Bandits and Experts in Metric Spaces

ROBERT KLEINBERG, Computer Science Department, Cornell University, USA
ALEKSANDRS SLIVKINS, Microsoft Research NYC, USA
ELI UPFAL, Computer Science Department, Brown University, USA

In a multi-armed bandit problem, an online algorithm chooses from a set of strategies in a sequence of trials to maximize the total payoff of the chosen strategies. While the performance of bandit algorithms with a small finite strategy set is well understood, bandit problems with large strategy sets are still a topic of active investigation, motivated by practical applications, such as online auctions and web advertisement. The goal of such research is to identify broad and natural classes of strategy sets and payoff functions that enable the design of efficient solutions.

In this work, we study a general setting for the multi-armed bandit problem, in which the strategies form a metric space, and the payoff function satisfies a Lipschitz condition with respect to the metric. We refer to this problem as the *Lipschitz MAB problem*. We present a solution for the multi-armed bandit problem in this setting. That is, for every metric space, we define an isometry invariant that bounds from below the performance of Lipschitz MAB algorithms for this metric space, and we present an algorithm that comes arbitrarily close to meeting this bound. Furthermore, our technique gives even better results for benign payoff functions. We also address the full-feedback ("best expert") version of the problem, where after every round the payoffs from all arms are revealed.

CCS Concepts: • **Theory of computation** → **Online learning algorithms**; **Online learning theory**; **Regret bounds**;

Additional Key Words and Phrases: Multi-armed bandits, regret, online learning, metric spaces, covering dimension, Lipschitz-continuity

ACM Reference format:
Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. 2019. Bandits and Experts in Metric Spaces. *J. ACM* 66, 4, Article 30 (May 2019), 77 pages.
https://doi.org/10.1145/3299873

This is a merged and definitive version of Kleinberg, Slivkins, and Upfal [64] and Kleinberg and Slivkins [63]. Compared to the conference publications, it contains full proofs and a significantly revised presentation. In particular, it develops new terminology and modifies the proof outlines to unify the technical exposition of the two papers. The paper also features an updated discussion of the follow-up work and open questions. All results on the zooming algorithm and the max-min-covering dimension are from Kleinberg et al. [64]; all results on regret dichotomies and on Lipschitz experts are from Kleinberg and Slivkins [63].

Supported in part by NSF Awards No. CCF-0643934 and No. CCF-0729102. Some of the work was done while the author was visiting Microsoft Research.
Some of the work was done while the author was a postdoctoral research associate at Brown University.
Supported in part by NSF Awards No. CCR-0121154 and No. DMI-0600384, and ONR Award No. N000140610607.

Authors' addresses: R. Kleinberg, Department of Computer Science, Cornell University, 317 Gates Hall, Ithaca, NY 14853, USA; email: rdk@cs.cornell.edu; A. Slivkins, Microsoft Research NYC, 641 6th Ave, 7th floor, New York, NY 10011, USA; email: slivkins@microsoft.com; E. Upfal, Brown University, P.O. Box 1910, Providence, RI 02912, USA; email: eli@cs.brown.edu.

0004-5411/2019/05-ART30 $15.00
https://doi.org/10.1145/3299873

1 INTRODUCTION

In a multi-armed bandit problem, an online algorithm must iteratively choose from a set of possible strategies (also called "arms") in a sequence of n trials to maximize the total payoff of the chosen strategies. These problems are the principal theoretical tool for modeling the exploration/exploitation tradeoffs inherent in sequential decision-making under uncertainty. Studied intensively for decades [22, 27, 35, 49, 84, 97], bandit problems are having an increasingly visible impact on computer science because of their diverse applications, including online auctions, adaptive routing, and the theory of learning in games. The performance of a multi-armed bandit algorithm is often evaluated in terms of its *regret*, defined as the gap between the expected payoff of the algorithm and that of an optimal strategy. While the performance of bandit algorithms with a small finite strategy set is well understood, bandit problems with exponentially or infinitely large strategy sets are still a topic of active investigation.

