt{L^2 + \sum_{t=1}^T f_t(\mathbf{w})})$  with respect to any  $\mathbf{w}$ , see, e.g., [18, Theorem 2]. Then, plugging in  $\alpha = 1/2$  in Theorem 8 concludes the proof.  $\square$

To the best of our knowledge, Corollary 10 is the first first-order SA-Regret bound for OCO.

**Discussion** Note that one can obtain the same result using the data streaming intervals (DS) [9, 8] in place of the geometric covering intervals (GC). Section A.3 elaborates on this with a lemma stating that DS induces a partition of an interval  $I$  in a very similar way to GC (a sequence of intervals of doubling lengths).

Our improved bound has another interesting implication. In designing strongly-adaptive algorithms for LEA, there is a well known technique called “restarts” or “sleeping experts” that has time complexity  $O(NT^2)$  [9, 14], and several studies used DS or GC to reduce the time complexity to  $O(NT \log T)$  [9, 8, 5]. However, it was unclear whether it is possible to achieve both an  $m$ -shift regret of  $O(\sqrt{mT}(\log N + \log T))$  and a time complexity of  $O(NT \log T)$  without knowing  $m$ . Indeed, every study on  $m$ -shift regret with time  $O(NT \log T)$  results in suboptimal  $m$ -shift regret bounds [5, 8, 9], to our knowledge. Furthermore, some studies (e.g., [14, Section 5]) speculated that perhaps applying the data streaming technique would increase its SA-Regret by a logarithmic factor. Our analysis implies that one can reduce the overall time complexity to  $O(NT \log T)$  without sacrificing the order of SA-Regret and  $m$ -shift regret.<sup>7</sup>

## 5. Experiments

We now turn to an empirical evaluation of algorithms for changing environments. We compare the performance of the meta algorithms under two online learning problems: (i) learning with expert advice (LEA) and (ii) metric learning (ML). We compare CBCE with SAOL [5] and AdaNormalHedge.TV (ATV) [14]. Although ATV was originally designed for LEA only, it is not hard to extend it to a meta algorithm and show that it has the same order of SA-Regret as CBCE using the same techniques. We run CBCE with both KT and AN potential, which are denoted by CBCE(KT) and CBCE(AN) respectively.

For our empirical study, we replace the geometric covering intervals (GC) with the data streaming intervals (DS) [9, 8]. Let  $u(t)$  be a number such that  $2^{u(t)}$  is the largest power of 2 that divides  $t$ ; e.g.,  $u(12) = 2$ . The data streaming intervals are  $\mathcal{J} = \{[t..(t + g \cdot 2^{u(t)} - 1)] : t = 1, 2, \dots\}$  for some  $g \geq 1$ . DS is an attractive alternative, unlike GC, (i) DS initiates one and only one black-box run at each time, and (ii) it is more flexible in that the parameter  $g$  can be increased to enjoy smaller regret in practice while increasing the time complexity by a constant factor.

For both ATV and CBCE, we set the prior  $\pi$  over the black-box runs as the uniform distribution. Note that this does not break the theoretical guarantees since the number of black-box runs is never actually infinite; we used  $\bar{\pi}$  (13) for ease of exposition.

<sup>7</sup>Note that we still pay an extra logarithmic factor when it comes to the *first-order* regret bound as discussed right below Corollary 9.

