

# Sharp Dichotomies for Regret Minimization in Metric Spaces\*

Robert Kleinberg<sup>†</sup>

Aleksandrs Slivkins<sup>‡</sup>

## Abstract

The Lipschitz multi-armed bandit (MAB) problem generalizes the classical multi-armed bandit problem by assuming one is given side information consisting of a priori upper bounds on the difference in expected payoff between certain pairs of strategies. Classical results of Lai-Robbins [28] and Auer et al. [3] imply a logarithmic regret bound for the Lipschitz MAB problem on finite metric spaces. Recent results on continuum-armed bandit problems and their generalizations imply lower bounds of  $\sqrt{t}$ , or stronger, for many infinite metric spaces such as the unit interval. Is this dichotomy universal? We prove that the answer is yes: for every metric space, the optimal regret of a Lipschitz MAB algorithm is either bounded above by any  $f \in \omega(\log t)$ , or bounded below by any  $g \in o(\sqrt{t})$ . Perhaps surprisingly, this dichotomy does not coincide with the distinction between finite and infinite metric spaces; instead it depends on whether the completion of the metric space is compact and countable. Our proof connects upper and lower bound techniques in online learning with classical topological notions such as perfect sets and the Cantor-Bendixson theorem.

We also consider the *full-feedback* (a.k.a., *best-expert*) version of Lipschitz MAB problem, termed the *Lipschitz experts problem*, and show that this problem exhibits a similar dichotomy. We proceed to give nearly matching upper and lower bounds on regret in the Lipschitz experts problem on uncountable metric spaces. These bounds are of the form  $\Theta(t^\gamma)$ , where the exponent  $\gamma \in [\frac{1}{2}, 1]$  depends on the metric space. To characterize this dependence, we introduce a novel dimensionality notion tailored to the experts problem. Finally, we show that both Lipschitz bandits and Lipschitz experts problems become completely intractable (in the sense that no algorithm has regret  $o(t)$ ) if and only if the completion of the metric space is non-compact.

\*Full version of this paper is available at [www.arxiv.org](http://www.arxiv.org).

<sup>†</sup>Computer Science Department, Cornell University, Ithaca, NY 14853. Email: [rdk@cs.cornell.edu](mailto:rdk@cs.cornell.edu). Supported by NSF awards CCF-0643934 and IIS-0905467, an Air Force Office of Scientific Research grant, a Microsoft Research New Faculty Fellowship, and an Alfred P. Sloan Foundation Fellowship.

<sup>‡</sup>Microsoft Research, Mountain View, CA 94043. Email: [slivkins@microsoft.com](mailto:slivkins@microsoft.com).

## 1 Introduction

Multi-armed bandit (henceforth, MAB) problems have been studied for more than fifty years as a clean abstract setting for analyzing the exploration-exploitation tradeoffs that are common in sequential decision making. In the *stochastic MAB problem*, an algorithm must repeatedly choose from a fixed set of strategies (a.k.a. “arms”), each time receiving a random payoff whose distribution depends on the strategy selected.<sup>1</sup> The performance of MAB algorithms is commonly evaluated in terms of *regret*: the difference in expected payoff between the algorithm’s choices and always playing one fixed strategy. In addition to their many applications — which range from experimental design to online auctions and web advertising — another appealing feature of multi-armed bandit algorithms is that they are surprisingly efficient in terms of the growth rate of regret: for finite-armed bandit problems, algorithms whose regret at time  $t$  scales as  $O(\log t)$  have been known for more than two decades, beginning with the seminal work of Lai and Robbins [28] and extended in subsequent work such as [3].

Many of the applications of MAB problems — especially the computer science applications such as online auctions, web advertising, or adaptive routing — require considering strategy sets which are very large or even infinite. For infinite strategy sets the  $O(\log t)$  bound does not apply, while for very large finite sets the  $O(\cdot)$  notation masks a prohibitively large constant. Indeed, without making any assumptions about the strategies and their payoffs, bandit problems with large strategy sets allow for no non-trivial solutions — any MAB algorithm performs as badly, on some inputs, as random guessing. This motivates the study of bandit problems in which the strategy set is large but one is given *side information* constraining the form of the payoffs. Such problems have become the subject of quite intensive study in recent years, e.g. [7, 1, 4, 24, 25, 30, 6, 15, 13, 11, 5, 14, 23, 22, 21].

The *Lipschitz MAB problem* is a version of the stochastic MAB problem in which the side information consists of *a priori* upper bounds on the difference in

<sup>1</sup>More precisely, the payoff of each arm is an independent sample from a fixed distribution with bounded support.

expected payoff between certain pairs of strategies. This models situations where the decision maker has access to some similarity information about strategies which ensures that similar strategies obtain similar payoffs. Abstractly, the similarity information may be modeled as defining a *metric space* structure on the strategy set, and the side constraints imply that the expected payoff function  $\mu$  is a Lipschitz function (with Lipschitz constant 1) on this metric space.

The Lipschitz MAB problem was introduced by Kleinberg et al. [27].<sup>2</sup> Preceding work [1, 5, 11, 24, 31] has studied the problem in a few specific metric spaces such as a one-dimensional real interval. The prior work considered regret  $R(t)$  as a function of time  $t$ , and focused on the asymptotic dependence of  $R(t)$  on, loosely speaking, the dimensionality of the metric space. Various upper and lower bounds of the form  $R(t) = \Theta(t^\gamma)$  were proved, where the exponent  $\gamma < 1$  depends on the metric space. In particular, if the metric space is the interval  $[0, 1]$  with the standard metric  $d(x, y) = |x - y|$ , then there exists an algorithm with regret  $R(t) = \tilde{O}(t^{2/3})$ , and this bound is tight up to polylog factors [24]. More generally, for an arbitrary infinite metric space  $(X, d)$  one can define an isometry invariant  $\gamma = \gamma(X, d) \in [\frac{1}{2}, 1]$  such that there exists an algorithm with regret  $R(t) = \tilde{O}(t^\gamma)$ , which is tight up to polylog factors if  $\gamma > \frac{1}{2}$ ; see [27].

The following picture emerges. Although algorithms with regret  $R(t) = O(t^\gamma)$ ,  $\gamma < 1$  are known for most metric spaces, existing work unfortunately provides no examples of infinite metric spaces admitting bandit algorithms satisfying the Lai-Robbins regret bound  $R(t) = O(\log t)$ , although this bound holds for all finite metrics. In fact, for most metric spaces that have been studied (such as the unit interval) this possibility is *excluded* by known lower bounds of the form  $R(t) \notin o(t^\gamma)$ , where  $\gamma \geq \frac{1}{2}$ . Therefore we ask,

- Is  $\tilde{O}(\sqrt{t})$  regret the best possible for an infinite metric space? Alternatively, are there infinite metric spaces for which one can achieve regret  $O(\log t)$ ? Is there any metric space for which the best possible regret is *between*  $O(\log t)$  and  $\tilde{O}(\sqrt{t})$ ?

**Our contributions.** To make the above issue more concrete, we put forward the following definition.

**DEFINITION 1.1.** *Consider the Lipschitz MAB problem on a fixed metric space. A bandit algorithm is  $f(t)$ -tractable if for any problem instance  $\mathcal{I}$  the algorithm's*

*regret is  $R(t) = O_{\mathcal{I}}(f(t))$ .*<sup>3</sup> *The problem is  $f(t)$ -tractable if such an algorithm exists.*

We settle the questions listed above by proving the following dichotomy.

**THEOREM 1.1.** *Consider the Lipschitz MAB problem on a fixed metric space  $(X, d)$ . Then the following dichotomy holds: either the problem is  $f(t)$ -tractable for every  $f \in \omega(\log t)$ , or it is not  $g(t)$ -tractable for any  $g \in o(\sqrt{t})$ . In fact, the former occurs if and only if the completion of  $X$  is a compact metric space with countably many points.*

It is worth mentioning that the regret bound  $R(t) = O_{\mathcal{I}}(\log t)$  is the best possible, even for two-armed bandit problems, by a lower bound of Lai and Robbins [28]. Thus our upper bound for Lipschitz MAB problems in compact, countable metric spaces is nearly the best possible bound for such spaces, modulo the gap between “ $f(t) = \log t$ ” and “ $\forall f \in \omega(\log t)$ ”. Furthermore, we show that this gap is inevitable for infinite metric spaces:

**THEOREM 1.2.** *For every infinite metric space  $(X, d)$ , the Lipschitz MAB problem on  $(X, d)$  is not  $(\log t)$ -tractable.*

We turn our attention to the *full-feedback* version of the Lipschitz MAB problem. For any MAB problem there exists a corresponding *full-feedback* problem in which after each round, the payoffs from all strategies are revealed.<sup>4</sup> Such settings have been extensively studied in the online learning literature under the name *best experts problems* [9, 10, 32]. In particular, for a finite set of strategies one can achieve a *constant* regret [26] when payoffs are i.i.d. over time.

In addition to the full feedback, one could also consider a version in which the payoffs are revealed for some but not all strategies. Specifically, we define the *double feedback*, where in each round the algorithm selects two strategies: the “bet” for which it receives the payoff, and the “free peek”. After the round, the payoffs are revealed for both strategies. By abuse of notation, we will treat the bandit setting as a special case of the experts setting.

The experts version of the Lipschitz MAB problem, called the *Lipschitz experts problem*, is defined in the obvious way: a problem instance is specified by a triple  $(X, d, \mathbb{P})$ , where  $(X, d)$  is a metric space and  $\mathbb{P}$  is a Borel probability measure on the set  $[0, 1]^X$  of payoff functions on  $X$  (with the Borel  $\sigma$ -algebra induced by

<sup>2</sup>Megiddo and Hazan [22] consider a somewhat related (but technically very different) setting which combines full feedback, contextual “hints”, convex payoffs, and (essentially) a similarity metric space on the contexts.

<sup>3</sup>Notation  $O_{\mathcal{I}}(f(t))$  means  $C f(t)$  for some constant  $C = C(\mathcal{I})$ .

<sup>4</sup>Formally, an algorithm can query any finite number of arms.

the product topology on  $[0, 1]^X$  such that the expected payoff function  $x \mapsto E_{f \in \mathbb{P}}[f(x)]$  is a Lipschitz function on  $(X, d)$ . In each round an algorithm is presented with an i.i.d. sample from  $\mathbb{P}$ . The metric structure of  $(X, d)$  is known to the algorithm, the measure  $\mathbb{P}$  is not. We show that the Lipschitz experts problem exhibits a dichotomy similar to the one in Theorem 1.1. We formulate the upper bound for the double feedback, and the lower bound for the full feedback, thus avoiding the issue of what it means for an algorithm to receive feedback for infinitely many strategies.

**THEOREM 1.3.** *The Lipschitz experts problem on a fixed metric space  $(X, d)$  is either 1-tractable, even with double feedback, or it is not  $g(t)$ -tractable for any  $g \in o(\sqrt{t})$ , even with full feedback. The former occurs if and only if the completion of  $X$  is a compact metric space with countably many points.*

Theorems 1.1 and 1.3 assert a dichotomy between metric spaces on which the Lipschitz MAB/experts problem is very tractable, and those on which it is somewhat tractable. Let us consider the opposite end of the “tractability spectrum” and ask for which metric spaces the problem becomes completely intractable. We obtain a precise characterization: the problem is completely intractable if and only if the metric space is not pre-compact. Moreover, our upper bound is for the bandit setting, whereas the lower bound is for full feedback.

**THEOREM 1.4.** *The Lipschitz experts problem on a fixed metric space  $(X, d)$  is either  $f(t)$ -tractable for some  $f \in o(t)$ , even in the bandit setting, or it is not  $g(t)$ -tractable for any  $g \in o(t)$ , even with full feedback. The former occurs if and only if the completion of  $X$  is a compact metric space.*

Consider the full-feedback Lipschitz experts problem. In view of the  $\sqrt{t}$  lower bound from Theorems 1.3, we are interested in matching upper bounds. Gupta et al. [20] observed that such bounds hold for every metric space  $(X, d)$  of finite covering dimension: namely, the Lipschitz experts problem on  $(X, d)$  is  $\sqrt{t}$ -tractable. (Their algorithm is a version of the “naive algorithm” from [24, 27].) Therefore it is natural to ask whether there exist metric spaces for which the optimal regret in the Lipschitz experts problem is *between*  $\sqrt{t}$  and  $t$ . We settle this question by proving a characterization with nearly matching upper and lower bounds in terms of a novel dimensionality notion tailored to the experts problem.

**THEOREM 1.5.** *For any metric space  $(X, d)$ , there exist an isometry invariant  $b = b(X, d)$  such that the full-feedback Lipschitz experts problem on  $(X, d)$  is  $(t^\gamma)$ -tractable for any  $\gamma > \frac{b+1}{b+2}$ , and not  $(t^\gamma)$ -tractable for any  $\gamma < \frac{b-1}{b}$ . Depending on the metric space,  $b(X, d)$  can take any value on a dense subset of  $[0, \infty)$ .*

The lower bound in Theorem 1.5 holds for a restricted version of full-feedback Lipschitz experts problem in which a problem instance  $(X, d, \mathbb{P})$  satisfies a further property that each function  $f \in \text{support}(\mathbb{P})$  is itself a Lipschitz function on  $(X, d)$ . We term this version the *uniformly Lipschitz experts problem* (with full feedback). In fact, for this version we obtain a matching upper bound.

**THEOREM 1.6.** *Consider the uniformly Lipschitz experts problem with full feedback. Fix an uncountable metric space  $(X, d)$ . Let  $b = b(X, d)$  the isometry invariant from Theorem 1.5. Then the problem on  $(X, d)$  is  $(t^\gamma)$ -tractable for any  $\gamma > \max(\frac{b-1}{b}, \frac{1}{2})$ , and not  $(t^\gamma)$ -tractable for any  $\gamma < \max(\frac{b-1}{b}, \frac{1}{2})$ .*

**Connection to point-set topology.** The main technical contribution of this paper is an interplay of online learning and point-set topology, which requires novel algorithmic and lower-bounding techniques. In particular, the connection to topology is essential in the (joint) proof of the two main results (Theorem 1.1 and Theorem 1.3). There, we identify a simple topological property (*well-orderability*) which entails the algorithmic result, and another topological property (*perfectness*) which entails the lower bound.

**DEFINITION 1.2.** *Consider a topological space  $X$ .  $X$  is called perfect if it contains no isolated points. A topological well-ordering of  $X$  is a well-ordering  $(X, \prec)$  such that every initial segment thereof is an open set. If such  $\prec$  exists,  $X$  is called well-orderable. A metric space  $(X, d)$  is called well-orderable if and only if its metric topology is well-orderable.*

Perfect spaces are a classical notion in point-set topology. Topological well-orderings are implicit in the work of Cantor [8], but the particular definition given here is new, to the best of our knowledge.

The proof of Theorems 1.1 and 1.3 (for compact metric spaces) consists of three parts: the algorithmic result for a compact well-orderable metric space, the lower bound for a metric space with a perfect subspace, and the following lemma that ties together the two topological properties.

LEMMA 1.1. *For any compact metric space  $(X, d)$ , the following are equivalent: (i)  $X$  is countable, (ii)  $(X, d)$  is well-orderable, (iii) no subspace of  $(X, d)$  is perfect.<sup>5</sup>*

Lemma 1.1 follows from classical theorems of Cantor-Bendixson [8] and Mazurkiewicz-Sierpinski [29]. We provide a proof in the full version of the paper.

To reduce the Lipschitz MAB problem to complete metric spaces we show that the problem is  $f(t)$ -tractable on a given metric space if and only if it is  $f(t)$ -tractable on the completion thereof. Same is true for the double-feedback Lipschitz experts problem, and the “only if” direction holds for the full-feedback Lipschitz experts problem. Then the main dichotomy results follow from the lower bound in Theorem 1.4.

**Accessing the metric space.** We define a bandit algorithm as a (possibly randomized) Borel measurable function that maps a history of past observations  $(x_i, r_i) \in X \times [0, 1]$  to a strategy  $x \in X$  to be played in the current period. An experts algorithm is similarly defined as a (possibly randomized) Borel measurable function mapping the observation history to a strategy  $x \in X$  to be played in the current period. (Or, in the case of the double feedback model, a pair of strategies representing the “bet” and “free peek”.) The observation history is either a sequence of elements of  $[0, 1]^X$  in the full feedback model, or a sequence of quadruples  $(x_i, r_i, x'_i, r'_i) \in (X \times [0, 1])^2$  in the double feedback model.

These definitions abstract away a potentially thorny issue of representing and accessing an infinite metric space. For our algorithmic results, we handle this issue as follows: the metric space is accessed via well-defined calls to a suitable *oracle*. Moreover, the main algorithmic result in Theorems 1.1 and 1.3 requires an oracle which represents the well-ordering. We also provide an extension in Section 6: an  $\omega(\log t)$ -tractability result for a wide family of metric spaces – including, for example, compact metric spaces with a finite number of limit points – for which a more intuitive oracle access suffices. These are the metric spaces with a finite *Cantor-Bendixson rank*, a classic notion from point-set topology.