Absent any assumptions about the strategies and their payoffs, bandit problems with large strategy sets allow for no non-trivial solutions—any multi-armed bandit algorithm performs as badly, on some inputs, as random guessing. But in most applications it is natural to assume a structured class of payoff functions, which often enables the design of efficient learning algorithms [61]. In this article, we consider a broad and natural class of problems in which the structure is induced by a metric on the space of strategies. While bandit problems have been studied in a few specific metric spaces (such as a one-dimensional interval) [4, 14, 38, 60, 81], the case of general metric spaces has not been treated before, despite being very natural for bandit problems.

As a motivating example, consider the problem faced by a website choosing from a database of thousands of banner ads to display to users, with the aim of maximizing the click-through rate of the ads displayed by matching ads to users' characterizations and the web content that they are currently watching. Independently experimenting with each advertisement is infeasible, or at least highly inefficient, since the number of ads is too large. Instead, the advertisements are usually organized into a taxonomy based on metadata (such as the category of product being advertised), which allows a similarity measure to be defined. The website can then attempt to optimize its learning algorithm by generalizing from experiments with one ad to make inferences about the performance of similar ads [81, 82].

Another motivating example is revenue-management problems (e.g., see References [24, 66]). Consider a monopolistic seller with unlimited inventory of many digital products, such as songs, movies, or software. Customers arrive over time, and the seller can give customized offers to each arriving customer to maximize the revenue. The space of possible offers is very large, both in terms of possible product bundles and in terms of the possible prices, so experimenting with each and every offer is inefficient. Instead, the seller may be able to use experiments with one offer to make inferences about similar offers.

Abstractly, we have a bandit problem of the following form: there is a strategy set X, with an unknown payoff function $\mu : X \to [0,1]$ satisfying a set of predefined constraints of the form $|\mu(x) - \mu(y)| \leq \delta(x,y)$ for some $x, y \in X$ and $\delta(x,y) > 0$. In each period the algorithm chooses a point $x \in X$ and receives payoff—a number in the $[0,1]$ interval—sampled independently from some distribution $\mathbb{P}_x$ whose expectation is $\mu(x)$.

A moment's thought reveals that this abstract problem can be regarded as a bandit problem in a metric space. Specifically, define $\mathcal{D}(x,y)$ to be the infimum of the quantity $\sum_i \delta(x_i, x_{i+1})$ over all finite paths $(x = x_0, x_1, \ldots, x_k = y)$ in X. Then $\mathcal{D}$ is a metric and the constraints $|\mu(x) - \mu(y)| < \delta(x,y)$ may be summarized (equivalently reformulated) as follows:

$$|\mu(x) - \mu(y)| \leq \mathcal{D}(x,y) \quad \text{for all } x, y \in X. \tag{1}$$

In other words, μ is a Lipschitz function (of Lipschitz constant 1) on the metric space $(X, \mathcal{D})$.

We assume that an algorithm is given the metric space $(X, \mathcal{D})$ as an input, with a promise that the payoff function μ satisfies Equation (1). We refer to this problem as the *Lipschitz MAB problem* on $(X, \mathcal{D})$, and we refer to the ordered triple $(X, \mathcal{D}, \mu)$ as an *instance* of the Lipschitz MAB problem.[1]

1.1 Prior Work

While our work is the first to treat the Lipschitz MAB problem in general metric spaces, special cases of the problem are implicit in prior work on the continuum-armed bandit problem [4, 14, 38, 60]—which corresponds to the space $[0, 1]$ under the metric $\ell_1^{1/d}$, $d \geq 1$—and the experimental work on "bandits for taxonomies" [81], which corresponds to the case in which $(X, \mathcal{D})$ is a tree metric.[2] Also, Hazan and Megiddo [53] considered a contextual bandit setting with a metric space on contexts rather than arms.