### 5.1. Learning with expert advice (LEA)

We consider LEA with the linear loss. That is, the loss function at time  $t$  is  $h_t(\mathbf{p}) = \boldsymbol{\ell}_t^\top \mathbf{p}$ . We draw linear losses  $\boldsymbol{\ell}_t \in [0, 1]^N, \forall t = 1, \dots, 600$  for  $N = 1000$  experts by setting  $\ell_{t,i}$  as the absolute value of an i.i.d. sample from  $\mathcal{N}(0, 0.5^2)$ . Then, for time  $t \in [1, 300]$ , we reduce loss of expert 1 by subtracting  $1/2$  from its loss:  $\ell_{t,1} \leftarrow [\ell_{t,1} - 1/2]_+$ . For time  $t \in [301, 600]$  and  $t \in [601, 900]$ , we perform the same for expert 2 and 3, respectively. Thus, the best expert is 1, 2, and 3 for time segment  $[1, 300]$ ,  $[301, 600]$ , and  $[601, 900]$ , respectively. Finally, we cap  $\ell_{t,i}$  below 1:  $\ell_{t,i} \leftarrow \min\{\ell_{t,i}, 1\}$ . We use the data streaming intervals with  $g = 2$ . In all our experiments, DS with  $g = 2$  outperforms GC while spending roughly the same time.

For each meta algorithm, we use Sleeping CB with the AN potential as the black-box,<sup>8</sup> where  $\mathcal{I}_{t,i} = 1$  for all  $t \geq 1$  and  $i \in [N]$  as there are no sleeping experts in this experiment. We warm-start each black-box run at time  $t \geq 2$  by setting its prior to the decision  $\mathbf{p}_{t-1}$  chosen by the meta algorithm at time step  $t - 1$ . We repeat the experiment 200 times and plot their average loss by computing moving mean with window size 10 in Figure 3(a). The overall winner is CBCE(AN). While CBCE(KT) catches up with the environmental change faster than CBCE(AN), CBCE(AN) shows smaller loss than CBCE(AN) once the change settles down. ATV is outperformed by both CBCEs but outperforms SAOL. Note that SAOL with GC intervals (SAOL-GC) tends to incur larger loss than the SAOL with DS. We observe that this is true for every meta algorithm, so we omit the result here to avoid clutter. We also run Fixed Share using the parameters recommended by Corollary 5.1 of [3], which requires knowing the target time horizon  $T = 900$  and the true number of switches  $m = 2$ . Such a strong assumption is often unrealistic in practice. Furthermore, we observe that Fixed Share is the slowest in adapting to the environmental changes. Nevertheless, Fixed Share can be attractive since (i) after the switch has settled down its loss is competitive to CBCE(AN), and (ii) its time complexity ( $O(NT)$ ) is lower than other algorithms ( $O(NT \log T)$ ).

### 5.2. Metric learning

We consider the problem of learning squared Mahalanobis distance from pairwise comparisons using the mirror descent algorithm [13]. The data point at time  $t$  is  $(\mathbf{z}_t^{(1)}, \mathbf{z}_t^{(2)}, y_t)$ , where  $y_t \in \{1, -1\}$  indicates whether or not  $\mathbf{z}_t^{(1)} \in \mathbb{R}^d$  and  $\mathbf{z}_t^{(2)} \in \mathbb{R}^d$  belongs to the same class. The goal is to learn a squared Mahalanobis distance parameterized by a positive semi-definite matrix  $\mathbf{M}$  and a bias  $\mu$  that have small loss  $f_t([\mathbf{M}; \mu]) :=$

$$[1 - y_t(\mu - (\mathbf{z}_t^{(1)} - \mathbf{z}_t^{(2)})^\top \mathbf{M}(\mathbf{z}_t^{(1)} - \mathbf{z}_t^{(2)}))]_+ + \rho \|\mathbf{M}\|_* ,$$

<sup>8</sup> We also experimented with Sleeping CB with the KT potential, but we found that it works slightly worse than the AN potential in general. We omit this result here to avoid clutter.