**Related work and discussion.** Algorithms for the stochastic MAB problem admit regret guarantees of the form  $R(t) = O(f(t))$ , which are of two types depending on whether the constant in  $O()$  is allowed to depend on the problem instance. For instance, UCB1 [3] admits an instance-specific guarantee  $R(t) = O(\log t)$ , whereas the best-known instance-independent guarantee for this algorithm is only  $R(t) = O(\sqrt{kt \log t})$ , where  $k$  is the number of arms. Accordingly, a lower

bound for the instance-independent version has to show that for any algorithm and a given time  $t$ , there exists a problem instance whose regret is large at this time, whereas for the instance-specific version one needs a much more ambitious argument: for any algorithm there exists a problem instance whose regret is large *infinitely often*. In this paper, we focus on instance-specific guarantees.

Apart from the stochastic MAB problem considered in this paper, several other MAB formulations have been studied in the literature (see [10] for a survey). Early work [17, 16] has focused on Bayesian formulations in which Bayesian priors on payoffs are known, and goal is to maximize the payoff in expectation over these priors. In these formulations, an MAB instance is a *Markov Decision Process* (MDP) in which each arm is represented by a Markov Chain with rewards on states, and the transition happens whenever the arm is played. In the more “difficult” *restless bandits* [33] formulations, the state also changes when the arm is passive, according to another transition matrix. In the theoretical computer science literature, recent work in this vein includes [18, 19]. Interestingly, these Bayesian formulations have an offline flavor: given the MDP, one needs to efficiently compute a (nearly) optimal mapping from states to actions. Contrasting the Bayesian formulations in which the probabilistic model is fully specified, the *adversarial MAB problem* [4, 2, 21] makes no stochastic assumptions whatsoever. Instead, it makes a very pessimistic assumption that payoffs are chosen by an adversary that has access to the algorithm’s code but not to its random seed. As in the stochastic MAB problem, the goal is to minimize regret. For any fixed (finite) number of arms, the best possible regret in this setting is  $R(t) = O(\sqrt{t})$  [4]. For infinite strategy sets, one often considers the *linear MAB problem* in which strategies lie in a convex subset of  $\mathbb{R}^d$ , and in each round the payoffs form a linear function, e.g. [30, 6, 14, 21].

It is an open question whether the ideas from the Lipschitz MAB problem extend to the above formulations. The adversarial version of the Lipschitz MAB problem is well-defined, but to the best of our knowledge, the only known result is the “naive” algorithm from [24]. One could define the stochastic version of the linear MAB problem (in which the expected payoffs form a fixed time-invariant linear function), which can be viewed as a special case of the Lipschitz MAB problem. However, this view is not likely to be fruitful because in the Lipschitz MAB problem measuring a payoff of one arm is useless for estimating the payoffs of distant arms, whereas in prior work on the linear MAB problem inferences about distant arms are crucial. For Bayesian MAB problems with limited similarity information, it

<sup>5</sup>Absent compactness, (ii)  $\iff$  (iii) and (i)  $\implies$  (ii), but (ii)  $\not\implies$  (i).

is not clear how to model this information, mainly because in the Bayesian setting similarity between arms is naturally represented via correlated priors rather than a metric space.

**Organization of the paper.** Preliminaries are in Section 2. We present a joint proof for the two main results (Theorems 1.1 and 1.3). The lower bound is proved in Section 3 and the algorithmic results are in Section 4. Coupled with the topological equivalence (Lemma 1.1), this gives the proof for compact metric spaces. A complementary (log  $t$ )-intractability result for infinite metric spaces (Theorem 1.2) is in Section 5. The  $\omega(\log t)$ -tractability result via simpler oracle access (for metric spaces of finite Cantor-Bendixson rank) is in Section 6. The boundary-of-tractability result (Theorems 1.4) is in Section 7. The full-feedback Lipschitz experts problem in a (very) high dimension (including Theorems 1.5 and 1.6) is discussed in Sections 8 and 9.

Some of the proofs are moved to appendices. In Appendix A we reduce the problem to that on complete metric spaces. All KL-divergence arguments (which underlie our lower bounds) are gathered in Appendix B.

## 2 Preliminaries.

This section contains various definitions which make the paper essentially self-contained (the only exception being *ordinal numbers* which are used in Section 9.2). In particular, the paper uses notions from General Topology which are typically covered in any introductory text or course on the subject.

**Lipschitz MAB problem.** Consider the Lipschitz MAB problem on a metric space  $(X, d)$  with payoff function  $\mu$ . The payoff from each arm  $x \in X$  is an independent sample from a fixed (time-invariant) distribution with support in  $[0, 1]$  and expectation  $\mu(x)$  such that  $|\mu(x) - \mu(y)| \leq d(x, y)$  for all  $x, y \in X$ . For  $S \subset X$  denote  $\sup(\mu, S) = \sup_{x \in S} \mu(x)$  and similarly  $\operatorname{argmax}(\mu, S) = \operatorname{argmax}_{x \in S} \mu(x)$ . Given a bandit algorithm  $\mathcal{A}$ , let  $P_{(\mathcal{A}, \mu)}(t)$  be the expected reward collected by the algorithm in the first  $t$  rounds on the problem instance  $(X, d, \mu)$ . The *regret* of algorithm  $\mathcal{A}$  in  $t$  rounds is  $R_{(\mathcal{A}, \mu)}(t) = \sup(\mu, X)t - P_{(\mathcal{A}, \mu)}(t)$ . Given a Lipschitz experts algorithm  $\mathcal{A}$  and a problem instance  $(X, d, \mathbb{P})$ , the notations  $P_{(\mathcal{A}, \mathbb{P})}(t)$  and  $R_{(\mathcal{A}, \mathbb{P})}(t)$  — denoting expected reward and regret — are defined analogously.

**Metric topology.** Let  $(X, d)$  be a metric space. An open ball in  $(X, d)$  is denoted  $B(x_0, r) = \{x \in X : d(x, x_0) < r\}$ , where  $x_0 \in X$  is the center, and  $r \geq 0$  is the radius. A *Cauchy sequence* in  $(X, d)$  is a sequence such that for every  $\delta > 0$ , there is an open ball of radius  $\delta$  containing all but finitely many points of the sequence. We say  $X$  is *complete* if every Cauchy sequence has a limit point in  $X$ . For two Cauchy

sequences  $\mathbf{x} = x_1, x_2, \dots$  and  $\mathbf{y} = y_1, y_2, \dots$  the *distance*  $d(\mathbf{x}, \mathbf{y}) = \lim_{i \rightarrow \infty} d(x_i, y_i)$  is well-defined. Two Cauchy sequences are declared to be equivalent if their distance is 0. The equivalence classes of Cauchy sequences form a metric space  $(X^*, d)$  called the *completion* of  $(X, d)$ . The subspace of all constant sequences is identified with  $(X, d)$ : formally, it is a dense subspace of  $(X^*, d)$  which is isometric to  $(X, d)$ . A metric space  $(X, d)$  is *compact* if every collection of open balls covering  $(X, d)$  has a finite subcollection that also covers  $(X, d)$ . Every compact metric space is complete, but not vice-versa.

Let  $X$  be a set. A family  $\mathcal{F}$  of subsets of  $X$  is called a *topology* if it contains  $\emptyset$  and  $X$  and is closed under arbitrary unions and finite intersections. When a specific topology is fixed and clear from the context, the elements of  $\mathcal{F}$  are called *open sets*, and their complements are called *closed sets*. Throughout this paper, these terms will refer to the *metric topology* of the underlying metric space, the smallest topology that contains all open balls (namely, the intersection of all such topologies). A point  $x$  is called *isolated* if the singleton set  $\{x\}$  is open. A function between topological spaces is *continuous* if the inverse image of every open set is open.

**Set theory.** Let  $S$  be a set. A *well-ordering* on a set  $S$  is a total order on  $S$  with the property that every non-empty subset of  $S$  has a least element in this order. Each set can be well-ordered. (More precisely, this statement is equivalent to the Axiom of Choice.)

In Section 9.2 use *ordinals*, a.k.a. *ordinal numbers*, are a classical concept in set theory that, in some sense, extend natural numbers beyond infinity. Understanding this paper requires only the basic notions about ordinals, namely the standard (von Neumann) definition of ordinals, successor and limit ordinals, and transfinite induction. The necessary material can be found in any introductory text on Mathematical Logic and Set Theory, and also on *Wikipedia*.

## 3 Lower bounds via a perfect subspace

In this section we prove the following lower bound:

**THEOREM 3.1.** *Consider the Lipschitz experts problem on a metric space  $(X, d)$  which has a perfect subspace. Then the problem is not  $g$ -tractable for any  $g \in o(\sqrt{t})$ . In fact, a much stronger result holds: there exist a distribution  $\mathcal{P}$  over problem instances  $\mu$  such that for any experts algorithm  $\mathcal{A}$  we have*

$$(3.1) \quad (\forall g \in o(\sqrt{t})) \quad \Pr_{\mu \in \mathcal{P}} [R_{(\mathcal{A}, \mu)}(t) = O_\mu(g(t))] = 0.$$

Let us construct the desired distribution over problem instances. First, we use the existence of a perfect subspace to construct a useful system of balls.

DEFINITION 3.1. A ball-tree on a metric space  $(X, d)$  is a complete infinite binary tree whose nodes are pairs  $(x, r)$ , where  $x \in X$  is the “center” and  $r \in (0, 1]$  is the “radius”, such that:

- if  $(x, r)$  is a parent of  $(x', r')$  then  $d(x, x') + r' < \frac{r}{2}$ ,
- for siblings  $(x, r_x)$  and  $(y, r_y)$ ,  $r_x + r_y < d(x, y)$ .

In a ball-tree, each tree node  $(x, r)$  corresponds to a ball  $B(x, r)$  so that each child is a subset of its parent and any two siblings are disjoint.<sup>6</sup>

LEMMA 3.1. For any metric space with a perfect subspace there exists a ball-tree.

*Proof.* Consider a metric space  $(X, d)$  with a perfect subspace  $(Y, d)$ . Let us construct the ball-tree recursively, maintaining the invariant that for each tree node  $(y, r)$  we have  $y \in Y$ . Pick an arbitrary  $y \in Y$  and let the root be  $(y, 1)$ . Suppose we have constructed a tree node  $(y, r)$ ,  $y \in Y$ . Since  $Y$  is perfect, the ball  $B(y, r/4)$  contains another point  $y' \in Y$ . Let  $r' = d(y, y')/2$  and define the two children of  $(y, r)$  as  $(y, r')$  and  $B(y', r')$ .

Now let us use the ball-tree to construct the distribution on payoff functions. Consider a metric space  $(X, d)$  with a fixed ball-tree  $T$ . For each  $i \geq 1$ , let  $D_i$  be the set of all depth- $i$  tree nodes, and let  $r_i^* = \min\{r : (x, r) \in D_i\}$  be the smallest radius among these nodes. Note that  $r_i^* \leq 2^{-i}$ . Choose a number  $n_i$  large enough that  $g(n) < \frac{1}{8i} r_i^* \sqrt{n}$  for all  $n > n_i$ , and let  $\delta_i = n_i^{-1/2}$ . For each tree node  $w = (x_0, r_0)$  define a function  $F_w : X \rightarrow [0, 1]$  by

$$(3.2) \quad F_w(x) = \begin{cases} \min\{r_0 - d(x, x_0), r_0/2\} & \text{if } x \in B(x_0, r_0), \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that  $F_w$  is a Lipschitz function on  $(X, d)$ . A leaf in a ball-tree is an infinite path from the root:  $\mathbf{w} = (w_0, w_1, w_2, \dots)$ , where  $w \in D_i$  for all  $i$ . A lineage in a ball-tree is a set of tree nodes containing at most one child of each node; if it contains exactly one child of each node then we call it a complete lineage. For each complete lineage  $\lambda$  there is an associated leaf  $\mathbf{w}(\lambda)$  defined by  $\mathbf{w} = (w_0, w_1, \dots)$  where  $w_0$  is the root and for  $i > 0$ ,  $w_i$  is the unique child of  $w_{i-1}$  that belongs to  $\lambda$ . Let us use a lineage in the ball-tree to define a probability measure  $\mathbb{P}_\lambda$  on payoff functions via the following sampling rule.

<sup>6</sup>Defining internal nodes as balls rather than  $(x, r)$  pairs could lead to confusion later in the construction because a ball in a metric space is a set of points, and as such does not necessarily have a unique center or radius.

First every node  $w$  independently samples a random sign  $\text{sign}(w) \in \{+1, -1\}$ , assigning probability  $(1 + \delta_i)/2$  to  $+1$  if  $w \in \lambda \cap D_i$ , and choosing the sign uniformly at random otherwise. Now define a payoff function  $\pi$  associated with this sign pattern as follows:  $\pi = \frac{1}{2} + \sum_{w \in T \setminus D_0} \text{sign}(w) F_w$ . By construction,  $\pi$  is a Lipschitz function taking values in  $[0, 1]$ . Let  $\mathcal{P}_T$  be the distribution over problem instances  $\mathbb{P}_\lambda$  in which  $\lambda$  is a complete lineage sampled uniformly at random; that is, each node samples one of its children independently and uniformly at random, and  $\lambda$  is the set of sampled children. This completes our construction.

*Remark.* Let  $\lambda$  be a complete lineage in the ball-tree, and let  $\mathbf{w}(\lambda) = (w_0, w_1, w_2, \dots)$  be its associated leaf, where  $w_i = (x_i, r_i)$  for all  $i$ . Suppose for some  $i$  we have  $x \in B(x_i, r_i/2)$  and  $x \notin B(x_{i+1}, r_{i+1})$ . Then the expected payoff function  $\mu_\lambda = \mathbf{E}[\pi]$  associated to the measure  $\mathbb{P}_\lambda$  satisfies  $\mu_\lambda(x) = \frac{1}{2} + \sum_{j=1}^i r_j^* \delta_j / 4$ .

LEMMA 3.2. Consider a metric space  $(X, d)$  with a ball-tree  $T$ . Then (3.1) holds with  $\mathcal{P} = \mathcal{P}_T$ .

To prove this lemma, we define a notion called an  $(\epsilon, \delta, k)$ -ensemble, which is a collection of  $k$  payoff distributions that are nearly indistinguishable from the standpoint of an online learning algorithm. To this end, we consider a more general setting than the one in the Lipschitz experts problem. In the feasible experts problem, one is given a set  $X$  (not necessarily a metric space) along with a collection  $\mathcal{D}$  of Borel probability measures on the set  $[0, 1]^X$  of functions  $\pi : X \rightarrow [0, 1]$ . A problem instance of the feasible experts problem consists of a triple  $(X, \mathcal{D}, \mathbb{P})$  where  $X$  and  $\mathcal{D}$  are known to the algorithm, and  $\mathbb{P} \in \mathcal{D}$  is not.

DEFINITION 3.2. Consider a set  $X$  and a  $(k + 1)$ -tuple  $\vec{\mathbb{P}} = (\mathbb{P}_0, \mathbb{P}_1, \dots, \mathbb{P}_k)$  of Borel probability measures on  $[0, 1]^X$ , the set of  $[0, 1]$ -valued payoff functions  $\pi$  on  $X$ . For  $0 \leq i \leq k$  and  $x \in X$ , let  $\mu_i(x)$  denote the expectation of  $\pi(x)$  under measure  $\mathbb{P}_i$ . We say that  $\vec{\mathbb{P}}$  is an  $(\epsilon, \delta, k)$ -ensemble if there exist pairwise disjoint subsets  $S_1, S_2, \dots, S_k \subseteq X$  for which the following properties hold:

1. for every  $i$  and every event  $\mathcal{E}$  in the Borel  $\sigma$ -algebra of  $[0, 1]^X$ , we have  $1 - \delta < \mathbb{P}_0(\mathcal{E})/\mathbb{P}_i(\mathcal{E}) < 1 + \delta$ ,
2. for every  $i > 0$ , we have

$$\sup(\mu_i, S_i) - \sup(\mu_i, X \setminus S_i) \geq \epsilon.$$

THEOREM 3.2. Consider the feasible experts problem on  $(X, \mathcal{D})$ . Let  $\vec{\mathbb{P}}$  be an  $(\epsilon, \delta, k)$ -ensemble with  $\{\mathbb{P}_1, \dots, \mathbb{P}_k\} \subseteq \mathcal{D}$  and  $0 < \epsilon, \delta < 1/2$ . Then for

any  $t < \ln(17k)/(2\delta^2)$  and any experts algorithm  $\mathcal{A}$ , at least half of the measures  $\mathbb{P}_i$  have the property that  $R_{(\mathcal{A}, \mathbb{P}_i)}(t) \geq \epsilon t/2$ .