Before describing our results in greater detail, it is helpful to put them in context by recounting the nearly optimal bounds for the one-dimensional continuum-armed bandit problem, a problem first formulated in Agrawal [4] and solved (up to logarithmic factors) by various authors [14, 38, 60]. In the following theorem and throughout this article, the *regret* of a multi-armed bandit algorithm $\mathcal{A}$ running on an instance $(X, \mathcal{D}, \mu)$ is defined to be the function $R_{\mathcal{A}}(t)$, which measures the difference between its expected payoff at time t and the quantity $t \cdot \sup_{x \in X} \mu(x)$. The latter quantity is the expected payoff of always playing an arm $x \in \text{argmax}\mu(x)$ if such arm exists. In regret-minimization, the main issue is typically how regret scales with t.

Theorem 1.1 ([14, 38, 60][3]). *For any $d \geq \mathbb{N}$, consider the Lipschitz MAB problem on $([0, 1], \ell_1^{1/d})$, $d \geq 1$. There is an algorithm whose regret on any instance μ satisfies $R(t) = \tilde{O}(t^\gamma)$ for every t, where $\gamma = \frac{d+1}{d+2}$. No such algorithm exists for any $\gamma < \frac{d+1}{d+2}$.*

In fact, if the time horizon t is known in advance, the upper bound in the theorem can be achieved by an extremely naïve algorithm, which uses an optimal k-armed bandit algorithm, such as the `UCB1` algorithm [11], to choose arms from the set $S = \{0, \frac{1}{k}, \frac{2}{k}, \ldots, 1\}$, for a suitable choice of the parameter k. Here the arms in S partition the strategy set in a uniform (and non-adaptive) way; hence, we call this algorithm `UniformMesh`.

1.2 Initial Result

We make an initial observation that the analysis of algorithm `UniformMesh` in Theorem 1.1 only relies on the covering properties of the metric space, rather than on its real-valued structure, and (with minor modifications) can be extended to any metric space of constant covering dimension d.

Theorem 1.2. *Consider the Lipschitz MAB problem on a metric space of covering dimension $d \geq 0$. There is an algorithm whose regret on any instance μ satisfies $R(t) = \tilde{O}(t^\gamma)$ for every t, where $\gamma = \frac{d+1}{d+2}$.*

The covering dimension is a standard notion that summarizes covering properties of a metric space. It is defined as the smallest (infimum) number $d \geq 0$ such that X can be covered by $O(\delta^{-d})$ sets of diameter δ, for each $\delta > 0$. We denote it $\mathtt{COV}(X, \mathcal{D})$, or $\mathtt{COV}(X)$ when the metric $\mathcal{D}$ is clear

[1] Formally, the problem instance also includes the parameterized family of reward distributions $\mathbb{P}_x$. To simplify exposition, we assume this family is fixed throughout, and therefore can be suppressed from the notation. When the metric space $(X, \mathcal{D})$ is understood from context, we may also refer to μ as an instance.

[2] Throughout the article, ℓ_p, $p \geq 1$ denotes a metric on a finite-dimensional real space given by $\ell_p(x, y) = \|x - y\|_p$.

[3] Auer et al. [14] and Cope [38] also achieve $\tilde{O}(\sqrt{T})$ regret under additional assumptions on the shape of the function near its optimum.

from the context. The covering dimension generalizes the Euclidean dimension, in the sense that the covering dimension of $([0,1]^d, \ell_p)$, $p \geq 1$ is d. Unlike the Euclidean dimension, the covering dimension can take fractional values. Theorem 1.2 generalizes the upper bound in Theorem 1.1, because the covering dimension of $([0,1], \ell_1^{1/d})$ is d, for any $d \geq 1$.

1.3 Present Scope

This article is a comprehensive study of the Lipschitz MAB problem in arbitrary metric spaces.