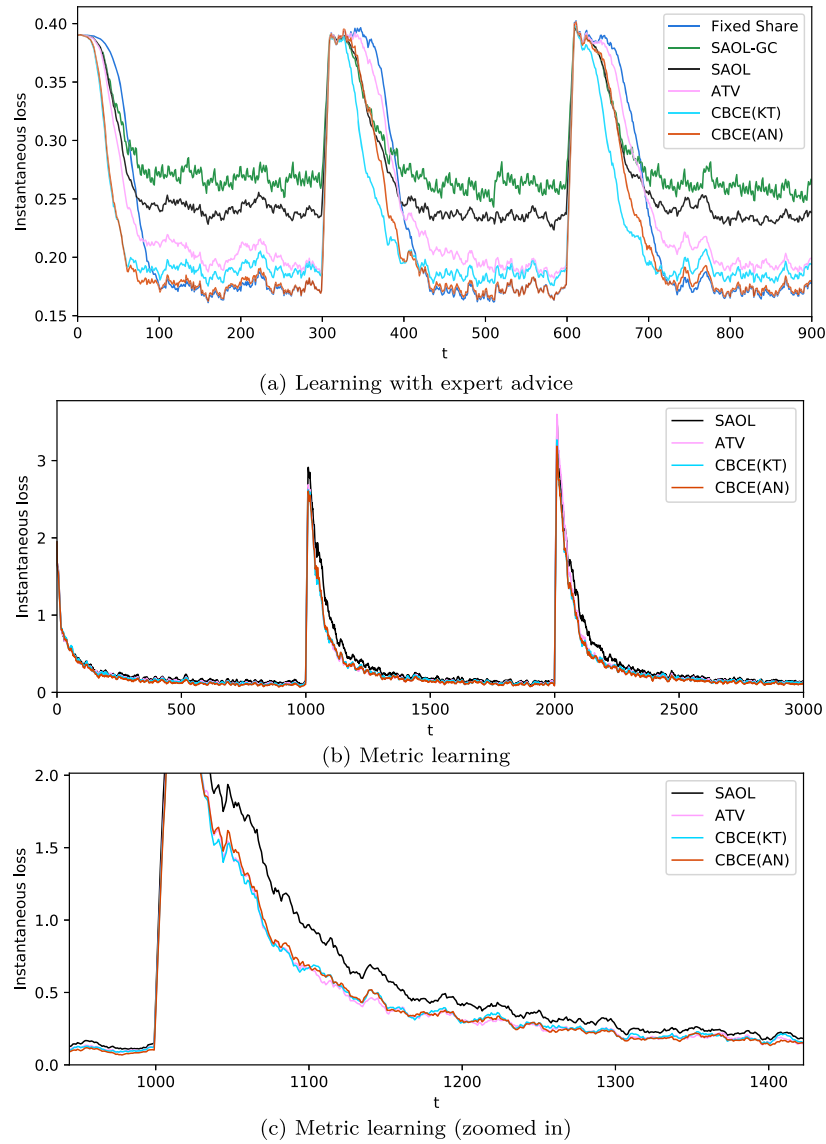


FIG 3. *Experiment results: Our method CBCE outperforms several baseline methods.*

where  $\mu$  is the bias parameter and  $\|\cdot\|_*$  is the trace norm. Such a formulation encourages predicting  $y_t$  with a large margin and with a low rank matrix  $\mathbf{M}$ . A learned matrix  $\mathbf{M}$  that has low rank can be useful in a number of machine learning tasks; e.g., distance-based classifications, clusterings, and low-dimensional embeddings. We refer to [13] for details.

We create a scenario that exhibits shifts in the metric, which is inspired by [7]. Specifically, we create a mixture of three Gaussians in  $\mathbb{R}^3$  whose means

are well-separated, and mixture weights are .5, .3, and .2. We draw 2000 points from it while keeping a record of their memberships. We repeat this three times independently and concatenate these three vectors to have 2000 9-dimensional vectors. Finally, we append to each point a 16-dimensional vector filled with Gaussian noise to have 25-dimensional vectors. Such a construction implies that for each point there are three independent cluster memberships. We run each algorithm for 1500 time steps. For time 1 to 500, we randomly pick a pair of points from the data pool and assign  $y_t = 1$  ( $y_t = -1$ ) if the pair belongs to the same (different) cluster under the first clustering. For time 501 to 1000 (1001 to 1500), we perform the same but under the second (third) clustering. In this way, a learner must track the change in the metric, especially the important low-dimensional subspaces for each time segment.