*Remarks.* For space reasons, the proof of this theorem has been moved to the appendix. The proof of Theorem 3.1 uses Theorem 3.2 for  $k = 2$ , and the proof of Theorem 1.5 will use it again for large  $k$ .

*Proof of Lemma 3.2:* Consider the ball-tree  $T$ . For each  $i \geq 1$ , recall that  $D_i$  is the set of all depth- $i$  nodes in  $T$ , and that  $r_i^* = \min\{r : (x, r) \in D_i\}$  is the smallest radius among these nodes. Let  $\mathbf{P}$  be the set of all probability measures induced by the lineages of  $T$ . For each complete lineage  $\lambda$  and tree node  $w$  in  $T$ , let  $w_1, w_2$  denote the children of  $w$ , let  $i$  denote their depth, and let  $w'$  denote the unique element of  $\{w_1, w_2\} \cap \lambda$ . The three lineages  $\lambda_0 = \lambda \setminus \{w'\}$ ,  $\lambda_1 = \lambda_0 \cup \{w_1\}$ ,  $\lambda_2 = \lambda_0 \cup \{w_2\}$  define a triple of probability measures  $\vec{\mathbb{P}} = (\mathbb{P}_{\lambda_0}, \mathbb{P}_{\lambda_1}, \mathbb{P}_{\lambda_2})$  that constitute a  $(\epsilon, \delta_i, 2)$ -ensemble where  $\epsilon = r_i^* \delta_i/4$ .

Let us fix an experts algorithm  $\mathcal{A}$ . By Theorem 3.2 there exists  $\alpha(w) \in \{\mathbb{P}_{\lambda_1}, \mathbb{P}_{\lambda_2}\}$  such that for any  $t_i$  satisfying  $1/\delta_i^2 < t < \ln(34)/(2\delta_i^2)$ ,

$$R_{(\mathcal{A}, \alpha(w))}(t_i) \geq \epsilon t_i/2 = r_i^* \delta_i t_i/8 > \frac{1}{8} r_i^* \sqrt{t},$$

Recalling the definition of  $n_i = \delta_i^{-2}$ , we see that  $i \cdot g(t_i) < \frac{1}{8} r_i^* \sqrt{t_i} < R_{(\mathcal{A}, \alpha(w))}(t_i)$ .

For each  $i$ , let us define  $\mathcal{E}_i$  to be the set of input distributions  $\mathbb{P}_\lambda$  such that  $\lambda$  is a complete lineage whose associated leaf  $\mathbf{w}(\lambda) = (w_0, w_1, \dots)$  satisfies  $w_i = \alpha(w_{i-1})$ . Interpreting these sets as random events under the probability distribution  $\mathcal{P}_T$ , we have proved the following: there exists a sequence of events  $\mathcal{E}_i$ ,  $i \in \mathbb{N}$  and a sequence of times  $t_i \rightarrow \infty$  such that for each  $i$  we have (i)  $\Pr[\mathcal{E}_i | \sigma(\mathcal{E}_1, \dots, \mathcal{E}_{i-1})] = \frac{1}{2}$  and (ii)  $R_{(\mathcal{A}, \mathbb{P})}(t_i) > i \cdot g(t_i)$  for any  $\mathbb{P} \in \mathcal{E}_i$ .

Now, let us fix an experts algorithm  $\mathcal{A}$ . For each complete lineage  $\lambda$ , define

$$C_\lambda := \inf\{C \leq \infty : R_{(\mathcal{A}, \mathbb{P}_\lambda)}(t) \leq C g(t) \text{ for all } t\}.$$

Note that  $R_{(\mathcal{A}, \mathbb{P}_\lambda)}(t) = O_\mu(g(t))$  if and only if  $C_\lambda < \infty$ . We claim that  $\Pr[C_\lambda < \infty] = 0$  where the probability is over the random choice of complete lineage  $\lambda$ . Indeed, if infinitely many events  $\mathcal{E}_i$  happen, then event  $\{C_\mu < C\}$  does not. But the probability that infinitely many events  $\mathcal{E}_i$  happen is 1, because for every positive integer  $n$ ,  $\Pr[\bigcap_{i=n}^\infty \mathcal{E}_i] = \prod_{i=n}^\infty \Pr[\mathcal{E}_i | \bigcap_{j=n}^{i-1} \mathcal{E}_j] = 0$ .

#### 4 Tractability via well-orderable metric spaces

In this section we prove the main algorithmic result.

**THEOREM 4.1.** *Consider a compact well-orderable metric space  $(X, d)$ . Then:*

- (a) *the Lipschitz MAB problem on  $(X, d)$  is  $f$ -tractable for every  $f \in \omega(\log t)$ ;*
- (b) *the Lipschitz experts problem on  $(X, d)$  is 1-tractable, even with a double feedback.*

We present a joint exposition for both the bandit and the experts version. Let us consider the Lipschitz MAB/experts problem on a compact metric space  $(X, d)$  with a topological well-ordering  $\prec$  and a payoff function  $\mu$ . For each strategy  $x \in X$ , let  $S(x) = \{y \preceq x : y \in X\}$  be the corresponding initial segment of the well-ordering  $(X, \prec)$ . Let  $\mu^* = \sup(\mu, X)$  denote the maximal payoff. Call a strategy  $x \in X$  *optimal* if  $\mu(x) = \mu^*$ . We rely on the following structural lemma:

**LEMMA 4.1.** *There exists an optimal strategy  $x^* \in X$  such that  $\sup(\mu, X \setminus S(x^*)) < \mu^*$ .*

*Proof.* Let  $X^*$  be the set of all optimal strategies. Since  $\mu$  is a continuous real-valued function on a compact space  $X$ , it attains its maximum, i.e.  $X^*$  is non-empty, and furthermore  $X^*$  is closed. Note that  $\{S(x) : x \in X^*\}$  is an open cover for  $X^*$ . Since  $X^*$  is compact (as a closed subset of a compact set) this cover contains a finite subcover, call it  $\{S(x) : x \in Y^*\}$ . Then the  $\prec$ -maximal element of  $Y^*$  is the  $\prec$ -maximal element of  $X^*$ . The initial segment  $S(x^*)$  is open, so its complement  $Y = X \setminus S(x^*)$  is closed and therefore compact. It follows that  $\mu$  attains its maximum on  $Y$ , say at a point  $y^* \in Y$ . By the choice of  $y^*$  we have  $x^* \prec y^*$ , so by the choice of  $x^*$  we have  $\mu(x^*) > \mu(y^*)$ .

In the rest of this section we let  $x^*$  be the strategy from Lemma 4.1. Our algorithm is geared towards finding  $x^*$  eventually, and playing it from then on. The idea is that if we cover  $X$  with balls of a sufficiently small radius, any strategy in a ball containing  $x^*$  has a significantly larger payoff than any strategy in a ball that overlaps with  $X \setminus S(x^*)$ .

The algorithm accesses the metric space and the well-ordering via the following two oracles.

**DEFINITION 4.1.** *A  $\delta$ -covering of a metric space  $(X, d)$  is a subset  $S \subset X$  such that each point in  $X$  lies within distance  $\delta$  from some point in  $S$ . An oracle  $\mathcal{O} = \mathcal{O}(k)$  is a covering oracle for  $(X, d)$  if it inputs  $k \in \mathbb{N}$  and outputs a pair  $(\delta, S)$  where  $\delta = \delta_{\mathcal{O}}(k)$  is a positive number and  $S$  is a  $\delta$ -covering of  $X$  consisting of at most  $k$  points. Here  $\delta_{\mathcal{O}}(\cdot)$  is any function such that  $\delta_{\mathcal{O}}(k) \rightarrow 0$  as  $k \rightarrow \infty$ .*

**DEFINITION 4.2.** *Given a metric space  $(X, d)$  and a total order  $(X, \prec)$ , the ordering oracle inputs a finite collection of balls (given by the centers and the radii),*

and returns the  $\prec$ -maximal element covered by the closure of these balls, if such element exists, and an arbitrary point in  $X$  otherwise.

Our algorithm is based on the following *exploration subroutine* `EXPL()`.

**ALGORITHM 4.1.** Subroutine `EXPL( $k, n, r$ )`: inputs  $k, n \in \mathbb{N}$  and  $r \in (0, 1)$ , outputs a point in  $X$ .

First it calls the covering oracle  $\mathcal{O}(k)$  and receives a  $\delta$ -covering  $S$  of  $X$  consisting of at most  $k$  points. Then it plays each strategy  $x \in S$  exactly  $n$  times; let  $\mu_{\text{av}}(x)$  be the sample average. Let us say that  $x$  a *loser* if  $\mu_{\text{av}}(y) - \mu_{\text{av}}(x) > 2r + \delta$  for some  $y \in S$ . Finally, it calls the ordering oracle with the collection of all closed balls  $\bar{B}(x, \delta)$  such that  $x$  is not a loser, and outputs the point  $x_{\text{or}} \in X$  returned by this oracle call.

Clearly, `EXPL( $k, n, r$ )` takes at most  $kn$  rounds to complete. We show that for sufficiently large  $k, n$  and sufficiently small  $r$  it returns  $x^*$  with high probability.

**LEMMA 4.2.** Fix a problem instance and let  $x^*$  be the optimal strategy from Lemma 4.1. Consider increasing functions  $k, n, T : \mathbb{N} \rightarrow \mathbb{N}$  such that  $r(t) := 4\sqrt{(\log T(t))/n(t)} \rightarrow 0$ . Then for any sufficiently large  $t$ , with probability at least  $1 - T^{-2}(t)$ , the subroutine `EXPL( $k(t), n(t), r(t)$ )` returns  $x^*$ .

*Proof.* Let us use the notation from Algorithm 4.1. Fix  $t$  and consider a run of `EXPL( $k(t), n(t), r(t)$ )`. Call this run *clean* if for each  $x \in S$  we have  $|\mu_{\text{av}}(x) - \mu(x)| \leq r(t)$ . By Chernoff Bounds, this happens with probability at least  $1 - T^{-2}(t)$ . In the rest of the proof, let us assume that the run is clean.

Let  $\bar{B}$  be the union of the closed balls  $\bar{B}(x, \delta)$ ,  $x \in S^*$ . Then the ordering oracle returns the  $\prec$ -maximal point in  $\bar{B}$  if such point exists. We will show that  $x^* \in \bar{B} \subset S(x^*)$  for any sufficiently large  $t$ , which will imply the lemma.

We claim that  $x^* \in \bar{B}$ . Since  $S$  is a  $\delta$ -covering, there exists  $y^* \in S$  such that  $d(x^*, y^*) \leq \delta$ . Let us fix one such  $y^*$ . It suffices to prove that  $y^*$  is not a loser. Indeed, if  $\mu_{\text{av}}(y) - \mu_{\text{av}}(y^*) > 2r(t) + \delta$  for some  $y \in S$  then  $\mu(y) > \mu(y^*) + \delta \geq \mu^*$ , contradiction. Claim proved.

Let  $\mu_0 = \sup(\mu, X \setminus S(x^*))$  and let  $r_0 = (\mu^* - \mu_0)/7$ . Let us assume that  $t$  is sufficiently large so that  $r(t) < r_0$  and  $\delta = \delta_{\mathcal{O}}(k(t)) < r_0$ , where  $\delta_{\mathcal{O}}(\cdot)$  is from the definition of the covering oracle.

We claim that  $\bar{B} \subset S(x^*)$ . Indeed, consider  $x \in S$  and  $y \in X \setminus S(x^*)$  such that  $d(x, y) \leq \delta$ . It suffices to prove that  $x$  is a loser. Consider some  $y^* \in S$  such that

$d(x^*, y^*) \leq \delta$ . Then by the Lipschitz condition

$$\begin{aligned} \mu_{\text{av}}(y^*) &\geq \mu(y^*) - r_0 \geq \mu^* - 2r_0, \\ \mu_{\text{av}}(x) &\leq \mu(x) + r_0 \leq \mu(y) + r_0 \\ &\leq \mu_0 + 2r_0 \leq \mu^* - 5r_0 \\ \mu_{\text{av}}(y^*) - \mu_{\text{av}}(x) &\geq 3r_0 > 2r(t) + \delta. \end{aligned}$$

*Proof of Theorem 4.1:* Let us fix a function  $f \in \omega(\log t)$ . Then  $f(t) = \alpha(t) \log(t)$  where  $\alpha(t) \rightarrow \infty$ . Without loss of generality, assume that  $\alpha(t)$  is non-decreasing. (If not, then instead of  $f(t)$  use  $g(t) = \beta(t) \log(t)$ , where  $\beta(t) = \inf\{\alpha(t') : t' \geq t\}$ .)

For part (a), define  $k_t = \lfloor \sqrt{g(t)/\log t} \rfloor$ ,  $n_t = \lfloor k_t \log t \rfloor$ , and  $r_t = 4\sqrt{(\log t)/n_t}$ . Note that  $r_t \rightarrow 0$ .

The algorithm proceeds in phases of a doubly exponential length<sup>7</sup>. A given phase  $i = 1, 2, 3, \dots$  lasts for  $T = 2^{2^i}$  rounds. In this phase, first we call the exploration subroutine `EXPL( $k_T, n_T, r_T$ )`. Let  $x_{\text{or}} \in X$  be the point returned by this subroutine. Then we play  $x_{\text{or}}$  till the end of the phase. This completes the description of the algorithm.

Fix a problem instance  $\mathcal{I}$ . Let  $W_i$  be the total reward accumulated by the algorithm in phase  $i$ , and let  $R_i = 2^{2^i} \mu^* - W_i$  be the corresponding share of regret. By Lemma 4.2 there exists  $i_0 = i_0(\mathcal{I})$  such that for any phase  $i \geq i_0$  we have, letting  $T = 2^{2^i}$  be the phase duration, that  $R_i \leq k_T n_T \leq g(T)$  with probability at least  $1 - T^{-2}$ , and therefore  $E[R_i] \leq g(T) + T^{-1}$ . For any  $t > t_0 = 2^{2^{i_0}}$  it follows by summing over  $i \in \{i_0, i_0 + 1, \dots, \lceil \log \log t \rceil\}$  that  $R_{\mathcal{A}, \mathcal{I}}(t) = O(t_0 + g(t))$ . Note that we have used the fact that  $\alpha(t)$  is non-decreasing.

For part (b), we separate exploration and exploitation. For exploration, we run `EXPL()` on the *free peeks*. For exploitation, we use the point returned by `EXPL()` in the previous phase. Specifically, define  $k_t = n_t = \lfloor \sqrt{t} \rfloor$ , and  $r_t = 4\sqrt{(t^{1/4})/n_t}$ . The algorithm proceeds in phases of exponential length. A given phase  $i = 1, 2, 3, \dots$  lasts for  $T = 2^i$  rounds. In this phase, we run the exploration subroutine `EXPL( $k_T, n_T, r_T$ )` on the *free peeks*. In each round, we *bet* on the point returned by `EXPL()` in the previous phase. This completes the description of the algorithm.

By Lemma 4.2 there exists  $i_0 = i_0(\mathcal{I})$  such that in any phase  $i \geq i_0$  the algorithm incurs zero regret with probability at least  $1 - e^{-\Omega(i)}$ . Thus the total regret after  $t > 2^{i_0}$  rounds is at most  $t_0 + O(1)$ .

<sup>7</sup>The doubly exponential phase length is necessary in order to get  $f$ -tractability. If we employed the more familiar *doubling trick* of using phase length  $2^i$  (as in [4, 24, 27] for example) then the algorithm would only be  $f(t) \log t$ -tractable.

## 5 The $(\log t)$ -intractability for infinite metric spaces: proof of Theorem 1.2

Consider an infinite metric space  $(X, d)$ . In view of Theorem 1.4, we can assume that the completion  $X^*$  of  $X$  is compact. It follows that there exists  $x^* \in X^*$  such that  $x_i \rightarrow x^*$  for some sequence  $x_1, x_2, \dots \in X$ . Let  $r_i = d(x_i, x^*)$ . Without loss of generality, assume that  $r_{i+1} < \frac{1}{2} r_i$  for each  $i$ , and the diameter of  $X$  is 1.

Let us define an ensemble of payoff functions  $\mu_i : X \rightarrow [0, 1]$ ,  $i \in \mathbb{N}$ , where  $\mu_0$  is the “baseline” function, and for each  $i \geq 1$  function  $\mu_i$  is the “counterexample” in which a neighborhood of  $x_i$  has slightly higher payoffs. The “baseline” is defined by  $\mu_0(x) = \frac{1}{2} - \frac{d(x, x^*)}{8}$ , and the “counterexamples” are given by

$$\mu_i(x) = \mu_0(x) + \nu_i(x),$$

where  $\nu_i(x) = \frac{3}{4} \max(0, \frac{r_i}{3} - d(x, x^*))$ .