While the regret bound in Theorem 1.1 is essentially optimal when the metric space is $([0,1], \ell_1^{1/d})$, it is strikingly odd that it is achieved by such a simple algorithm as `UniformMesh`. In particular, the algorithm approximates the strategy set by a fixed mesh S and does not refine this mesh as it gains information about the location of the optimal arm. Moreover, the metric contains seemingly useful proximity information, but the algorithm ignores this information after choosing its initial mesh. Is this really the best algorithm?

A closer examination of the lower bound proof raises further reasons for suspicion: it is based on a contrived, highly singular payoff function μ that alternates between being constant on some distance scales and being very steep on other (much smaller) distance scales, to create a multi-scale "needle in haystack" phenomenon, which nearly obliterates the usefulness of the proximity information contained in the metric $\ell_1^{1/d}$. Can we expect algorithms to do better when the payoff function is more benign?[4] For the Lipschitz MAB problem on $([0,1], \ell_1)$, the question was answered affirmatively in Auer et al. [14], Cope [38] for some classes of instances, with algorithms that are tuned to the specific classes.

We are concerned with the following two directions motivated by the discussion above:

(Q1) *Per-metric optimality.* What is the best possible bound on regret for a given metric space? (Implicitly, such regret bound is worst-case over all payoff functions consistent with this metric space.) Is `UniformMesh`, as naive as it is, really an optimal algorithm? Is covering dimension an appropriate structure to characterize such worst-case regret bounds?

(Q2) *Benign problem instances.* Is it possible to take advantage of benign payoff functions? What structures would be useful to characterize benign payoff functions and the corresponding better-than-worst-case regret bounds? What algorithmic techniques would help?

Theorem 1.2 calibrates our intuition: For relatively "rich" metric spaces such as $([0,1], \ell_1^{1/d})$, we expect regret bounds of the form $\tilde{O}(t^\gamma)$, for some constant $\gamma \in (0,1)$, which depends on the metric space, and perhaps also on the problem instance. Henceforth, we will call this *polynomial regret.* Apart from metric spaces that admit polynomial regret, we are interested in the extremes: metric spaces for which the Lipschitz MAB problem becomes very easy or very difficult.

It is known that one can achieve logarithmic regret as long as the number of arms is finite [11, 69].[5] However, all prior results for infinite metric spaces had regret $O(t^\gamma)$, $\gamma \geq \frac{1}{2}$. We view problem instances with $O(\log t)$ regret as "very tractable," and those with regret t^γ, $\gamma \geq \frac{1}{2}$, as "somewhat tractable." It is natural to ask what is the transition between the two.

(Q3) 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 $\tilde{O}(\sqrt{t})$?

[4] Here and elsewhere, we use "benign" as a non-technical term.

[5] The constant in front of the $\log(t)$ increases with the number of arms and also depends on instance-specific parameters. $O(\log t)$ regret is optimal even for two arms [69].

On the opposite end of the "tractability spectrum" of the Lipschitz MAB problem, there are metric spaces of infinite covering dimension, for which no algorithm can have regret of the form $O(t^\gamma)$, $\gamma < 1$. Intuitively, such metric spaces are intractable. Formally, will define "intractable" metric spaces is those that do not admit sub-linear regret.

(Q4) Which metric spaces are tractable, i.e., admit $o(t)$ regret?

We are also interested in the full-feedback version of the Lipschitz MAB problem, where after each round the payoff for each arm can be queried by the algorithm. Such settings have been extensively studied in the online learning literature under the name *best experts problems* [34, 35, 100]. Accordingly, we call our setting the *Lipschitz experts problem.* To the best of our knowledge, prior work for this setting includes the following two results: constant regret for a finite set of arms [62], and $\tilde{O}(\sqrt{t})$ regret for metric spaces of bounded covering dimension [50]; the latter result uses a version of the `UniformMesh`. We are interested in per-metric optimality (Q1), including the extreme versions (Q3) and (Q4). For polynomial regret the goal is to handle metric spaces of infinite covering dimension.