Since the loss of the metric learning is unbounded, we scale the loss by multiplying  $1/5$  and then capping it above at 1 as in [7]. Although the randomized decision discussed in Section 4 can be used to maintain the theoretical guarantee, we stick to the weighted average since the event that the loss being capped at 1 is rare in our experiments. As in our LEA experiment, we use the data streaming intervals with  $g = 2$  and initialize each black-box algorithm with the decision of the meta algorithm at the previous time step. We repeat the experiment 200 times and plot their average loss in Figure 3(b) by moving mean with window size 10. While we observe that CBCE(KT), CBCE(AN), and ATV are indistinguishable (see Figure 3(c)), all these methods outperform SAOL. We have verified that visible gaps in Figure 3 are statistically significant. This confirms the improved regret bound of CBCE and ATV.

## 6. Future work

Among a number of interesting directions, we are interested in reducing the time complexity in online learning within a changing environment. For LEA, Fixed Share has the best time complexity. However, Fixed Share is inherently not parameter-free; especially, it requires the knowledge of the number of shifts  $m$ . Achieving the best  $m$ -shift regret bound without knowing  $m$  or the best SA-Regret bound in time  $O(NT)$  would be an interesting future work. The same direction is interesting for the online convex optimization (OCO) problem. It would be interesting if an OCO algorithm such as online gradient descent can have the same SA-Regret as CBCE(OGD) without paying extra order of computation.

## Appendix

### A.1. The coin betting potential

We precisely define the coin betting potential. In this paper, we set  $\epsilon = 1$  throughout. For technical reasons, we define the potential function that takes a form of  $F_t(x; y_{1:t})$ . We then define  $F_t(y_{1:t}) := \bar{F}_t(\sum_{s=1}^t y_s; y_{1:t})$ .



**Definition 11.** (Coin Betting Potential [16]) Let  $\epsilon > 0$ . Let  $\{\bar{F}_t\}_{t=0}^\infty$  be a sequence of functions  $\bar{F}_t : (-a_t, a_t) \times [-1, 1]^t \rightarrow \mathbb{R}_+$  where  $a_t > t$ . The sequence  $\{\bar{F}_t\}_{t=0}^\infty$  is called a **sequence of coin betting potentials for initial endowment  $\epsilon$** , if it satisfies the following three conditions:

- (a)  $\bar{F}_0(0; \cdot) = \epsilon$ .
- (b) For every  $t \geq 0$ ,  $\bar{F}_t(x; y_{1:t})$  is even, logarithmically convex, strictly increasing on  $[0, a_t)$  in the first argument, and  $\lim_{x \rightarrow a_t} \bar{F}_t(x; y_{1:t}) = +\infty$ .
- (c) Define  $\beta_t := \frac{\bar{F}_t(x+1; y_{1:t-1}, 1) - \bar{F}_t(x-1; y_{1:t-1}, -1)}{\bar{F}_t(x+1; y_{1:t-1}, 1) + \bar{F}_t(x-1; y_{1:t-1}, -1)}$ . For every  $t \geq 1$  every  $x \in [-(t-1), (t-1)]$  and every  $g \in [-1, 1]$ ,

$$(1 + g\beta_t)\bar{F}_{t-1}(x; y_{1:t-1}) \geq \bar{F}_t(x + g; y_{1:t-1}, g) .$$

We now describe how the conditions for the coin betting potential lead to a lowerbound on the wealth:

$$\text{Wealth}_t \geq F_t(g_{1:t})$$

for any  $g_1, g_2, \dots, g_t \in [-1, 1]$ . We use induction. First, verify that  $\text{Wealth}_0 \geq F_0(\cdot) = \epsilon$ , trivially. Assuming  $\text{Wealth}_{t-1} \geq F_{t-1}(g_{1:t-1})$ ,

$$\begin{aligned} \text{Wealth}_t &= \text{Wealth}_{t-1} + w_t g_t = (1 + g_t \beta_t) \text{Wealth}_{t-1} \\ &\geq (1 + g_t \beta_t) F_{t-1}(g_{1:t-1}) \stackrel{\text{Def. 11(c)}}{\geq} F_t(g_{1:t-1}, g_t) = F_t(g_{1:t}) . \end{aligned}$$