Note that both  $\mu_0$  and  $\nu_i$  are  $\frac{1}{8}$ -Lipschitz and  $\frac{3}{4}$ -Lipschitz w.r.t.  $(X, d)$ , respectively, so  $\mu_i$  is  $\frac{7}{8}$ -Lipschitz w.r.t.  $(X, d)$ . Let us fix a MAB algorithm  $\mathcal{A}$  and assume that it is  $(\log t)$ -tractable. Then for each  $i \geq 0$  there exists a constant  $C_i$  such that  $R_{(\mathcal{A}, \mu_i)}(t) < C_i \log t$  for all times  $t$ . We will show that this is not possible.

Intuitively, the ability of an algorithm to distinguish between payoff functions  $\mu_0$  and  $\mu_i$ ,  $i \geq 1$  depends on the number of samples in the ball  $B_i = B(x_i, r_i/3)$ . (This is because  $\mu_0 = \mu_i$  outside  $B_i$ .) In particular, the number of samples itself cannot be too different under  $\mu_0$  and under  $\mu_i$ , unless it is large. To formalize this idea, let  $N_i(t)$  be the number of times algorithm  $\mathcal{A}$  selects a strategy in the ball  $B_i$  during the first  $t$  rounds, and let  $\sigma(N_i(t))$  be the corresponding  $\sigma$ -algebra. Let  $\mathbb{P}_i[\cdot]$  and  $\mathbb{E}_i[\cdot]$  be, respectively, the distribution and expectation induced by  $\mu_i$ . Then we can connect  $\mathbb{E}_0[N_i(t)]$  with the probability of any event  $S \in \sigma(N_i(t))$  as follows.

CLAIM 5.1. For any  $i \geq 1$  and any event  $S \in \sigma(N_i(t))$  it is the case that

$$(5.3) \quad \mathbb{P}_i[S] < \frac{1}{3} \leq \mathbb{P}_0[S] \\ \Rightarrow -\ln(\mathbb{P}_i[S]) - \frac{3}{e} \leq O(r_i^2) \mathbb{E}_0[N_i(t)].$$

Claim 5.1 is proved using KL-divergence techniques, see Appendix B for details. To complete the proof of the theorem, we claim that for each  $i \geq 1$  it is the case that  $\mathbb{E}_0[N_i(t)] \geq \Omega(r_i^{-2} \log t)$  for any sufficiently large  $t$ . Indeed, fix  $i$  and let  $S = \{N_i(t) < r_i^{-2} \log t\}$ . Since

$$C_i \log t > R_{(\mathcal{A}, \mu_i)}(t) \geq \mathbb{P}_i(S) (t - r_i^{-2} \log t) \frac{r_i}{8},$$

it follows that  $\mathbb{P}_i(S) < t^{-1/2} < \frac{1}{3}$  for any sufficiently large  $t$ . Then by Claim 5.1 either  $\mathbb{P}_0(S) < \frac{1}{3}$  or the

consequent in (5.3) holds. In both cases  $\mathbb{E}_0[N_i(t)] \geq \Omega(r_i^{-2} \log t)$ . Claim proved.

Finally, the fact that  $\mu_0(x^*) - \mu_0(x) \geq r_i/12$  for every  $x \in B_i$  implies that

$$R_{(\mathcal{A}, \mu_0)}(t) \geq \frac{r_i}{12} \mathbb{E}_0[N_i(t)] \geq \Omega(r_i^{-1} \log t)$$

which completes the proof since  $r_i^{-1} \rightarrow \infty$  as  $i \rightarrow \infty$ .

## 6 Tractability via more intuitive oracle access

In Theorem 4.1, the algorithm accesses the metric space via two oracles: a very intuitive *covering oracle*, and a less intuitive *ordering oracle*. In this section we show that for a wide family of metric spaces — including, for example, compact metric spaces with a finite number of limit points — the ordering oracle is not needed: we provide an algorithm which accesses the metric space via a finite set of covering oracles. We will consider metric spaces of finite *Cantor-Bendixson rank*, a classic notion from point topology.

DEFINITION 6.1. Fix a metric space  $(X, d)$ . If for some  $x \in X$  there exists a sequence of points in  $X \setminus \{x\}$  which converges to  $x$ , then  $x$  is called a limit point. For  $S \subset X$  let  $\text{LIM}(S)$  denote the limit set: the set of all limit points of  $S$ . Let  $\text{LIM}(S, 0) = S$ , and  $\text{LIM}(S, i) = \text{LIM}(\text{LIM}(\dots \text{LIM}(S)))$ , where  $\text{LIM}(\cdot)$  is applied  $i$  times. The Cantor-Bendixson rank of  $(X, d)$  is defined as  $\sup\{n : \text{LIM}(X, n) \neq \emptyset\}$ .

Let us say that a *Cantor-Bendixson metric space* is one with a finite Cantor-Bendixson rank. In order to apply Theorem 4.1, we show that any such metric space is well-orderable.

LEMMA 6.1. Any Cantor-Bendixson metric space is well-orderable.

*Proof.* Any finite metric space is trivially well-orderable. To prove the lemma, it suffices to show the following: any metric space  $(X, d)$  is well-orderable if so is  $(\text{LIM}(X), d)$ .

Let  $X_1 = X \setminus \text{LIM}(X)$  and  $X_2 = \text{LIM}(X)$ . Suppose  $(X_2, d)$  admits a topological well-ordering  $\prec_2$ . Define a binary relation  $\prec$  on  $X$  as follows. Fix an arbitrary well-ordering  $\prec_1$  on  $X_1$ . For any  $x, y \in X$  posit  $x \prec y$  if either (i)  $x, y \in X_1$  and  $x \prec_1 y$ , or (ii)  $x, y \in X_2$  and  $x \prec_2 y$ , or (iii)  $x \in X_1$  and  $y \in X_2$ . It is easy to see that  $(X, \prec)$  is a well-ordering.

It remains to prove that an arbitrary initial segment  $Y = \{x \in X : x \prec y\}$  is open in  $(X, d)$ . We need to show that for each  $x \in Y$  there is a ball  $B(x, \epsilon)$ ,  $\epsilon > 0$  which is contained in  $Y$ . This is true if  $x \in X_1$  since by definition each such  $x$  is an isolated point in  $X$ . If

$x \in X_2$  then  $Y = X_1 \cup Y_2$  where  $Y_2 = \{x \in X_2 : x \prec_2 y\}$  is the initial segment of  $X_2$ . Since  $Y_2$  is open in  $(X_2, d)$ , there exists  $\epsilon > 0$  such that  $B_{X_2}(x, \epsilon) \subset Y_2$ . It follows that  $B_X(x, \epsilon) \subset B_{X_2}(x, \epsilon) \cup X_1 \subset Y$ .

The structure of a Cantor-Bendixson metric space is revealed by a partition of  $X$  into subsets  $X_i = \text{LIM}(X, i) \setminus \text{LIM}(X, i+1)$ ,  $0 \leq i \leq n$ . For a point  $x \in X_i$ , we define the *rank* to be  $i$ . The algorithm requires a covering oracle for each  $X_i$ .

**THEOREM 6.1.** *Consider the Lipschitz MAB/experts problem on a compact metric space  $(X, d)$  such that  $\text{LIM}_N(X) = \emptyset$  for some  $N$ . Let  $\mathcal{O}_i$  be the covering oracle for  $X_i = \text{LIM}(X, i) \setminus \text{LIM}(X, i+1)$ . Assume that access to the metric space is provided only via the collection of oracles  $\{\mathcal{O}_i\}_{i=0}^N$ . Then:*

- (a) *the Lipschitz MAB problem on  $(X, d)$  is  $f$ -tractable for every  $f \in \omega(\log t)$ ;*
- (b) *the Lipschitz experts problem on  $(X, d)$  is 1-tractable, even with a double feedback.*

In the rest of this section, consider the setting in Theorem 6.1. We describe the *exploration subroutine*  $\text{EXPL}'()$ , which is similar to  $\text{EXPL}()$  in Section 4 but does not use the ordering oracle. Then we prove a version of Lemma 4.2 for  $\text{EXPL}'()$ . Once we have this lemma, the proof of Theorem 6.1 is identical to that of Theorem 4.1 (and is omitted).

**ALGORITHM 6.1.** Subroutine  $\text{EXPL}'(k, n, r)$ : inputs  $k, n \in \mathbb{N}$  and  $r \in (0, 1)$ , outputs a point in  $X$ .

Call each covering oracle  $\mathcal{O}_i(k)$  and receive a  $\delta_i$ -covering  $S_i$  of  $X$  consisting of at most  $k$  points. Let  $S = \cup_{i=1}^n S_i$ . Play each strategy  $x \in S$  exactly  $n$  times; let  $\mu_{\text{av}}(x)$  be the corresponding sample average. For  $x, y \in S$ , let us say that  $x$  *dominates*  $y$  if  $\mu_{\text{av}}(x) - \mu_{\text{av}}(y) > 2r$ . Call  $x \in S$  a *winner* if  $x$  has a largest rank among the strategies that are not dominated by any other strategy. Output an arbitrary winner if a winner exists, else output an arbitrary point in  $S$ .

Clearly,  $\text{EXPL}(k, n, r)$  takes at most  $knN$  rounds to complete. We show that for sufficiently large  $k, n$  and sufficiently small  $r$  it returns an optimal strategy with high probability.

**LEMMA 6.2.** *Fix a problem instance. Consider increasing functions  $k, n, T : \mathbb{N} \rightarrow \mathbb{N}$  such that  $r(t) := 4\sqrt{(\log T(t))/n(t)} \rightarrow 0$ . Then for any sufficiently large  $t$ , with probability at least  $1 - T^{-2}(t)$ , the subroutine  $\text{EXPL}'(k(t), n(t), r(t))$  returns an optimal strategy.*

*Proof.* Use the notation from Algorithm 6.1. Fix  $t$  and consider a run of  $\text{EXPL}'(k(t), n(t), r(t))$ . Call this run

*clean* if for each  $x \in S$  we have  $|\mu_{\text{av}}(x) - \mu(x)| \leq r(t)$ . By Chernoff Bounds, this happens with probability at least  $1 - T^{-2}(t)$ . In the rest of the proof, let us assume that the run is clean.

Let us introduce some notation. Let  $\mu$  be the payoff function and let  $\mu^* = \sup(\mu, X)$ . Call  $x \in X$  *optimal* if  $\mu(x) = \mu^*$ . (There exists an optimal strategy since  $(X, d)$  is compact.) Let  $i^*$  be the largest rank of any optimal strategy. Let  $X^*$  be the set of all optimal strategies of rank  $i^*$ . Let  $Y = \text{LIM}(X, i^*)$ . Since each point  $x \in X_{i^*}$  is an isolated point in  $Y$ , there exists some  $r(x) > 0$  such that  $x$  is the only point of  $B(x, r(x))$  that lies in  $Y$ .

We claim that  $\sup(\mu, Y \setminus X^*) < \mu^*$ . Indeed, consider  $C = \cup_{x \in X^*} B(x, r(x))$ . This is an open set. Since  $Y$  is closed,  $Y \setminus C$  is closed, too, hence compact. Therefore there exists  $y \in Y \setminus C$  such that  $\mu(y) = \sup(\mu, Y \setminus C)$ . Since  $X^* \subset C$ ,  $\mu(y)$  is not optimal, i.e.  $\mu(y) < \mu^*$ . Finally, by definition of  $r(x)$  we have  $Y \setminus C = Y \setminus X^*$ . Claim proved.

Pick any  $x^* \in X^*$ . Let  $\mu_0 = \sup(\mu, Y \setminus X^*)$ . Assume that  $t$  is large enough so that  $r(t) < (\mu^* - \mu_0)/4$  and  $\delta_{i^*} < r(x^*)$ . Note, the  $\delta_{i^*}$ -covering  $S_{i^*}$  contains  $x^*$ .

Finally, we claim that in a clean phase,  $x^*$  is a winner, and all winners lie in  $X^*$ . Indeed, note that  $x^*$  dominates any non-optimal strategy  $y \in S$  of larger or equal rank, i.e. any  $y \in S \cap (Y \setminus X^*)$ . This is because  $\mu_{\text{av}}(x^*) - \mu_{\text{av}}(y) \geq \mu^* - \mu_0 - 2r > 2$ . The claim follows since any optimal strategy cannot be dominated by any other strategy.

## 7 Boundary of tractability: Theorem 1.4

In this section we prove Theorem 1.4. In Appendix A we reduce the theorem to that on complete metric spaces. We will use a basic fact that a complete metric space is compact if and only if for any  $r > 0$ , it can be covered by a finite number of balls of radius  $r$ .

**Algorithmic result.** We consider a compact metric space  $(X, d)$  and use an extension of the *naive algorithm* from [24, 27]. In each phase  $i$  (which lasts for  $t_i$  round) we fix a covering of  $X$  with  $N_i < \infty$  balls of radius  $2^{-i}$  (such covering exists by compactness), and run a fresh instance of the  $N_i$ -armed bandit algorithm UCB1 from [3] on the centers of these balls. (This algorithm is for the “basic” MAB problem, in the sense that it does not look at the distances in the metric space.) The phase durations  $t_i$  need to be tuned to the  $N_i$ ’s. In the setting considered in [24, 27] (bounded covering dimension) it suffices to tune each  $t_i$  to the corresponding  $N_i$  in a fairly natural way. The difficulty in the present setting is that there are no guarantees on how fast the  $N_i$ ’s grow. To take this into account, we fine-tune each  $t_i$  to (essentially) all covering numbers  $N_1, \dots, N_{i+1}$ .

Let  $R_k(t)$  be the expected regret accumulated by the algorithm in the first  $t$  rounds of phase  $k$ . Using the off-the-shelf regret guarantees for UCBL, it is easy to see [24, 27] that

$$(7.4) \quad R_k(t) \leq O(\sqrt{N_k t \log t}) + \epsilon_k t \leq \epsilon_k \max(t_k^*, t),$$

where  $t_k^* = 2 \frac{N_k}{\epsilon_k^2} \log \frac{N_k}{\epsilon_k}$ .

Let us specify phase durations  $t_i$ . They are defined very differently from the ones in [24, 27]. In particular, in [24, 27] each  $t_i$  is fine-tuned to the corresponding covering number  $N_i$  by setting  $t_i = t_i^*$ , and the analysis works out for metric spaces of bounded covering dimension. In our setting, we fine-tune each  $t_i$  to (essentially) all covering numbers  $N_1, \dots, N_{i+1}$ . Specifically, we define the  $t_i$ 's inductively as follows:

$$t_i = \min(t_i^*, t_{i+1}^*, 2 \sum_{j=1}^{i-1} t_j).$$

This completes the algorithm, call it  $\mathcal{A}$ .

**LEMMA 7.1.** *Consider the Lipschitz MAB problem on a compact and complete metric space  $(X, d)$ . Then  $R_{\mathcal{A}}(t) \leq 5\epsilon(t)t$ , where  $\epsilon(t) = \min\{2^{-k} : t \leq s_k\}$  and  $s_k = \sum_{i=1}^k t_i$ . In particular,  $R_{\mathcal{A}}(t) = o(t)$ .*

*Proof.* First we claim that  $R_{\mathcal{A}}(s_k) \leq 2\epsilon_k s_k$  for each  $k$ . Use induction on  $k$ . For the induction base, note that  $R_{\mathcal{A}}(s_1) = R_1(t_1) \leq \epsilon_1 t_1$  by (7.4). Assume the claim holds for some  $k-1$ . Then

$$\begin{aligned} R_{\mathcal{A}}(s_k) &= R_{\mathcal{A}}(s_{k-1}) + R_k(t_k) \\ &\leq 2\epsilon_{k-1} s_{k-1} + \epsilon_k t_k \\ &\leq 2\epsilon_k (s_{k-1} + t_k) = 2\epsilon_k s_k, \end{aligned}$$

claim proved. Note that we have used (7.4) and the facts that  $t_k \geq t_k^*$  and  $t_k \geq 2s_{k-1}$ .

For the general case, let  $T = s_{k-1} + t$ , where  $t \in (0, t_k)$ . Then by (7.4) we have that

$$\begin{aligned} R_k(t) &\leq \epsilon_k \max(t_k^*, t) \leq \epsilon_k \max(t_{k-1}, t) \leq \epsilon_k T, \\ R_{\mathcal{A}}(T) &= R_{\mathcal{A}}(s_{k-1}) + R_k(T) \leq 2\epsilon_{k-1} s_{k-1} + \epsilon_k T \leq 5\epsilon_k T. \end{aligned}$$

**Lower bound: proof sketch.** For the lower bound, we consider a metric space  $(X, d)$  with an infinitely many disjoint balls  $B(x_i, r_*)$  for some  $r_* > 0$ . For each ball  $i$  we define the *wedge function* supported on this ball:

$$G_{(i,r)}(x) = \begin{cases} \min\{r_* - d(x, x_i), r_* - r\} & \text{if } x \in B(x_i, r_*) \\ 0 & \text{otherwise.} \end{cases}$$

The balls are partitioned into two infinite sets: the *ordinary* and *special* balls. The random payoff function

is then defined by taking a constant function, adding the wedge function on each special ball, and randomly adding or subtracting the wedge function on each ordinary ball. Thus, the expected payoff is constant throughout the metric space except that it assumes higher values on the special balls. However, the algorithm has no chance of ever finding these balls, because at time  $t$  they are statistically indistinguishable from the  $2^{-t}$  fraction of ordinary balls that randomly happen to never subtract their wedge function during the first  $t$  steps of play.