1.4 Our Contributions: Lipschitz MAB Problem

We give a complete solution to (Q1), by describing for every metric space $(X, \mathcal{D})$ a family of algorithms that come arbitrarily close to achieving the best possible regret bound for this metric space. In particular, we resolve (Q3) and (Q4). We also give a satisfactory answer to (Q2); our solution is arbitrarily close to optimal in terms of the zooming dimension defined below.

Underpinning these contributions is a new algorithm, called the *zooming algorithm.* It maintains a mesh of "active arms," but (unlike `UniformMesh`) it adapts this mesh to the observed payoffs. It combines the upper confidence bound technique used in earlier bandit algorithms such as `UCB1` [11] with a novel *adaptive refinement* step that uses past history to refine the mesh ("zoom in") in regions with high observed payoffs. We show that the zooming algorithm can perform significantly better on benign problem instances. Moreover, it is a key ingredient in our design of a per-metric optimal bandit algorithm.

Benign problem instances. For every problem instance $(X, \mathcal{D}, \mu)$, we define a parameter called the *zooming dimension*, and use it to bound the performance of the zooming algorithm in a way that is often significantly stronger than the corresponding per-metric bound. Note that the zooming algorithm is *self-tuning*, i.e., it achieves this bound without requiring prior knowledge of the zooming dimension. Somewhat surprisingly, our regret bound for the zooming algorithm result has exactly the same "shape" as Theorem 1.2.

Theorem 1.3. *If d is the zooming dimension of a Lipschitz MAB instance, then at any time t the zooming algorithm suffers regret $\tilde{O}(t^\gamma)$, where $\gamma = \frac{d+1}{d+2}$.*

The exponent γ in the theorem is the best possible, as a function of d, in light of Theorem 1.1.

While covering dimension is about covering the entire metric space, zooming dimension focuses on covering near-optimal arms. The lower bounds in Theorems 1.1 and 1.5 are based on contrived examples with a high-dimensional set of near-optimal arms, which leads to the "needle-in-the-haystack" phenomenon. We sidestep these examples if the set of near-optimal arms is low-dimensional, in the sense that we make formal below. We define the *zooming dimension* of an instance $(X, \mathcal{D}, \mu)$ as the smallest d such that the following covering property holds: For every $\delta > 0$, we require only $O(\delta^{-d})$ sets of diameter $\delta/8$ to cover the set of arms whose expected payoff falls short of the optimum by an amount between δ and 2δ.

Zooming dimension is our way to quantify the benignness of a problem instance. It is trivially no larger than the covering dimension, and can be significantly smaller. Below let us give some examples:

- Suppose a low-dimensional region $S \subset X$ contains all arms with optimal or near-optimal payoffs. The zooming dimension of such problem instance is bounded from above by the covering dimension of S. For example, S can be a "thin" subtree of an infinitely deep tree.[6]
- Suppose the metric space is $([0,1], \ell_1^{1/d})$, $d \in \mathbb{N}$, and the expected payoff of each arm x is determined by its distance from the best arm: $\mu(x) = \max(0,\ \mu^* - \mathcal{D}(x, x^*))$ for some number $\mu^* \in (0,1]$ and some arm x^*. Then the zooming dimension is 0, whereas the covering dimension is d.
- Suppose the metric space is $([0,1]^d, \ell_2)$, $d \in \mathbb{N}$, and payoff function μ is C^2-smooth. Assume μ has a unique maximum x^* and is strongly concave in a neighborhood of x^*. Then the zooming dimension is $d/2$, whereas the covering dimension is d.

It turns out that the analysis of the zooming algorithm does not require the similarity function $\mathcal{D}$ to satisfy the triangle inequality, and needs only a relaxed version of the Lipschitz condition Equation (1).