**A.2. SA-Regret is stronger than m-shift regret**

A strongly-adaptive regret bound can be turned into an  $m$ -shift regret bound as follows. Let  $c > 0$ . We claim that:

$$\begin{aligned} & \left( \forall I = [I_1..I_2], R_I^A(\mathbf{w}) \leq c\sqrt{|I| \log(I_2)} \right) \\ & \implies m\text{-Shift-Regret}_T^A \leq c\sqrt{(m+1)T \log(T)} . \end{aligned}$$

To prove the claim, note that an  $m$ -shift sequence of experts  $\mathbf{w}_{1:T}$  can be partitioned into  $m+1$  contiguous blocks denoted by  $I^{(1)}, \dots, I^{(m+1)}$ . For example,  $(1, 1, 2, 2, 1)$  is 2-switch sequence whose partition  $\{[1, 2], [3, 4], [5]\}$ . Denote by  $\mathbf{w}_{I^{(k)}} \in \mathcal{W}$  the comparator in interval  $I^{(k)}$ :  $\mathbf{w}_t = \mathbf{w}_{I^{(k)}}, \forall t \in I^{(k)}$ . Then, using Cauchy-Schwartz inequality, we have

$$\begin{aligned} m\text{-Shift-Regret}_T^A &= \max_{\mathbf{w}_{1:T}: m\text{-shift seq.}} \sum_{k=1}^{m+1} R_{I^{(k)}}^A(\mathbf{w}_{I^{(k)}}) \\ &\leq \max_{\mathbf{w}_{1:T}: m\text{-shift seq.}} c\sqrt{\log T} \sum_{k=1}^{m+1} \sqrt{|I^{(k)}|} \\ &\leq \max_{\mathbf{w}_{1:T}: m\text{-shift seq.}} c\sqrt{\log T} \sqrt{m+1} \sqrt{\sum_{k=1}^{m+1} |I^{(k)}|} \\ &= c\sqrt{\log T} \sqrt{m+1} \sqrt{T} . \end{aligned} \tag{15}$$

**A.3. The data streaming intervals can replace the geometric covering intervals**

We show that the data streaming intervals achieves the same goal as the geometric covering intervals (GC). Let  $u(t)$  be a number such that  $2^{u(t)}$  is the largest power of 2 that divides  $t$ ; e.g.,  $u(12) = 2$ . The data streaming intervals (DS) are

$$\mathcal{J} = \{[t..(t + g \cdot 2^{u(t)} - 1)] : t = 1, 2, \dots\} . \tag{16}$$

For any interval  $J$ , we denote by  $J_1$  its starting time and by  $J_2$  its ending time. We say an interval  $J'$  is a prefix of  $J$  if  $J'_1 = J_1$  and  $J' \subseteq J$ .

We show that DS also partitions an interval  $I$  in Lemma 12.

**Lemma 12.** *Consider  $\mathcal{J}$  defined in (16) with  $g \geq 1$ . An interval  $[I_1..I_2] \subseteq [T]$  can be partitioned to a sequence of intervals  $\bar{J}^{(1)}, \bar{J}^{(2)}, \dots, \bar{J}^{(n)}$  such that*

1.  $\bar{J}^{(i)}$  is a prefix of some  $J \in \mathcal{J}$  for  $i = 1, \dots, n$ .
2.  $|\bar{J}^{(i+1)}|/|\bar{J}^{(i)}| \geq 2$  for  $i = 1, \dots, (n - 1)$ .

*Proof.* For simplicity, we assume  $g = 1$ ; we later explain how the analysis can be extended to  $g > 1$ . Let  $I_1 = 2^u \cdot k$  where  $2^u$  is the largest power of 2 that divides  $I_1$ . It follows that  $k$  is an odd number.