**Lower bound: full proof.** Suppose  $(X, d)$  is not compact. Fix  $r > 0$  such that  $X$  cannot be covered by a finite number of balls of radius  $r$ . There exists a countably infinite subset  $S \subset X$  such that the balls  $B(x, r)$ ,  $x \in S$  are mutually disjoint. (Such subset can be constructed inductively.) Number the elements of  $S$  as  $s_1, s_2, \dots$ , and denote the ball  $B(s_i, r)$  by  $B(i)$ .

Suppose there exists a Lipschitz experts algorithm  $\mathcal{A}$  that is  $g(t)$ -tractable for some  $g \in o(t)$ . Pick an increasing sequence  $t_1, t_2, \dots \in \mathbb{N}$  such that  $t_{k+1} > 2t_k \geq 10$  and  $g(t_k) < r_k t_k/k$  for each  $k$ , where  $r_k = r/2^{k+1}$ . Let  $m_0 = 0$  and  $m_k = \sum_{i=1}^k 4^{t_i}$  for  $k > 0$ , and let  $I_k = \{m_k + 1, \dots, m_{k+1}\}$ . The intervals  $I_k$  form a partition of  $\mathbb{N}$  into sets of sizes  $4^{t_1}, 4^{t_2}, \dots$ . For every  $i \in \mathbb{N}$ , let  $k$  be the unique value such that  $i \in I_k$  and define the following Lipschitz function supported in  $B(s_i, r)$ :

$$G_i(x) = \begin{cases} \min\{r - d(x, s_i), r - r_k\} & \text{if } x \in B(i) \\ 0 & \text{otherwise.} \end{cases}$$

If  $J \subseteq \mathbb{N}$  is any set of natural numbers, we can define a distribution  $\mathbb{P}_J$  on payoff functions by sampling independent, uniformly-random signs  $\sigma_i \in \{\pm 1\}$  for every  $i \in \mathbb{N}$  and defining the payoff function to be

$$\pi = \frac{1}{2} + \sum_{i \in J} G_i + \sum_{i \notin J} \sigma_i G_i.$$

Note that the distribution  $\mathbb{P}_J$  has expected payoff function  $\mu = \frac{1}{2} + \sum_{i \in J} G_i$ . Let us define a distribution  $\mathcal{P}$  over problem instances  $\mathbb{P}_J$  by letting  $J$  be a random subset of  $\mathbb{N}$  obtained by sampling exactly one element  $j_k$  of each set  $I_k$  uniformly at random, independently for each  $k$ .

Intuitively, consider an algorithm that is trying to discover the value of  $j_k$ . Every time a payoff function  $\pi_t$  is revealed, we get to see a random  $\{\pm 1\}$  sample at every element of  $I_k$  and we can eliminate the possibility that  $j_k$  is one of the elements that sampled  $-1$ . This filters out about half the elements of  $I_k$  in every time step, but  $|I_k| = 4^{t_k}$  so on average it takes  $2t_k$  steps before we can discover the identity of  $j_k$ . Until that time, whenever we play a strategy in  $\cup_{i \in I_k} B(i)$ , there

is a constant probability that our regret is at least  $r_k$ . Thus our regret is bounded below by  $r_k t_k \geq kg(t_k)$ . This rules out the possibility of a  $g(t)$ -tractable algorithm. The following lemma makes this argument precise.

LEMMA 7.2.  $\Pr_{\mathbb{P} \in \mathcal{P}}[R_{(\mathcal{A}, \mathbb{P})}(t) = O_\mu(g(t))] = 0$ .

*Proof.* Let  $j_1, j_2, \dots$  be the elements of the random set  $J$ , numbered so that  $j_k \in I_k$  for all  $k$ . For any  $i, t \in \mathbb{N}$ , let  $\sigma(i, t)$  denote the value of  $\sigma_i$  sampled at time  $t$  when sampling the sequence of i.i.d. payoff functions  $\pi_t$  from distribution  $\mathbb{P}_J$ . We know that  $\sigma(j_k, t) = 1$  for all  $t$ . In fact if  $S(k, t)$  denotes the set of all  $i \in I_k$  such that  $\sigma(i, 1) = \sigma(i, 2) = \dots = \sigma(i, t) = 1$  then conditional on the value of the set  $S(k, t)$ , the value of  $j_k$  is distributed uniformly at random in  $S(k, t)$ . As long as this set  $S(k, t)$  has at least  $n$  elements, the probability that the algorithm picks a strategy  $x_t$  belonging to  $B(j_k)$  at time  $t$  is bounded above by  $\frac{1}{n}$ , even if we condition on the event that  $x_t \in \cup_{i \in I_k} B(i)$ . For any given  $i \in I_k \setminus \{j_k\}$ , we have  $\mathbb{P}_J(i \in S(k, t)) = 2^{-t}$  and these events are independent for different values of  $i$ . Setting  $n = 2^{tk}$ , so that  $|I_k| = n^2$ , we have

$$\begin{aligned} \mathbb{P}_J[|S(k, t)| \leq n] &\leq \sum_{R \subset I_k, |R|=n} \mathbb{P}_J[S(k, t) \subseteq R] \\ &= \binom{n^2}{n} (1 - 2^{-t})^{n^2 - n} < \left(n^2 \cdot (1 - 2^{-t})^{n-1}\right)^n \\ (7.5) \quad &< \exp(n(2 \ln(n) - (n-1)/2^t)). \end{aligned}$$

As long as  $t \leq t_{k-1}$ , the relation  $t_k > 2t$  implies  $(n-1)/2^t > \sqrt{n}$  so the expression (7.5) is bounded above by  $\exp(-n\sqrt{n} + 2n \ln(n))$ , which equals  $\exp(-8^{tk} + 2 \ln(4)t_k 4^{tk})$  and is in turn bounded above by  $\exp(-8^{tk}/2)$ .

Let  $B(j_{>k}) = B(j_{k+1}) \cup B(j_{k+2}) \cup \dots$ , and let  $N(t, k)$  denote the random variable that counts the number of times  $\mathcal{A}$  selects a strategy in  $B(j_{>k})$  during rounds  $1, \dots, t$ . We have already demonstrated that for all  $t \leq t_k$ ,

$$(7.6) \quad \Pr_{\mathbb{P}_J \in \mathcal{P}}(x_t \in B(j_{>k})) \leq 2^{-tk+1} + \sum_{\ell > k} \exp(-8^{t\ell}/2) < 2^{1-tk+1},$$

where the term  $2^{-tk+1}$  accounts for the event that  $S(\ell, t)$  has at least  $2^{tk+1}$  elements, where  $\ell$  is the index of the set  $I_\ell$  containing the number  $i$  such that  $x_t \in B(i)$ , if such an  $i$  exists. Equation (7.6) implies the bound  $\mathbb{E}_{\mathbb{P}_J \in \mathcal{P}}[N(t_k, k)] < t_k \cdot 2^{1-tk+1}$ . By Markov's inequality, the probability that  $N(t_k, k) > t_k/2$  is less than  $2^{2-tk+1}$ . By Borel-Cantelli, almost surely the number of  $k$  such that  $N(t_k, k) \leq t_k/2$  is finite. The algorithm's expected regret at time  $t$  is bounded below by  $r_k(t_k - N(t_k, k))$ , so with probability 1, for all but finitely many  $k$  we have  $R_{(\mathcal{A}, \mathbb{P}_J)}(t_k) \geq r_k t_k / 2 \geq (k/2)g(t_k)$ . This establishes that  $\mathcal{A}$  is not  $g(t)$ -tractable.

## 8 Lipschitz experts in a (very) high dimension

In this section we discuss the full-feedback Lipschitz experts problem in (very) high dimensional metric spaces. We posit a new notion of dimensionality which is well-suited to describe regret in such problems, and analyze the performance of a simple algorithm in terms of this notion.

Fix a metric space  $(X, d)$ . For a subset  $Y \subset X$  and  $\delta > 0$ , a  $\delta$ -covering of  $Y$  is a collection of sets of diameter at most  $\delta$  whose union contains  $Y$ . A subset  $S \subset X$  is a  $\delta$ -hitting set for  $Y$  if  $Y \subset \cup_{x \in S} B(x, \delta)$ . (So if  $S$  is a hitting set for some  $\delta$ -covering of  $Y$  then it is a  $\delta$ -hitting set for  $Y$ .)

Let  $N_\delta(Y)$  be the minimal size (cardinality) of a  $\delta$ -covering of  $Y$ , i.e. the smallest number of sets of diameter at most  $\delta$  sufficient to cover  $Y$ . The standard definition of the *covering dimension* is

$$(8.7) \quad \text{Cov}(Y) = \limsup_{\delta > 0} \frac{\log N_\delta(Y)}{\log(1/\delta)}.$$

Covering dimension and its refinements have been essential in the study of the Lipschitz MAB problem [27]. However, for the full-feedback Lipschitz experts problem the metrics with bounded covering dimension are too "easy". We need to consider a much broader class of metrics that satisfy a non-trivial bound on what we call the *log-covering-dimension*:

$$(8.8) \quad \text{LCD}(Y) = \limsup_{\delta > 0} \frac{\log \log N_\delta(Y)}{\log(1/\delta)}.$$

To give an example of a metric space with a non-trivial log-covering dimension, let us consider a *uniform tree* – a rooted tree in which all nodes at the same level have the same number of children. An  $\epsilon$ -uniform tree metric is a metric on the leaves of an infinitely deep uniform tree, in which the distance between two leaves is  $\epsilon^{-i}$ , where  $i$  is the level of their least common ancestor. It is easy to see that an  $\epsilon$ -uniform tree metric such that the branching factor at each level  $i$  is  $\exp(\epsilon^{-ib}(2^b - 1))$  has log-covering dimension  $b$ .

For another example, consider the space of all probability distributions over a given metric space  $(X, d)$  under the Wasserstein  $W_1$  metric, a.k.a. the Earthmover distance. It is not hard to show that the log-covering dimension of this space is at most the covering dimension of  $(X, d)$ . Indeed, if  $(X, d)$  has diameter  $D$  and is covered by balls of radius  $\epsilon/2$  centered at points  $x_1, \dots, x_M$ , then the Wasserstein metric is covered by balls of radius  $\epsilon$  centered at the distributions whose support is contained in  $\{x_1, \dots, x_M\}$  and whose probabilities are rational numbers with denominator  $\lceil 2DM/\epsilon \rceil$ . See the full version for details, as well as a matching lower bound on the log-covering dimension of the Wasserstein metric.

To see how the log-covering dimension is relevant to the Lipschitz experts problem, consider the following simple algorithm, called `NAIVEEXPERTS(b)`.<sup>8</sup> The algorithm is parameterized by  $b > 0$ . It runs in phases. Each phase  $i$  lasts for  $T = 2^i$  rounds, and outputs its *best guess*  $x_i^* \in X$ , which is played throughout phase  $i + 1$ . During phase  $i$ , the algorithm picks a  $\delta$ -hitting set for  $X$  of size at most  $N_\delta(X)$ , for  $\delta = T^{-1/(b+2)}$ . By the end of the phase,  $x_i^*$  as defined as the point in  $S$  with the highest sample average (breaking ties arbitrarily). This completes the description of the algorithm.

It is easy to see that the regret of `NAIVEEXPERTS` is naturally described in terms of the log-covering dimension. The proof is based the argument from [24]. We restate it here for the sake of completeness, and to explain how the new dimensionality notion is used.

**THEOREM 8.1.** *Consider the full-feedback Lipschitz experts problem on a metric space  $(X, d)$ . For each  $b > \text{LCD}(X)$ , algorithm `NAIVEEXPERTS(b)` achieves regret  $R(t) = O(t^{1-1/(b+2)})$ .*

*Proof.* Let  $N_\delta = N_\delta(X)$ , and let  $\mu$  be the expected payoff function. Consider a given phase  $i$  of the algorithm. Let  $T = 2^i$  be the phase duration. Let  $\delta = T^{-1/(b+2)}$ , and let  $S \subset X$  the  $\delta$ -hitting set chosen in this phase. Note that for any sufficiently large  $T$  it is the case that  $N_\delta < 2^{\delta^{-b}}$ . For each  $x \in S$ , let  $\mu_T(x)$  be the sample average of the feedback from  $x$  by the end of the phase. Then by Chernoff bounds,

$$(8.9) \quad \Pr[|\mu_T(x) - \mu(x)| < r_T] > 1 - (TN_\delta)^{-3},$$

where  $r_T = \sqrt{8 \log(TN_\delta)/T} < 2\delta$ . Note that  $\delta$  is chosen specifically to ensure that  $r_T \leq O(\delta)$ .

We can neglect the regret incurred when the event in (8.9) does not hold for some  $x \in S$ . From now on, let us assume that the event in (8.9) holds for all  $x \in S$ . Let  $x^*$  be an optimal strategy, and  $x_i^* = \text{argmax}_{x \in S} \mu_T(x)$  be the “best guess”. Let  $x \in S$  be a point that covers  $x^*$ . Then

$$\begin{aligned} \mu(x_i^*) &\geq \mu_T(x_i^*) - 2\delta \geq \mu_T(x) - 2\delta \\ &\geq \mu(x) - 4\delta \geq \mu(x^*) - 5\delta. \end{aligned}$$

Thus the total regret  $R_{i+1}$  from phase  $i + 1$  is

$$R_{i+1} \leq 2^{i+1} (\mu(x^*) - \mu(x_i^*)) \leq O(\delta T) = O(T^{1-1/(2+b)}).$$

Thus the total regret summed over phases is as claimed.

<sup>8</sup>This algorithm is a version of the “naive algorithm” [24, 27] for the Lipschitz MAB problem. A similar algorithm has been used by [20] to obtain regret  $R(T) = O(\sqrt{T})$  for metric spaces of finite covering dimension.

## 8.1 The uniformly Lipschitz experts problem

We now turn our attention to a restricted version of the full-feedback Lipschitz experts problem in which a problem instance  $(X, d, \mathbb{P})$  satisfies a further property that each function  $f \in \text{support}(\mathbb{P})$  is itself a Lipschitz function on  $(X, d)$ . We show that for this version, `NAIVEEXPERTS` obtains a significantly better regret guarantee, via a more involved analysis. In fact, (as we will see in the next section) for a wide class of metric spaces, including  $\epsilon$ -uniform tree metrics, there is a matching upper bound.

**THEOREM 8.2.** *Consider the uniformly Lipschitz experts problem with full feedback. Fix a metric space  $(X, d)$ . For each  $b > \text{LCD}(X)$  such that  $b \geq 2$ , `NAIVEEXPERTS(b-2)` achieves regret  $R(t) = O(t^{1-1/b})$ .*

*Proof.* The preliminaries are similar to those in the proof of Theorem 8.1. For simplicity, assume  $b \geq 2$ . Let  $N_\delta = N_\delta(X)$ , and let  $\mu$  be the expected payoff function. Consider a given phase  $i$  of the algorithm. Let  $T = 2^i$  be the phase duration. Let  $\delta = T^{-1/b}$ , and let  $S$  be the  $\delta$ -hitting set chosen in this phase. (The specific choice of  $\delta$  is the only difference between the algorithm here and the algorithm in Theorem 8.1.) Note that  $|S| \leq N_\delta$ , and for any sufficiently large  $T$  it is the case that  $N_\delta < 2^{\delta^{-b}}$ .

The rest of the analysis holds for any set  $S$  such that  $|S| \leq N_\delta$ . (That is, it is not essential that  $S$  is a  $\delta$ -hitting set for  $X$ .) For each  $x \in S$ , let  $\nu(x)$  be the sample average of the feedback from  $x$  by the end of the phase. Let  $y_i^* = \text{argmax}(\mu, S)$  be the optimal strategy in the chosen sample, and let  $x_i^* = \text{argmax}(\nu, S)$  be the algorithm’s “best guess”. The crux is to show that

$$(8.10) \quad \Pr[\mu(y_i^*) - \mu(x_i^*) \leq O(\delta \log T)] > 1 - T^{-3}.$$

Once (8.10) is established, the remaining steps is exactly as the proof of Theorem 8.1.