Theorem 1.4 (Informal). *The upper bound in Theorem 1.3 holds in a more general setting where the similarity function $\mathcal{D}$ does not satisfy the triangle inequality, and Lipschitz condition Equation (1) is relaxed to hold only if one of the two arms is optimal.*

In addition to the two theorems above, we apply the zooming algorithm to the following special cases, deriving improved or otherwise non-trivial regret bounds: (i) the maximal payoff is near 1, (ii) $\mu(x) = 1 - f(\mathcal{D}(x, S))$, where S is a "target set" that is not revealed to the algorithm, (iii) the reward from playing each arm x is $\mu(x)$ plus an independent, benignly distributed noise.

In particular, we obtain an improved regret rate if the rewards are deterministic. This corollary is related to the literature on global Lipschitz optimization (e.g., see Floudas [45]), and extends this literature by relaxing the Lipschitz assumption as in Theorem 1.4.

While our definitions and results so far have been tailored to infinite strategy sets, they can be extended to the finite case as well. We use a more precise, *non-asymptotic* version of the zooming dimension, so that all results on the zooming algorithm are meaningful for both finite and infinite strategy sets.

Per-metric optimality: full characterization. We are interested in *per-metric optimal* regret bounds: best possible regret bounds for a given metric space. We prove several theorems, which jointly provide a full characterization of per-metric optimal regret bounds for any given metric space $(X, \mathcal{D})$. To state polynomial regret bounds in this characterization, we define a parameter of the metric space called *max-min-covering dimension* (`MaxMinCOV`). Our characterization is summarized in the table below.

Table 1 should be interpreted as follows. We consider regret bounds with an instance-dependent constant, i.e., those of the form $R(t) \leq C_{\mathcal{I}}\, f(t)$, for some function $f : \mathbb{N} \to \mathbb{R}$ and a constant $C_{\mathcal{I}}$ that can depend on the problem instance $\mathcal{I}$; we denote this as $R(t) = O_{\mathcal{I}}(f(t))$. Let us say that the Lipschitz MAB problem on a given metric space is $f(t)$-tractable if there exists an algorithm whose regret satisfies $R(t) = O_{\mathcal{I}}(f(t))$. Then, Lai and Robbins [69] and Auer et al. [11] show that

[6] Consider an infinitely deep rooted tree and let arms correspond to ends of the tree (i.e., infinite paths away from the root). The distance between two ends decreases exponentially in the height of their least common ancestor (i.e., the deepest vertex belonging to both paths). Suppose there is a subtree in which the branching factor is smaller than elsewhere in the tree. Then, we can take S to be the set of ends of this subtree.

Table 1. Per-metric Optimal Regret Bounds for Lipschitz MAB

If the metric completion of $(X, \mathcal{D})$ is …	then regret can be …	but not …
finite	$O(\log t)$	$o(\log t)$
compact and countable	$\omega(\log t)$	$O(\log t)$
compact and uncountable		
$\texttt{MaxMinCOV} = 0$	$\tilde{O}(t^\gamma), \gamma > \frac{1}{2}$	$o(\sqrt{t})$
$\texttt{MaxMinCOV} = d \in (0, \infty)$	$\tilde{O}(t^\gamma), \gamma > \frac{d+1}{d+2}$	$o(t^\gamma), \gamma < \frac{d+1}{d+2}$
$\texttt{MaxMinCOV} = \infty$	$o(t)$	$O(t^\gamma), \gamma < 1$
non-compact	$O(t)$	$o(t)$

the problem is $\log(t)$-tractable if the metric space has finitely many points (here the instance-dependent constant $C_{\mathcal{I}}$ is essential), and not $f(t)$-tractable for any $f(t) = o(\log t)$. Thus, the first row of Table 1 reads $O(\log t)$ and $o(\log t)$, respectively; other rows should be interpreted similarly.

In what follows, we discuss the individual results that comprise the characterization in Table 1.