Let  $J \in \mathcal{J}$  be the data streaming interval that starts from  $I_1$ . The length  $|J|$  is  $2^u$  by the definition, and  $J_2$  is  $I_1 + 2^u - 1$ . Define  $\bar{J}^{(1)} := J$ .

Then, consider the next interval  $J' \in \mathcal{J}$  starting from time  $I_1 + 2^u$ . Note

$$J'_1 = I_1 + 2^u = 2^u \cdot k + 2^u = 2^{u+1} \cdot \frac{k + 1}{2}$$

Note that  $\frac{k+1}{2}$  is an integer since  $k$  is odd. Therefore,  $J'_1 = 2^{u'} \cdot k'$  where  $u' > u$ . It follows that the length of  $J'$  is

$$|J'| = 2^{u'} \geq 2 \cdot 2^u .$$

Then, define  $\bar{J}^{(2)} := J'$ .

We repeat this process until  $I$  is completely covered by  $\bar{J}^{(1)}, \dots, \bar{J}^{(n)}$  for some  $n$ . Finally, modify the last interval  $\bar{J}^{(n)}$  to end at  $I_2$  which is still a prefix of some  $J \in \mathcal{J}$ . This completes the proof for  $g = 1$ .

For the case of  $g > 1$ , note that by setting  $g > 1$  we are only making the intervals longer. Observe that even if  $g > 1$ , the sequence of intervals  $\bar{J}^{(1)}, \dots, \bar{J}^{(n)}$  above are still prefixes of some intervals in  $\mathcal{J}$ . □

Note that, unlike the partition induced by GC in which interval lengths successively double then successively halve, the partition induced by DS just successively doubles its interval lengths except the last interval. One can use DS to decompose SA-Regret of  $\mathcal{M}(\mathcal{B})$ ; that is, in (2), replace  $\sum_{i=-a}^b$  with  $\sum_{i=1}^n$  and  $J^{(i)}$  with  $\bar{J}^{(i)}$ . Since the decomposition by DS has the same effect of ‘doubling lengths’, one can show that Theorem 6 holds true with DS, too, with slightly smaller constant factors.

#### A.4. A subtle difference between the geometric covering and data streaming intervals

There is a subtle difference between the geometric covering intervals (GC) and the data streaming intervals (DS).

As far as the black-box algorithm has an anytime regret bound, both GC and DS can be used to prove the overall regret bound as in Theorem 6. In our experiments, the blackbox algorithm has anytime regret bound, so using DS does not break the theoretical guarantee.

However, there exist algorithms with fixed-budget regret bounds only. That is, the algorithm needs to know the target time horizon  $T^*$ , and the regret bound exists after exactly  $T^*$  time steps only. When these algorithms are used as the black-box, there is no easy way to prove Theorem 6 with DS intervals. The good news, still, is that most online learning algorithms are equipped with anytime regret bounds, and one can often use a technique called ‘doubling-trick’ [3, Section 2.3] to turn an algorithm with a fixed budget regret into the one with an anytime regret bound.

#### A.5. Technical results

**Lemma 13.** Assume  $\text{Regret}_T(\mathbf{u}) \leq \sqrt{\left(\xi + \sum_{i=1}^N u_i \bar{Z}_{T,i}\right) A(\mathbf{u})}$  for some function  $A$ . Then,  $\text{Regret}_T(\mathbf{u}) = \sqrt{(\xi + 2 \sum_{i=1}^N u_i \tilde{L}_{T,i}) A(\mathbf{u})} + A(\mathbf{u})$ .