Proving (8.10) requires a new technique. The obvious approach – to use Chernoff Bounds for each  $x \in S$  separately and then take a Union Bound – does not work, essentially because one needs to take the Union Bound over too many points. Instead, we will use a more efficient version tail bound: for each  $x, y \in X$ , we will use Chernoff Bounds applied to the random variable  $f(x) - f(y)$ , where  $f \sim \mathbb{P}$  and  $(X, d, \mathbb{P})$  is the problem instance. For a more convenient notation, we define

$$\Delta(x, y) = [\mu(x) - \mu(y)] + [\nu(y) - \nu(x)],$$

Then for any  $N \in \mathbb{N}$  we have

$$(8.11) \quad \Pr \left[ |\Delta(x, y)| \leq d(x, y) \sqrt{8 \log(TN)/T} \right] > 1 - (TN)^{-3}.$$

The point is that the “slack” in the Chernoff Bound is scaled by the factor of  $d(x, y)$ . This is because each  $f \in \text{support}(\mathbb{P})$  is a Lipschitz function on  $(X, d)$ ,

In order to take advantage of (8.11), let us define the following structure that we call the *covering tree* of the metric space  $(X, d)$ . This structure consists of a rooted tree  $\mathcal{T}$  and non-empty subsets  $X(u) \subset X$  for each internal node  $u$ . Let  $V_{\mathcal{T}}$  be the set of all internal nodes. Let  $\mathcal{T}_j$  be the set of all level- $j$  internal nodes (so that  $\mathcal{T}_0$  is a singleton set containing the root). For each  $u \in V_{\mathcal{T}}$ , let  $\mathcal{C}(u)$  be the set of all children of  $u$ . For each node  $u \in \mathcal{T}_j$  the structure satisfies the following two properties: (i) set  $X(u)$  has diameter at most  $2^{-j}$ , (ii) the sets  $X(v)$ ,  $v \in \mathcal{C}(u)$  form a partition of  $X(u)$ . This completes the definition.

By definition of the covering number  $N_{\delta}(\cdot)$  there exist a covering tree  $\mathcal{T}$  in which each node  $u \in \mathcal{T}_j$  has fan-out  $N_{2^{-j}}(X(u))$ . Fix one such covering tree. For each node  $u \in V_{\mathcal{T}}$ , define

$$(8.12) \quad \begin{aligned} \sigma(u) &= \operatorname{argmax}(\mu, \mathcal{X}(u) \cap S) \\ \rho(u) &= \operatorname{argmax}(\nu, \mathcal{X}(u) \cap S), \end{aligned}$$

where the tie-breaking rule is the same as in the algorithm.

Let  $n = \lceil \log \frac{1}{\delta} \rceil$ . Let us say that phase  $i$  is *clean* if the following two properties hold:

- (i) for each node  $u \in V_{\mathcal{T}}$  any two children  $v, w \in \mathcal{C}(u)$  we have  $|\Delta(\sigma(v), \sigma(w))| \leq 4\delta$ .
- (ii) for any  $x, y \in S$  such that  $d(x, y) \leq \delta$  we have  $|\Delta(x, y)| \leq 4\delta$ .

CLAIM 8.1. *For any sufficiently large  $i$ , phase  $i$  is clean with probability at least  $1 - T^{-2}$ .*

*Proof.* To prove (i), let  $j$  be such that  $u \in \mathcal{T}_j$ . We consider each  $j$  separately. Note that (i) is trivial for  $j > n$ . Now fix  $j \leq n$  and apply the Chernoff-style bound (8.11) with  $N = |\mathcal{T}_j|$  and  $(x, y) = (\sigma(v), \sigma(w))$ . Since  $|\mathcal{T}_l| \leq 2^{2^{lb}} |\mathcal{T}_{l-1}|$  for each sufficiently large  $l$ , it follows that  $\log |\mathcal{T}_j| \leq C + \sum_{l=1}^j 2^{lb} \leq C + \frac{4}{3} 2^{jb}$ , where  $C$  is a constant that depends only on the metric space and  $b$ . It is easy to check that for any sufficiently large phase  $i$  (which, in turn, determines  $T$ ,  $\delta$  and  $n$ ), the “slack” in (8.11) is at most  $4\delta$ :

$$\begin{aligned} d(x, y) \sqrt{8 \log(TN)/T} \\ \leq 3 d(x, y) \sqrt{\log(N)/T} \leq 4 \cdot 2^{-j} \sqrt{2^{bj}/2^{bn}} \\ = 4\delta 2^{-(n-j)(b-2)/2} \leq 4\delta. \end{aligned}$$

Interestingly, the right-most inequality above is the only place in the proof where it is essential that  $b \geq 2$ .

To prove (ii), apply (8.11) with  $N = |S|$  similarly. Claim proved.

From now on we will consider clean phase. (We can ignore regret incurred in the event that the phase is not clean.) We focus on the quantity  $\Delta^*(u) = \Delta(\sigma(u), \rho(u))$ . Note that by definition  $\Delta^*(u) \geq 0$ . The central argument of this proof is the following upper bound on  $\Delta^*(u)$ .

CLAIM 8.2. *In a clean phase,  $\Delta^*(u) \leq O(\delta)(n - j)$  for each  $j \leq n$  and each  $u \in \mathcal{T}_j$ .*

*Proof.* Use induction on  $j$ . The base case  $j = n$  follows by part (ii) of the definition of the clean phase, since for  $u \in \mathcal{T}_n$  both  $\sigma(u)$  and  $\rho(u)$  lie in  $X(u)$ , the set of diameter at most  $\delta$ . For the induction step, assume the claim holds for each  $v \in \mathcal{T}_{j+1}$ , and let us prove it for some fixed  $u \in \mathcal{T}_j$ .

Pick children  $u, v \in \mathcal{C}(u)$  such that  $\sigma(u) \in X(v)$  and  $\rho(u) \in X(w)$ . Since the tie-breaking rules in (8.12) is fixed for all nodes in the covering tree, it follows that  $\sigma(u) = \sigma(v)$  and  $\rho(u) = \rho(w)$ . Then

$$\begin{aligned} \Delta^*(w) + \Delta(\sigma(v), \sigma(w)) \\ = \Delta(\sigma(w), \rho(u)) + \Delta(\sigma(u), \sigma(w)) \\ = \mu(\sigma(w)) - \mu(\rho(u)) + \nu(\rho(u)) - \rho(\sigma(w)) \\ \quad + \mu(\sigma(u)) - \mu(\sigma(w)) + \nu(\sigma(w)) - \nu(\sigma(u)) \\ = \Delta^*(u). \end{aligned}$$

Claim follows since  $\Delta^*(w) \leq O(\delta)(n - j - 1)$  by induction, and  $\Delta(\sigma(v), \sigma(w)) \leq 4\delta$  by part (i) in the definition of the clean phase.

To complete the proof of (8.10), let  $u_0$  be the root of the covering tree. Then  $y_i^* = \sigma(u_0)$  and  $x_i^* = \rho(u_0)$ . Therefore by Claim 8.2 (applied for  $\mathcal{T}_0 = \{u_0\}$ ) we have

$$O(\delta n) \geq \Delta^*(u_0) = \Delta^*(y_i^*, x_i^*) \geq \mu(y_i^*) - \mu(x_i^*).$$

## 9 Lipschitz experts: regret characterization

As it turns out, the log-covering dimension (8.8) is not the right notion to characterize regret for arbitrary metric spaces. We need a more refined version, similar to the *max-min-covering dimension* from [27]:

$$(9.13) \quad \text{MaxMinLCD}(X) = \sup_{Y \subset X} \inf_Z \text{LCD}(Z)$$

where the infimum is taken over all open non-empty subsets  $Z \subset Y$ . Note that in general  $\text{MaxMinLCD} \leq \text{LCD}(X)$ . The equality holds for “homogenous” metric spaces such as  $\epsilon$ -uniform tree metrics. We prove that Theorems 1.5 holds with  $b = \text{MaxMinLCD}(X)$ .

THEOREM 9.1. *Fix a metric space  $(X, d)$  and let  $b = \text{MaxMinLCD}(X)$ . The full-feedback Lipschitz experts problem on  $(X, d)$  is  $(t^\gamma)$ -tractable for any  $\gamma > \frac{b+1}{b+2}$ , and not  $(t^\gamma)$ -tractable for any  $\gamma < \frac{b-1}{b}$ .*

For the lower bound, we use a suitably “thick” version of the ball-tree from Section 3 in conjunction with the  $(\epsilon, \delta, k)$ -ensemble idea from Section 3, see Section 9.1. For the algorithmic result, we combine the “naive” experts algorithm (NAIVEEXPERTS) with (an extension of) the *transfinite fat decomposition* technique from [27], see Section 9.2.

The lower bound in Theorem 9.1 holds for the uniformly Lipschitz experts problem. It follows that the upper bound in Theorem 8.2 is optimal for metric spaces such that  $\text{MaxMinLCD}(X) = \text{LCD}(X)$ , e.g. for  $\epsilon$ -uniform tree metrics. In fact, we can plug the improved analysis of NAIVEEXPERTS from Theorem 8.2 into the algorithmic technique from Theorem 9.1 and obtain a matching upper bound in terms of the  $\text{MaxMinLCD}$ . Thus (in conjunction with Theorem 1.3) we have a complete characterization for regret:

**THEOREM 9.2.** *Consider the uniformly Lipschitz experts problem with full feedback. Fix a metric space  $(X, d)$  with uncountably many points, and let  $b = \text{MaxMinLCD}(X)$ . The problem on  $(X, d)$  is  $(t^\gamma)$ -tractable for any  $\gamma > \max(\frac{b-1}{b}, \frac{1}{2})$ , and not  $(t^\gamma)$ -tractable for any  $\gamma < \max(\frac{b-1}{b}, \frac{1}{2})$ .*

The proof of the upper bound in Theorem 9.2 proceeds exactly that in Theorem 9.1, except that we use a more efficient analysis of NAIVEEXPERTS.

### 9.1 The $\text{MaxMinLCD}$ lower bound (Theorem 9.2)

If  $\text{MaxMinLCD}(X) = d$ , and  $\gamma < \frac{d-1}{d}$ , let us first choose  $b < c < d$  such that  $\gamma < \frac{b-1}{b}$ . Let  $Y \subseteq X$  be a subspace such that  $c \leq \inf\{\text{LCD}(Z) : \text{open, nonempty } Z \subseteq Y\}$ . We will repeatedly use the following packing lemma that relies on the fact that  $b < \text{LCD}(U)$  for all nonempty  $U \subseteq Y$ .

**LEMMA 9.1.** *For any nonempty open  $U \subseteq Y$  and any  $r_0 > 0$  there exists  $r \in (0, r_0)$  such that  $U$  contains more than  $2^{r-b}$  disjoint balls of radius  $r$ .*

*Proof.* Let  $r < r_0$  be a positive number such that every covering of  $U$  requires more than  $2^{r-b}$  balls of radius  $2r$ . Such an  $r$  exists, because  $\text{LCD}(U) > b$ . Now let  $\mathcal{P} = \{B_1, B_2, \dots, B_M\}$  be any maximal collection of disjoint  $r$ -balls. For every  $y \in Y$  there must exist some ball  $B_i$  ( $1 \leq i \leq M$ ) whose center is within distance  $2r$  of  $y$ , as otherwise  $B(y, r)$  would be disjoint from every element of  $\mathcal{P}$  contradicting the maximality of that collection. If we enlarge each ball  $B_i$  to a ball  $B_i^+$  of radius  $2r$ , then every  $y \in Y$  is contained in one of the balls  $\{B_i^+ | 1 \leq i \leq M\}$ , i.e. they form a covering of  $Y$ . Hence  $M \geq 2^{r-b}$  as desired.

Using the packing lemma we recursively construct an infinite sequence of sets  $\mathcal{B}_0, \mathcal{B}_1, \dots$  each consisting of finitely many disjoint open balls of equal radius  $r_i$  in  $Y$ . Let  $\mathcal{B}_0 = \{Y\}$  and let  $r_0 = 1/4$ . If  $i > 0$ , let  $r_i < r_{i-1}/4$  be a positive number small enough that for every ball  $B = B(x, r_{i-1}) \in \mathcal{B}_{i-1}$ , the sub-ball  $B(x, r_{i-1}/2)$  contains  $n_i = \lceil 2^{r_i-b} \rceil$  disjoint balls of radius  $r_i$ . Let  $\mathcal{B}_i(B)$  denote this collection of disjoint balls and let  $\mathcal{B}_i = \bigcup_{B \in \mathcal{B}_{i-1}} \mathcal{B}_i(B)$ . For each ball  $B = B(x, r) \in \mathcal{B}_i$ , define a “bump function” supported in  $B$  by

$$G_B(x) = \begin{cases} \min\{r - d(x, s_i), r/2\} & \text{if } x \in B \\ 0 & \text{otherwise.} \end{cases}$$

Let  $\mathcal{B}_\infty = \bigcup_{i=0}^\infty \mathcal{B}_i$ . Note that  $\mathcal{B}_\infty$  is analogous to the ball-tree defined in Section 3.

For a mapping  $\sigma : \mathcal{B}_\infty \rightarrow \{\pm 1\}$ , define a function  $\pi : X \rightarrow [0, 1]$  by

$$(9.14) \quad \pi(x) = \frac{1}{2} + \sum_{B \in \mathcal{B}_\infty} \sigma(B) G_B(x).$$

The sum converges absolutely because every  $x \in X$  belongs to at most ball in  $\mathcal{B}_i$  for each  $i$ , so the absolute value of the infinite sum in (9.14) is bounded above by  $\sum_{i=0}^\infty r_i < 1/3$ . Moreover, one can verify that our construction ensures that  $\pi(x)$  is a Lipschitz function of  $x$  with Lipschitz constant 1.

If  $\mathcal{Q}$  is any subset of  $\mathcal{B}_\infty$ , one can define a problem instance of the full-feedback Lipschitz experts problem: a distribution  $\mathbb{P}_\mathcal{Q}$  on payoff functions  $\pi$  by sampling  $\sigma(B) \in \{\pm 1\}$  uniformly at random for  $B \notin \mathcal{Q}$  and performing biased sampling of  $\sigma(B) \in \{\pm 1\}$  with  $E[\sigma(B)] = 1/3$  when  $B \in \mathcal{Q}$ , and then defining  $\pi$  using (9.14). Note that the distribution  $\mathbb{P}_\mathcal{Q}$  has expected payoff function  $\mu = \frac{1}{2} + \sum_{B \in \mathcal{Q}} G_B/3$ .

In proving the lower bound, we will consider the distribution  $\mathcal{P}$  on Lipschitz experts problem instances  $\mathbb{P}_\mathcal{Q}$  where  $\mathcal{Q}$  is a random subset of  $\mathcal{B}_\infty$  obtained by sampling one ball  $B_0 \in \mathcal{B}_0$  uniformly at random, and also sampling one element  $Q(B) \in \mathcal{B}_i(B)$  uniformly at random and independently for each  $B \in \mathcal{B}_{i-1}$ . By analogy with the notion of complete lineage defined in Section 3, we will refer to any such set  $\mathcal{Q}$  as a complete lineage in  $\mathcal{B}_\infty$ . Given a complete lineage  $\mathcal{Q}$ , we can define an infinite nested sequence of balls  $B_0 \supset B_1 \supset \dots$  by specifying that  $B_{i+1} = Q(B_i)$  for each  $i$ .

If  $\mu$  is the expectation of a random payoff function  $\pi$  sampled from  $\mathbb{P}_\mathcal{Q}$ , then  $\mu$  achieves its maximum value  $\frac{1}{2} + \frac{1}{6} \sum_{i=0}^\infty r_i$  at the unique point  $x^* \in \bigcap_{i=0}^\infty B_i$ . At any point  $x \notin B_j$ , we have

$$\mu(x^*) - \mu(x) \geq \left(\frac{1}{6} \sum_{i=j}^\infty r_i\right) - \left(\frac{1}{4} \sum_{i=j+1}^\infty r_i\right) = \frac{1}{6} r_j.$$

We now finish the lower bound proof as in the proof of Lemma 3.2. For each complete lineage  $\mathcal{Q}$  and ball

$B \in \mathcal{B}_{i-1}$ , let  $B^1, B^2, \dots, B^{n_i}$  be the elements of  $\mathcal{B}_i(B)$ . Consider the sets  $\mathcal{Q}_0 = \mathcal{Q} \setminus \mathcal{Q}(B)$  and  $\mathcal{Q}_j = \mathcal{Q}_0 \cup \{B^j\}$  for  $j = 1, 2, \dots, n_i$ . The distributions  $(\mathbb{P}_{\mathcal{Q}_0}, \mathbb{P}_{\mathcal{Q}_1}, \dots, \mathbb{P}_{\mathcal{Q}_{n_i}})$  constitute an  $(\epsilon, \delta, k)$ -ensemble for  $\epsilon = \frac{1}{6}rk$ ,  $\delta = \frac{1}{2}$ , and  $k = n_i$ . Consequently, for  $t_i = r_i^{-b}$ , the inequality  $t_i < \ln(17k)/2\delta^2$  holds, and we obtain a lower bound

$$R_{(\mathcal{A}, \mathbb{P}_{\mathcal{Q}_j})}(t_i) > \epsilon t_i/2 = \Omega(r_i^{1-b}) = \Omega(t_i^{(b-1)/b})$$

for at least half of the distributions  $\mathbb{P}_{\mathcal{Q}_j}$  in the ensemble. Recalling that  $\gamma < \frac{b-1}{b}$ , we see that the problem is not  $t^\gamma$ -tractable.