Per-metric optimality: polynomial regret bounds. The definition of the max-min-covering dimension arises naturally as one tries to extend the lower bound from Kleinberg [60] to general metric spaces. The min-covering dimension of a subset $Y \subset X$ is the smallest covering dimension of any non-empty subset $U \subset Y$, which is open in the metric topology of $(Y, \mathcal{D})$. Further, the *max-min-covering dimension* of X, denoted $\texttt{MaxMinCOV}(X)$, is the largest min-covering dimension of any subset $Y \subset X$. In a formula:

$$\texttt{MaxMinCOV}(X) = \sup_{Y \subset X} \left(\inf_{\text{non-empty } U \subset Y:\ U \text{ is open in } (Y, \mathcal{D})} \texttt{COV}(U) \right).$$

We find that $\texttt{MaxMinCOV}$ is precisely the right notion to characterize per-metric optimal regret.

Theorem 1.5. *Consider the Lipschitz MAB problem on a compact metric space with* $d = \texttt{MaxMinCOV}(X)$. *If* $\gamma > \frac{d+1}{d+2}$, *then there exists a bandit algorithm* $\mathcal{A}$ *such that for every problem instance* $\mathcal{I}$ *its regret satisfies* $R(t) = O_{\mathcal{I}}(t^\gamma)$ *for all* t. *No such algorithm exists if* $d > 0$ *and* $\gamma < \frac{d+1}{d+2}$.

The fact that the above result allows an instance-dependent constant makes the corresponding lower bound more challenging: one needs to show that for any algorithm there exists a problem instance whose regret is at least t^γ *infinitely often*, whereas without an instance-dependent constant it suffices to show this for any one time t. The former requires a problem instance with infinitely many arms, whereas the latter can be accomplished via a simple problem instance with finitely many arms, and in fact is already done in Theorem 1.1.

In general, $\texttt{MaxMinCOV}(X)$ is bounded from above by the covering dimension of X. For metric spaces that are highly homogeneous, in the sense that any two ϵ-balls are isometric to one another, the two dimensions are equal, and the upper bound in the theorem can be achieved using a generalization of the UniformMesh algorithm described earlier. The difficulty in Theorem 1.5 lies in dealing with inhomogeneities in the metric space. It is important to treat the problem at this level of generality, because some of the most natural applications of the Lipschitz MAB problem, e.g., the web advertising problem described earlier, are based on highly inhomogeneous metric spaces.[7]

[7] For example, in web taxonomies, it is unreasonable to expect different categories at the same level of a topic hierarchy to have roughly the same number of descendants. Thus, in a natural interpretation of taxonomy as a metric space in Lipschitz MAB—where each subtree is a ball whose radius equals or upper-bounds the maximal difference between expected rewards in the said subtree—balls may be very different from one another.

The simplest scenario in which we improve over Theorem 1.2 involves a point $x \in X$ and a number $\epsilon > 0$ such that cutting out any open neighborhood of x reduces the covering dimension by at least ϵ. We think of such x as a "fat point" in the metric space. This example can be extended to a "fat region" $S \subset X$ such that $\mathtt{COV}(S) < \mathtt{COV}(X)$ and cutting out any open superset of S reduces the covering dimension by at least ϵ. One can show that $\mathtt{MaxMinCOV}(X) \leq \max\{\mathtt{COV}(S), \mathtt{COV}(X) - \epsilon\}$.

A "fat region" S becomes an obstacle for the zooming algorithm if it contains an optimal arm, in which case the algorithm needs to instantiate too many active arms in the vicinity of S. To deal with this, we impose a *quota* on the number of active arms outside S. The downside is that the set $X \setminus S$ is insufficiently covered by active arms. However, this downside does not have much impact on performance if an optimal arm lies in S. And if S does not contain an optimal arm, then the zooming algorithm learns this fact eventually, in the sense that it stops refining the mesh of active arms on some open neighborhood U of S. From then on, the algorithm essentially limits itself to $X \setminus U$, which is a comparatively low-dimensional set.