*Proof.* We closely follow the proof of Luo & Schapire [14, Theorem 2]. We first claim that  $\sum_{i=1}^N u_i \bar{Z}_{t,i} \leq \text{Regret}_T(\mathbf{u}) + 2 \sum_{i=1}^N u_i \tilde{L}_{T,i}$ . The proof is as follows:

$$\begin{aligned}
 & \sum_{i=1}^N u_i \bar{Z}_{T,i} \\
 &= \sum_i u_i \sum_{t \in [T]} |r_{t,i}| \mathbf{1}\{w_{t,i} > 0\} + |[r_{t,i}]_+| \mathbf{1}\{w_{t,i} \leq 0\} \\
 &= \sum_i u_i \sum_t |r_{t,i}| \mathbf{1}\{w_{t,i} > 0\} + [r_{t,i}]_+ \mathbf{1}\{w_{t,i} \leq 0\} \\
 &= \sum_i u_i \sum_t r_{t,i} \mathbf{1}\{w_{t,i} > 0 \wedge r_{t,i} > 0\} + (-r_{t,i}) \mathbf{1}\{w_{t,i} > 0 \wedge r_{t,i} \leq 0\} \\
 & \quad + [r_{t,i}]_+ \mathbf{1}\{w_{t,i} \leq 0\} \\
 &= \sum_i u_i \sum_t r_{t,i} \mathbf{1}\{w_{t,i} > 0 \wedge r_{t,i} > 0\} + (-r_{t,i}) \mathbf{1}\{w_{t,i} > 0 \wedge r_{t,i} \leq 0\} \\
 & \quad + [r_{t,i}]_+ \mathbf{1}\{w_{t,i} \leq 0\} \\
 & \quad + r_{t,i} \mathbf{1}\{w_{t,i} > 0 \wedge r_{t,i} \leq 0\} - r_{t,i} \mathbf{1}\{w_{t,i} > 0 \wedge r_{t,i} \leq 0\} \\
 & \quad + [-r_{t,i}]_+ \mathbf{1}\{w_{t,i} \leq 0\} - [-r_{t,i}]_+ \mathbf{1}\{w_{t,i} \leq 0\} \\
 &= \sum_i u_i \sum_t r_{t,i}
 \end{aligned}$$

$$\begin{aligned}
& + \sum_i u_i \sum_t 2(-r_{t,i})\mathbf{1}\{w_{t,i} > 0 \wedge r_{t,i} \leq 0\} + [-r_{t,i}]_+\mathbf{1}\{w_{t,i} \leq 0\} \\
& = \text{Regret}_T(\mathbf{u}) + \sum_i u_i \sum_t 2[-r_{t,i}]_+\mathbf{1}\{w_{t,i} > 0\} + [-r_{t,i}]_+\mathbf{1}\{w_{t,i} \leq 0\} \\
& \leq \text{Regret}_T(\mathbf{u}) + 2 \sum_i u_i \tilde{L}_{T,i}
\end{aligned}$$

Let us simply use the notations  $R$  in place of  $\text{Regret}_T(\mathbf{u})$ ,  $A$  in place of  $A(\mathbf{u})$ , and  $\tilde{L}$  in place of  $\sum_i u_i \tilde{L}_{T,i}$ . It is safe to assume that  $R \geq 0$  since otherwise the statement of the Theorem is trivial. Then, by the assumption of the theorem,

$$\begin{aligned}
R & \leq \sqrt{(\xi + R + 2L)A} \\
\iff R^2 & \leq (\xi + R + 2L)A \\
\implies R & \leq \frac{1}{2}(A + \sqrt{A^2 + 4(\xi + 2L)A}) \\
& \leq \frac{1}{2}(A + A + 2\sqrt{(\xi + 2L)A}) \\
& = A + \sqrt{(\xi + 2L)A}. \quad \square
\end{aligned}$$

## Acknowledgements

This work was supported by NSF Award IIS-1447449 and NIH Award 1 U54 AI117924-01. The authors thank András György for providing constructive feedback and Kristjan Greenewald for providing the metric learning code.