## 9.2 The MaxMinLCD upper bound: proofs for Theorem 9.1 and Theorem 9.2

First, let us incorporate the analysis from Section 8 via the following lemma.

**LEMMA 9.2.** *Consider an instance  $(X, d, \mathbb{P})$  of the full-feedback Lipschitz experts problem, and let  $x^* \in X$  be an optimal point. Fix subset  $U \subset X$  which contains  $x^*$ , and let  $b > \text{LCD}(U)$ . Then for any sufficiently large  $T$  and  $\delta = T^{-1/(b+2)}$  the following holds:*

- (a) *Let  $S$  be a  $\delta$ -hitting set for  $U$ ,  $|S| \leq N_\delta(U)$ . Consider the feedback of all points in  $S$  over  $T$  rounds; let  $x$  be the point in  $S$  with the largest sample average (break ties arbitrarily). Then*

$$\Pr[\mu(x^*) - \mu(x) < O(\delta \log T)] > 1 - T^{-2}.$$

- (b) *For the uniformly Lipschitz experts problem and  $b \geq 2$ , property (a) holds for  $\delta = T^{-1/b}$ .*

**Transfinite LCD decomposition.** We redefine the transfinite fat decomposition from [27] with respect to the log-covering dimension rather than the covering dimension.

**DEFINITION 9.1.** *Fix a metric space  $(X, d)$ . Let  $\beta$  denote an arbitrary ordinal. A transfinite LCD decomposition of depth  $\beta$  and dimension  $b$  is a transfinite sequence  $\{S_\lambda\}_{0 \leq \lambda \leq \beta}$  of closed subsets of  $X$  such that:*

- (a)  $S_0 = X$ ,  $S_\beta = \emptyset$ , and  $S_\nu \supseteq S_\lambda$  whenever  $\nu < \lambda$ .  
 (b) if  $V \subset X$  is closed, then the set  $\{\text{ordinals } \nu \leq \beta: V \text{ intersects } S_\nu\}$  has a maximum element.  
 (c) for any ordinal  $\lambda \leq \beta$  and any open set  $U \subset X$  containing  $S_{\lambda+1}$  we have  $\text{LCD}(S_\lambda \setminus U) \leq b$ .

The existence of suitable decompositions and the connection to MaxMinLCD is derived exactly as in Proposition 3.15 in [27].

**LEMMA 9.3.** *For every compact metric space  $(X, d)$ ,  $\text{MaxMinLCD}(X)$  is equal to the infimum of all  $b$  such that  $X$  has a transfinite LCD decomposition of dimension  $b$ .*

In what follows, let us fix metric space  $(X, d)$  and  $b > \text{MaxMinLCD}(X)$ , and let  $\{S_\lambda\}_{0 \leq \lambda \leq \beta}$  be a transfinite LCD decomposition of depth  $\beta$  and dimension  $b$ . For each  $x \in X$ , let the *depth* of  $x$  be the maximal ordinal  $\lambda$  such that  $x \in S_\lambda$ . (Such an ordinal exists by Definition 9.1(b).)

**Access to the metric space.** The algorithm requires two oracles: the *depth oracle*  $\text{Depth}(\cdot)$  and the *covering oracle*  $\text{Cover}(\cdot)$ . Both oracles input a finite collection  $\mathcal{F}$  of open balls  $B_0, B_1, \dots, B_n$ , given via the centers and the radii, and return a point in  $X$ . Let  $B$  be the union of these balls, and let  $\bar{B}$  be the closure of  $B$ . A call to oracle  $\text{Depth}(\mathcal{F})$  returns an arbitrary point  $x \in \bar{B} \cap S_\lambda$ , where  $\lambda$  is the maximum ordinal such that  $S_\lambda$  intersects  $\bar{B}$ . (Such an ordinal exists by Definition 9.1(b).) Given a point  $y^* \in X$  of depth  $\lambda$ , a call to oracle  $\text{Cover}(y^*, \mathcal{F})$  either reports that  $B$  covers  $S_\lambda$ , or it returns an arbitrary point  $x \in S_\lambda \setminus B$ . A call to  $\text{Cover}(\emptyset, \mathcal{F})$  is equivalent to the call  $\text{Cover}(y^*, \mathcal{F})$  for some  $y^* \in S_0$ .

The covering oracle will be used to construct  $\delta$ -nets as follows. First, using successive calls to  $\text{Cover}(\emptyset, \mathcal{F})$  one can construct a  $\delta$ -net for  $X$ . Second, given a point  $y^* \in X$  of depth  $\lambda$  and a collection of open balls whose union is  $B$ , using successive calls to  $\text{Cover}(y^*, \cdot)$  one can construct a  $\delta$ -net for  $S_\lambda \setminus B$ . The second usage is geared towards the scenario when  $S_{\lambda+1} \subseteq B$  and for some optimal strategy  $x^*$  we have  $x^* \in S_\lambda \setminus B$ . Then by Definition 9.1(c) we have  $\text{LCD}(S_\lambda \setminus B) < b$ , and one can apply Lemma 9.2.

**The algorithm.** Our algorithm proceeds in phases  $i = 1, 2, 3, \dots$  of  $2^i$  rounds each. Each phase  $i$  outputs two strategies:  $x_i^*, y_i^* \in X$  that we call the *best guess* and the *depth estimate*. Throughout phase  $i$ , the algorithm plays the best guess  $x_{i-1}^*$  from the previous phase. The depth estimate  $y_{i-1}^*$  is used “as if” its depth is equal to the depth of some optimal strategy. (We show that for a large enough  $i$  this is indeed the case with a very high probability.)

In the end of the phase, an algorithm selects a finite set  $A_i \subset X$  of *active points*, as described below. Once this set is chosen,  $x_i^*$  is defined simply as a point in  $A_i$  with the largest sample average of the feedback (breaking ties arbitrarily). It remains to define  $y_i^*$  and  $A_i$  itself.

Let  $T = 2^i$  be the phase duration. Using the covering oracle, the algorithm constructs (roughly) the finest  $r$ -net containing at most  $2^{\sqrt{T}}$  points. Specifically, the algorithm constructs  $2^{-j}$ -nets  $\mathcal{N}_j$ , for  $j = 0, 1, 2, \dots$ , until it finds the largest  $j$  such that  $\mathcal{N}_j$  contains at most  $2^{\sqrt{T}}$  points. Let  $r = 2^{-j}$  and  $\mathcal{N} = \mathcal{N}_j$ .

For each  $x \in X$ , let  $\mu_T(x)$  be the sample average of

the feedback during this phase. Let

$$\Delta_T(x) = \mu_T^* - \mu_T(x), \quad \text{where} \quad \mu_T^* = \max(\mu_T, \mathcal{N})$$

Define the depth estimate  $y_i^*$  to be the output of the oracle call  $\text{Depth}(\mathcal{F})$ , where

$$\mathcal{F} = \{B(x, r) : x \in \mathcal{N} \text{ and } \Delta_T(x) < r\}.$$

Finally, let us specify  $A_i$ . Let  $B$  be the union of balls

$$(9.15) \quad \{B(x, r) : x \in \mathcal{N} \text{ and } \Delta_T(x) > 2(r_T + r)\},$$

where  $r_T = \sqrt{8 \log(T|\mathcal{N}|)/T}$  is chosen so that by Chernoff Bounds for each  $x \in \mathcal{N}$  we have

$$(9.16) \quad \Pr[|\mu_T(x) - \mu(x)| < r_T] > 1 - (T|\mathcal{N}|)^{-3}.$$

Let  $\delta = T^{-1/b}$  for the uniformly Lipschitz experts problem, and  $\delta = T^{-1/(b+2)}$  otherwise. Let  $Q_T = 2^{\delta^{-b}}$  be the *quota* on the number of active points. Given a point  $y_{i-1}^*$  whose depth is (say)  $\lambda$ , algorithm uses the covering oracle to construct a  $\delta$ -net  $\mathcal{N}'$  for  $S_\lambda \setminus B$ . Define  $A_i$  as  $\mathcal{N}'$  or an arbitrary  $Q_T$ -point subset thereof, whichever is smaller.<sup>9</sup>

**(Very high-level) sketch of the analysis.** The proof roughly follows that of Theorem 3.16 in [27]. Call a phase *clean* if the event in (9.16) holds for all  $x \in \mathcal{N}_i$  and the appropriate version of this event holds for all  $x \in A_i$ . (The regret from phases which are not clean is negligible). On a very high level, the proof consists of two steps. First we show that for a sufficiently large  $i$ , if phase  $i$  is clean then the depth estimate  $y_i^*$  is correct, in the sense that it is indeed equal to the depth of some optimal strategy. The argument is similar to the one in Lemma 4.2. Second, we show that for a sufficiently large  $i$ , if the depth estimate  $y_{i-1}^*$  is “correct” (i.e. its depth is equal to that of some optimal strategy), and phase  $i$  is clean, then the “best guess”  $x_i^*$  is good, namely  $\mu(x_i^*)$  is within  $O(\delta \log T)$  of the optimum. The reason is that, letting  $\lambda$  be the depth of  $y_{i-1}^*$ , one can show that for a sufficiently large  $T$  the set  $B$  (defined in (9.15)) contains  $S_{\lambda+1}$  and does not contain some optimal strategy. By definition of the transfinite LCD decomposition we have  $\text{LCD}(S_\lambda \setminus U) < b$ , so in our construction the quota  $Q_T$  on the number of active points permits  $A_i$  to be a  $\delta$ -cover of  $S_\lambda \setminus U$ . Now we can use Lemma 9.2 to guarantee the “quality” of  $x_i^*$ . The final regret computation is similar to the one in the proof of Theorem 8.1.

## References

<sup>9</sup>The interesting case here is  $|\mathcal{N}'| \leq Q_T$ . If  $\mathcal{N}'$  contains too many points, the choice of  $A_i$  is not essential for the analysis.

- [1] Rajeev Agrawal. The continuum-armed bandit problem. *SIAM J. Control and Optimization*, 33(6):1926–1951, 1995.
- [2] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *J. Machine Learning Research*, 3:397–422, 2002. Preliminary version in *41st IEEE FOCS*, 2000.
- [3] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002. Preliminary version in *15th ICML*, 1998.
- [4] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, 32(1):48–77, 2002. Preliminary version in *36th IEEE FOCS*, 1995.
- [5] Peter Auer, Ronald Ortner, and Csaba Szepesvári. Improved Rates for the Stochastic Continuum-Armed Bandit Problem. In *20th Conference on Learning Theory (COLT)*, pages 454–468, 2007.
- [6] Baruch Awerbuch and Robert Kleinberg. Online linear optimization and adaptive routing. *Journal of Computer and System Sciences*, 74(1):97–114, February 2008. Preliminary version appeared in *36th ACM STOC*, 2004.
- [7] Jeffrey Banks and Rangarajan Sundaram. Denumerable-armed bandits. *Econometrica*, 60(5):1071–1096, 1992.
- [8] G. Cantor. Über unendliche, lineare Punktmannichfaltigkeiten, 4. *Mathematische Annalen*, 21:51–58, 1883. In G. Cantor, *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, Berlin: Teubner, 1932; reprinted in 1980; reprint ed., Hildesheim: Olms, 1966.
- [9] Nicolò Cesa-Bianchi, Yoav Freund, David Haussler, David P. Helmbold, Robert E. Schapire, and Manfred K. Warmuth. How to use expert advice. *J. ACM*, 44(3):427–485, 1997.
- [10] Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.
- [11] Eric Cope. Regret and convergence bounds for immediate-reward reinforcement learning with continuous action spaces, 2004. Unpublished manuscript.
- [12] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1991.
- [13] Varsha Dani and Thomas P. Hayes. Robbing the bandit: less regret in online geometric optimization against an adaptive adversary. In *17th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 937–943, 2006.
- [14] Varsha Dani, Thomas P. Hayes, and Sham Kakade. The Price of Bandit Information for Online Optimization. In *20th Advances in Neural Information Processing Systems (NIPS)*, 2007.
- [15] Abraham Flaxman, Adam Kalai, and H. Brendan McMahan. Online Convex Optimization in the Bandit Setting: Gradient Descent without a Gradient. In *16th*

- ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 385–394, 2005.
- [16] J. C. Gittins. *Multi-Armed Bandit Allocation Indices*. John Wiley & Sons, 1989.
- [17] J. C. Gittins and D. M. Jones. A dynamic allocation index for the sequential design of experiments. In J. Gani et al., editor, *Progress in Statistics*, pages 241–266. North-Holland, 1974.
- [18] Sudipta Guha and Kamesh Munagala. Approximation algorithms for partial-information based stochastic control with Markovian rewards. In *48th Symp. on Foundations of Computer Science (FOCS)*, pages 483–493, 2007.
- [19] Sudipta Guha, Kamesh Munagala, and Peng Shi. Approximation algorithms for restless bandit problems. In *20th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 28–37, 2009.
- [20] Anupam Gupta, Mike Dinitz, and Kanat Tangwongsan. Private communication, 2007.
- [21] Elad Hazan and Satyen Kale. Better algorithms for benign bandits. In *20th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 38–47, 2009.
- [22] Elad Hazan and Nimrod Megiddo. Online Learning with Prior Information. In *20th Conference on Learning Theory (COLT)*, pages 499–513, 2007.
- [23] Sham M. Kakade, Adam T. Kalai, and Katrina Ligett. Playing Games with Approximation Algorithms. In *39th ACM Symp. on Theory of Computing (STOC)*, 2007.
- [24] Robert Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *18th Advances in Neural Information Processing Systems (NIPS)*, 2004. Full version appeared in the author’s thesis (MIT, 1995).
- [25] Robert Kleinberg. *Online Decision Problems with Large Strategy Sets*. PhD thesis, MIT, Boston, MA, 2005.
- [26] Robert Kleinberg, Alexandru Niculescu-Mizil, and Yogeshwar Sharma. Regret bounds for sleeping experts and bandits. In *21st Conference on Learning Theory (COLT)*, pages 425–436, 2008.
- [27] Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-Armed Bandits in Metric Spaces. In *40th ACM Symp. on Theory of Computing (STOC)*, pages 681–690, 2008.
- [28] T.L. Lai and Herbert Robbins. Asymptotically efficient Adaptive Allocation Rules. *Advances in Applied Mathematics*, 6:4–22, 1985.
- [29] S. Mazurkiewicz and W. Sierpinski. Contribution à la topologie des ensembles dénombrables. *Fund. Math.*, 1:17–27, 1920.
- [30] H. Brendan McMahan and Avrim Blum. Online Geometric Optimization in the Bandit Setting Against an Adaptive Adversary. In *17th Conference on Learning Theory (COLT)*, pages 109–123, 2004.
- [31] Sandeep Pandey, Deepak Agarwal, Deepayan Chakrabarti, and Vanja Josifovski. Bandits for Taxonomies: A Model-based Approach. In *SIAM Intl. Conf. on Data Mining (SDM)*, 2007.
- [32] V. Vovk. A game of prediction with expert advice. *J. Computer and System Sciences*, 56(2):153–173, 1998.
- [33] P. Whittle. Restless bandits: Activity allocation in a changing world. *J. of Appl. Prob.*, 25A:287–298, 1988.

## A Reduction to compact metric spaces

In this section we reduce the Lipschitz MAB problem to that on complete metric spaces.

LEMMA A.1. *The Lipschitz MAB problem on a metric space  $(X, d)$  is  $f(t)$ -tractable if and only if it is  $f(t)$ -tractable on the completion of  $(X, d)$ . Likewise for the Lipschitz experts problem with double feedback.*

*Proof.* Let  $(X, d)$  be a metric space with completion  $(Y, d)$ . Since  $Y$  contain an isometric copy of  $X$ , we will abuse notation and consider  $X$  as a subset of  $Y$ . We will present the proof the Lipschitz MAB problem; for the experts problem with double feedback, the proof is similar.

Given an algorithm  $\mathcal{A}_X$  which is  $f(t)$ -tractable for  $(X, d)$ , we may use it as a Lipschitz MAB algorithm for  $(Y, d)$  as well. (The algorithm has the property that it never selects a point of  $Y \setminus X$ , but this doesn’t prevent us from using it when the metric space is  $(Y, d)$ .) The fact that  $X$  is dense in  $Y$  implies that for every Lipschitz payoff function  $\mu$  defined on  $Y$ , we have  $\sup(\mu, X) = \sup(\mu, Y)$ . From this, it follows immediately that the regret of  $\mathcal{A}_X$ , when considered a Lipschitz MAB algorithm for  $(X, d)$ , is the same as its regret when considered as a Lipschitz MAB algorithm for  $(Y, d)$ .

Conversely, given an algorithm  $\mathcal{A}_Y$  which is  $f(t)$ -tractable for  $(Y, d)$ , we may design a Lipschitz MAB algorithm  $\mathcal{A}_X$  for  $(X, d)$  by running  $\mathcal{A}_Y$  and perturbing its output slightly. Specifically, for each point  $y \in Y$  and each  $t \in \mathbb{N}$  we fix  $x = x(y, t) \in X$  such that  $d(x, y) < 2^{-t}$ . If  $\mathcal{A}_Y$  recommends playing strategy  $y_t \in Y$  at time  $t$ , algorithm  $\mathcal{A}_X$  instead plays  $x = x(y, t)$ . Let  $\pi$  be the observed payoff. Algorithm  $\mathcal{A}_X$  draws an independent 0-1 random sample with expectation  $\pi$ , and reports this sample to  $\mathcal{A}_Y$ . This completes the description of the modified algorithm  $\mathcal{A}_X$ .

Suppose  $\mathcal{A}_X$  is not  $f(t)$ -tractable. Then for some problem instance  $\mathcal{I}$  on  $(Y, d)$ , letting  $R_X(t)$  be the expected regret of  $\mathcal{A}_X$  on this instance, we have that  $\sup_{t \in \mathbb{N}} R_X(t)/f(t) = \infty$ . Let  $\mu$  be the expected payoff function in  $\mathcal{I}$ . Consider the following two problem instances of a MAB problem on  $Y$ , called  $\mathcal{I}_1$  and  $\mathcal{I}_2$ , in which if point  $y \in Y$  is played at time  $t$ , the payoff is an independent 0-1 random sample with expectation  $\mu(y)$  and  $\mu(x(y, t))$ , respectively. Note that algorithm  $\mathcal{A}_Y$  is  $f(t)$ -tractable on  $\mathcal{I}_1$ , and its behavior on  $\mathcal{I}_2$  is identical

to that of  $\mathcal{A}_X$  on the original problem instance  $\mathcal{I}$ . It follows that by observing the payoffs of  $\mathcal{A}_Y$  one can tell apart  $\mathcal{I}_1$  and  $\mathcal{I}_2$  with high probability. Specifically, there is a “classifier”  $\mathcal{C}$  which queries one point in each round, such that for infinitely many times  $t$  it tell apart  $\mathcal{I}_1$  and  $\mathcal{I}_2$  with success probability  $p(t) \rightarrow 1$ . Now, the latter is information-theoretically impossible.

To see this, let  $H_t$  be the  $t$ -round history of the algorithm (the sequence of points queried, and outputs received), and consider the distribution of  $H_t$  under problem instances  $\mathcal{I}_\infty$  and  $\mathcal{I}_\epsilon$  (call these distributions  $q_1$  and  $q_2$ ). Let us consider and look at their KL-divergence. By the chain rule (See Lemma B.1), we can show that  $KL(q_1, q_2) < \frac{1}{2}$ . (We omit the details.) It follows that letting  $S_t$  be the event that  $\mathcal{C}$  classifies the instance as  $\mathcal{I}_1$  after round  $t$ , we have  $\mathbb{P}_{q_1}[S_t] - \mathbb{P}_{q_2}[S_t] \leq KL(q_1, q_2) \leq \frac{1}{2}$ . For any large enough time  $t$ ,  $\mathbb{P}_{q_1}[S_t] < \frac{1}{4}$ , in which case  $\mathcal{C}$  makes a mistake (on  $\mathcal{I}_2$ ) with constant probability.

**LEMMA A.2.** *Consider The experts MAB problem with full feedback. If it is  $f(t)$ -tractable on a metric space  $(X, d)$  then it is  $f(t)$ -tractable on the completion of  $(X, d)$ .*

*Proof.* Identical to the easy (“only if”) direction of Lemma A.1.

*Remark.* Lower bounds only require Lemma A.2, or the easy (“only if”) direction of Lemma A.1. For the upper bounds (algorithmic results), we can either quote the “if” direction of Lemma A.1, or prove the desired property directly for the specific type algorithms that we use (which is much easier but less elegant).

## B KL-divergence techniques

Our proof will use the notion of Kullback-Leibler divergence (or *KL-divergence*), defined as follows.

**DEFINITION B.1.** *Let  $\Omega$  be a finite set with two probability measures  $p, q$ . Their Kullback-Leibler divergence, or *KL-divergence*, is the sum*

$$KL(p; q) = \sum_{x \in \Omega} p(x) \ln \left( \frac{p(x)}{q(x)} \right),$$

with the convention that  $p(x) \ln(p(x)/q(x))$  is interpreted to be 0 when  $p(x) = 0$  and  $+\infty$  when  $p(x) > 0$  and  $q(x) = 0$ . If  $Y$  is a random variable defined on  $\Omega$  and taking values in some set  $\Gamma$ , the conditional Kullback-Leibler divergence of  $p$  and  $q$  given  $Y$  is the sum

$$KL(p; q | Y) = \sum_{x \in \Omega} p(x) \ln \left( \frac{p(x | Y=Y(x))}{q(x | Y=Y(x))} \right),$$

where terms containing  $\log(0)$  or  $\log(\infty)$  are handled according to the same convention as above.

The definition can be applied to an infinite sample space  $\Omega$  provided that  $q$  is absolutely continuous with respect to  $p$ . For details, see [25], Chapter 2.7. The following lemma summarizes some standard facts about KL-divergence; for proofs, see [12, 25].

**LEMMA B.1.** *Let  $p, q$  be two probability measures on a measure space  $(\Omega, \mathcal{F})$  and let  $Y$  be a random variable defined on  $\Omega$  and taking values in some finite set  $\Gamma$ . Define a pair of probability measures  $p_Y, q_Y$  on  $\Gamma$  by specifying that  $p_Y(y) = p(Y = y)$ ,  $q_Y(y) = q(Y = y)$  for each  $y \in \Gamma$ . Then  $KL(p; q | Y) \geq 0$  and*

$$KL(p; q) = KL(p; q | Y) + KL(p_Y; q_Y).$$

An easy corollary is the following lemma which expresses the KL-divergence of two distributions on sequences as a sum of conditional KL-divergences.

**LEMMA B.2.** *Let  $\Omega$  be a sample space, and suppose  $p, q$  are two probability measures on  $\Omega^n$ , the set of  $n$ -tuples of elements of  $\Omega$ . For a sample point  $\vec{\omega} \in \Omega^n$ , let  $\omega^i$  denote its first  $i$  components. If  $p^i, q^i$  denote the probability measures induced on  $\Omega^i$  by  $p$  (resp.  $q$ ) then*

$$KL(p; q) = \sum_{i=1}^n KL(p^i; q^i | \omega^{i-1}).$$

*Proof.* For  $m = 1, 2, \dots, n$ , the formula  $KL(p^m; q^m) = \sum_{i=1}^m KL(p^i; q^i | \omega^{i-1})$  follows by induction on  $m$ , using Lemma B.1.

The following three lemmas will also be useful in our lower bound argument. Here and henceforth we will use the following notational convention: for real numbers  $a, b \in [0, 1]$ ,  $KL(a; b)$  denotes the KL-divergence  $KL(p; q)$  where  $p, q$  are probability measures on  $\{0, 1\}$  such that  $p(\{1\}) = a$ ,  $q(\{1\}) = b$ . In other words,

$$KL(a; b) = a \ln \left( \frac{a}{b} \right) + (1 - a) \ln \left( \frac{1-a}{1-b} \right).$$

**LEMMA B.3.** *For any  $0 < \epsilon < y \leq 1$ ,*

$$KL(y - \epsilon; y) < \epsilon^2 / y(1 - y).$$

*Proof.* A calculation using the inequality  $\ln(1 + x) < x$  (valid for  $x > 0$ ) yields

$$\begin{aligned} KL(y - \epsilon; y) &= (y - \epsilon) \ln \left( \frac{y - \epsilon}{y} \right) + (1 - y + \epsilon) \ln \left( \frac{1 - y + \epsilon}{1 - y} \right) \\ &< (y - \epsilon) \left( \frac{y - \epsilon}{y} - 1 \right) + (1 - y + \epsilon) \left( \frac{1 - y + \epsilon}{1 - y} - 1 \right) \\ &= \frac{-\epsilon(y - \epsilon)}{y} + \frac{\epsilon(1 - y + \epsilon)}{1 - y} = \frac{\epsilon^2}{y(1 - y)}. \end{aligned}$$

**LEMMA B.4.** *Let  $\Omega$  be a sample space with two probability measures  $p, q$  whose KL-divergence is  $\kappa$ . For any event  $\mathcal{E}$ , the probabilities  $p(\mathcal{E})$ ,  $q(\mathcal{E})$  satisfy*

$$q(\mathcal{E}) \geq p(\mathcal{E}) \exp \left( -\frac{\kappa + 1/e}{p(\mathcal{E})} \right).$$

A consequence of the lemma, stated in less quantitative terms, is the following: if  $\kappa = KL(p; q)$  is bounded above and  $p(\mathcal{E})$  is bounded away from zero then  $q(\mathcal{E})$  is bounded away from zero.

*Proof.* Let  $a = p(\mathcal{E})$ ,  $b = q(\mathcal{E})$ ,  $c = (1 - a)/(1 - b)$ . Applying Lemma B.1 with  $Y$  as the indicator random variable of  $\mathcal{E}$  we obtain

$$\begin{aligned} \kappa &= KL(p; q) \geq KL(p_Y; q_Y) \\ &= a \ln\left(\frac{a}{b}\right) + (1 - a) \ln\left(\frac{1-a}{1-b}\right) \\ &= a \ln\left(\frac{a}{b}\right) + (1 - b) c \ln(c). \end{aligned}$$

Now using the inequality  $c \ln(c) \geq -1/e$ , (valid for all  $c \geq 0$ ) we obtain

$$\kappa \geq a \ln(a/b) - (1 - b)/e \geq a \ln(a/b) - 1/e.$$

The lemma follows by rearranging terms.

LEMMA B.5. *Let  $p, q$  be two probability measures, and suppose that for some  $\delta \in (0, \frac{1}{2}]$  they satisfy*

$$\forall \text{ events } \mathcal{E}, \quad 1 - \delta < \frac{q(\mathcal{E})}{p(\mathcal{E})} < 1 + \delta$$

*Then  $KL(p; q) < \delta^2$ .*

*Proof.* We will prove the lemma assuming the sample space is finite. The result for general measure spaces follows by taking a supremum.

For every  $x$  in the sample space  $\Omega$ , let  $r(x) = \frac{q(x)}{p(x)} - 1$  and note that  $|r(x)| < \delta$  for all  $x$ . Now we make use of the inequality  $\ln(1 + x) \leq x - x^2$ , valid for  $x \geq -\frac{1}{2}$ .

$$\begin{aligned} KL(p; q) &= \sum_x p(x) \ln\left(\frac{p(x)}{q(x)}\right) = \sum_x p(x) \ln\left(\frac{1}{1+r(x)}\right) \\ &= -\sum_x p(x) \ln(1 + r(x)) \\ &\leq -\sum_x p(x) [r(x) - (r(x))^2] \\ &< -(\sum_x p(x)r(x)) + \delta^2 (\sum_x p(x)) \\ &= -(\sum_x q(x) - p(x)) + \delta^2 = \delta^2. \end{aligned}$$

*Proof of Theorem 3.2:* Let  $\Omega = [0, 1]^X$ . Using Property 1 of an  $(\epsilon, \delta, k)$ -ensemble combined with Lemma B.5, we find that  $KL(\mathbb{P}_i; \mathbb{P}_0) < \delta^2$ .

Let  $\mathcal{A}$  be an experts algorithm whose random bits are drawn from a sample space  $\Gamma$  with probability measure  $\nu$ . For any positive integer  $s < \ln(17k)/2\delta^2$ , let  $p_i^s$  denote the measure  $\nu \times (\mathbb{P}_i)^s$  on the probability space  $\Gamma \times \Omega^s$ . By the chain rule for KL-divergence (Lemma B.2),  $KL(p_i^s; p_0^s) < s\delta^2 < \ln(17k)/2$ . Now let  $\mathcal{E}_i^s$  denote the event that  $\mathcal{A}$  selects a point  $x \in S_i$  at time  $s$ . If  $p_i^s(\mathcal{E}_i^s) \geq \frac{1}{2}$  then Lemma B.4 implies

$$\begin{aligned} p_0^s(\mathcal{E}_i^s) &\geq p_i^s(\mathcal{E}_i^s) \exp\left(-\frac{\ln(17k)/2 + 1/e}{p_i^s(\mathcal{E}_i^s)}\right) \\ &\geq \frac{1}{2} \exp\left(-\ln(k) + \ln(17) - \frac{2}{e}\right) > \frac{4}{k}. \end{aligned}$$

The events  $\{\mathcal{E}_i^s \mid 1 \leq i \leq k\}$  are mutually exclusive, so fewer than  $k/4$  of them can satisfy  $p_0^s(\mathcal{E}_i^s) > \frac{4}{k}$ . Consequently, fewer than  $k/4$  of them can satisfy  $p_i^s(\mathcal{E}_i^s) \geq \frac{1}{2}$ , a property we denote in this proof by saying that  $s$  is *satisfactory* for  $i$ . Now assume  $t < \ln(17k)/2\delta^2$ . For a uniformly random  $i \in \{1, \dots, k\}$ , the expected number of satisfactory  $s \in \{1, \dots, t\}$  is less than  $t/4$ , so by Markov's inequality, for at least half of the  $i \in \{1, \dots, k\}$ , the number of satisfactory  $s \in \{1, \dots, t\}$  is less than  $t/2$ . Property 2 of an  $(\epsilon, \delta, k)$ -ensemble guarantees that every unsatisfactory  $s$  contributes at least  $\epsilon$  to the regret of  $\mathcal{A}$  when the problem instance is  $\mathbb{P}_i$ . Therefore, at least half of the measures  $\mathbb{P}_i$  have the property that  $R_{(\mathcal{A}, \mathbb{P}_i)}(t) \geq \epsilon t/2$ .

**B.1 Proof of Claim 5.1** Recall that in Section 5 we defined a pair of payoff functions  $\mu_0, \mu_i$  and a ball  $B_i$  of radius  $r_i$  such that  $\mu_0 \equiv \mu_i$  on  $X \setminus B_i$ , while for  $x \in B_i$  we have  $\frac{3}{8} \leq \mu_0(x) \leq \mu_i(x) \leq \mu_0(x) + \frac{r_i}{4} \leq \frac{3}{4}$ . Thus, by Lemma B.3,  $KL(\mu_0(x); \mu_i(x)) < r_i^2/3$  for all  $x \in X$ , and  $KL(\mu_0(x); \mu_i(x)) = 0$  for  $x \notin B_i$ .

Represent the algorithm's choice and the payoff observed at any given time  $t$  by a pair  $(x_t, y_t)$ . Let  $\Omega = X \times [0, 1]$  denote the set of all such pairs. When a given algorithm  $\mathcal{A}$  plays against payoff functions  $\mu_0, \mu_i$ , this defines two different probability measures  $p_0^t, p_i^t$  on the set  $\Omega^t$  of possible  $t$ -step histories. Let  $\omega^t$  denote a sample point in  $\Omega^t$ . The bounds derived in the previous paragraph imply that for any non-negative integer  $s$ ,

$$(B.1) \quad KL(p_0^{s+1}; p_i^{s+1} \mid \omega^s) < \frac{1}{3} r_i^2 \mathbb{P}_0(x_{s+1} \in B_i).$$

Summing equation (B.1) for  $s = 0, 1, \dots, t - 1$  and applying Lemma B.2 we obtain

$$(B.2) \quad \begin{aligned} KL(p_0^t; p_i^t) &< \frac{1}{3} r_i^2 \sum_{s=1}^t \mathbb{P}_0(x_s \in B_i) \\ &= \frac{1}{3} r_i^2 \mathbb{E}_0(N_i(t)), \end{aligned}$$

where the last equation follows from the definition of  $N_i(t)$  as the number of times algorithm  $\mathcal{A}$  selects a strategy in  $B_i$  during the first  $t$  rounds.

The bound in Claim 5.1 now follows by applying Lemma B.4 with the event  $S$  playing the role of  $\mathcal{E}$ ,  $\mathbb{P}_0$  playing the role of  $p$ , and  $\mathbb{P}_i$  playing the role of  $q$ .