## References

- [1] Adamskiy, Dmitry, Koolen, Wouter M, Chernov, Alexey, and Vovk, Vladimir. A Closer Look at Adaptive Regret. In *Proceedings of the International Conference on Algorithmic Learning Theory (ALT)*, pp. 290–304, 2012. [MR3042897](#)
- [2] Blum, Avrim. Empirical Support for Winnow and Weighted-Majority Algorithms: Results on a Calendar Scheduling Domain. *Machine Learning*, 26(1):5–23, 1997.
- [3] Cesa-Bianchi, Nicolo and Lugosi, Gabor. *Prediction, Learning, and Games*. Cambridge University Press, 2006. [MR2409394](#)
- [4] Cesa-Bianchi, Nicolo, Gaillard, Pierre, Lugosi, Gábor, and Stoltz, Gilles. Mirror descent meets fixed share (and feels no regret). In *Advances in Neural Information Processing Systems (NIPS)*, pp. 980–988, 2012.
- [5] Daniely, Amit, Gonen, Alon, and Shalev-Shwartz, Shai. Strongly Adaptive Online Learning. *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 1–18, 2015. [MR3277164](#)
- [6] Freund, Yoav, Schapire, Robert E, Singer, Yoram, and Warmuth, Manfred K. Using and combining predictors that specialize. *Proceedings of the*

- ACM symposium on Theory of computing (STOC)*, 37(3):334–343, 1997. [MR1715643](#)
- [7] Greenewald, Kristjan, Kelley, Stephen, and Hero, Alfred O. Dynamic metric learning from pairwise comparisons. *54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2016.
- [8] György, András, Linder, Tamás, and Lugosi, Gábor. Efficient tracking of large classes of experts. *IEEE Transactions on Information Theory*, 58(11): 6709–6725, 2012. [MR2991803](#)
- [9] Hazan, Elad and Seshadhri, Comandur. Adaptive Algorithms for Online Decision Problems. *IBM Research Report*, 10418:1–19, 2007.
- [10] Herbster, Mark and Warmuth, Manfred K. Tracking the Best Expert. *Mach. Learn.*, 32(2):151–178, 1998.
- [11] Jun, Kwang-Sung, Orabona, Francesco, Wright, Stephen, and Willett, Rebecca. Improved Strongly Adaptive Online Learning using Coin Betting. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 54, pp. 943–951, 2017.
- [12] Krichevsky, Raphael E and Trofimov, Victor K. The performance of universal encoding. *IEEE Trans. Information Theory*, 27(2):199–206, 1981. [MR0633417](#)
- [13] Kunapuli, Gautam and Shavlik, Jude. Mirror descent for metric learning: A unified approach. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Database (ECML/PKDD)*, pp. 859–874, 2012.
- [14] Luo, Haipeng and Schapire, Robert E. Achieving All with No Parameters: AdaNormalHedge. In *Proceedings of the Conference on Learning Theory (COLT)*, pp. 1286–1304, 2015.
- [15] Orabona, F., Crammer, K., and Cesa-Bianchi, N. A generalized online mirror descent with applications to classification and regression. *Machine Learning*, 99(3):411–435, 2015. [MR3347600](#)
- [16] Orabona, Francesco and Pál, David. Coin betting and parameter-free online learning. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 577–585. 2016.
- [17] Orabona, Francesco and Tommasi, Tatiana. Backprop without learning rates through coin betting. In *Advances in Neural Information Processing Systems (NIPS)*, 2017.
- [18] Shalev-Shwartz, Shai. *Online Learning: Theory, Algorithms, and Applications*. PhD thesis, Hebrew University, 2007.
- [19] Shalev-Shwartz, Shai. Online Learning and Online Convex Optimization. *Found. Trends Mach. Learn.*, 4(2):107–194, 2012.
- [20] Veness, Joel, White, Martha, Bowling, Michael, and György, András. Partition tree weighting. In *Proceedings of the 2013 Data Compression Conference*, pp. 321–330. IEEE Computer Society, 2013.
- [21] Zhang, Lijun, Yang, Tianbao, Jin, Rong, and Zhou, Zhi-Hua. Strongly adaptive regret implies optimally dynamic regret. *CoRR*, abs/1701.07570, 2017.