The general algorithm in Theorem 1.5 combines the above "quota-limited zooming" idea with a delicate decomposition of the metric space that gradually "peels off" regions with abnormally high covering dimension. In the above example with a "fat region" S, the decomposition consists of two sets, X and S. In general, the decomposition is a decreasing sequence of subsets $X = S_0 \supset S_1 \supset \cdots$ where each S_i is a "fat region" with respect to S_{i-1}. If the sequence is finite, then the algorithm has a separate quota for each S_i.

Further, to handle arbitrary metric spaces, we allow this sequence to be infinite, and moreover *transfinitely* infinite, i.e., parameterized by ordinal numbers. The algorithm proceeds in phases. Each phase i begins by "guessing" an ordinal $\lambda = \lambda_i$ that represents the algorithm's estimate of the largest index of a set in the transfinite sequence that intersects the set of optimal arms. During a phase, the algorithm focuses on the set S_λ in the sequence, and has a quota on active arms not in S_λ. In the end of the phase it uses the observed payoffs to compute the next ordinal λ_{i+1}. The analysis of the algorithm shows that almost surely, the sequence of guesses $\lambda_1, \lambda_2, \ldots$ is eventually constant, and that the eventual value of this sequence is almost surely equal to the largest index of a set in the transfinite sequence that intersects the set of optimal arms. The regret bound then follows easily from our analysis of the zooming algorithm.

For the lower bound, we craft a new dimensionality notion ($\mathtt{MaxMinCOV}$), which captures the inhomogeneity of a metric space, and connect this notion with the maximal possible "strength" of the transfinite decomposition. Further, we connect $\mathtt{MaxMinCOV}$ with the existence of a certain structure in the metric space (a *ball-tree*), which supports our lower-bounding example. This relation between the two structures—d-dimensional transfinite decompositions and d-dimensional ball-trees—is a new result on metric topology, and as such it may be of independent interest.

While the lower bound is proved using the notion of Kullback-Leibler divergence (*KL-divergence*), our usage of the KL-divergence technique is encapsulated as a generic theorem statement (Theorem 5.7). A similar encapsulation (Theorem 6.12) is stated and proved for the full-feedback version. These theorems and the corresponding setup may be of independent interest. In particular, Theorem 5.7 has been used in Slivkins [91] to encapsulate a version of the KL-divergence argument that underlies a lower bound on regret in a contextual bandit setting.

Per-metric optimality: beyond polynomial regret. To resolve question (Q3), we show that the apparent gap between logarithmic and polynomial regret is inherent to the Lipschitz MAB problem.

Theorem 1.6. *For Lipschitz MAB on any fixed metric space $(X, \mathcal{D})$, the following dichotomy holds: either it 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 metric completion of* $(X, \mathcal{D})$ *is a compact metric space with countably many points.*

Thus, we establish the $\log(t)$ vs. $\sqrt{t}$ regret dichotomy, and moreover show that it is determined by some of the most basic set-theoretic and topological properties of the metric space. For compact metric spaces, the dichotomy corresponds to the transition from countable to uncountable strategy sets. This is also surprising; in particular, it was natural to conjecture that if the dichotomy exists and admits a simple characterization, it would correspond to the finite vs. infinite transition.

Given the $\Omega(\log t)$ lower bound in Lai and Robbins [69], our upper bound for the Lipschitz MAB problem 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 1.7. *The Lipschitz MAB problem on any infinite metric space is not* $(\log t)$*-tractable.*

To answer question (Q4), we show that the tractability of the Lipschitz MAB problem on a complete metric space hinges on the compactness of the metric space.

Theorem 1.8. *The Lipschitz MAB problem on a fixed metric space* $(X, \mathcal{D})$ *is* $f(t)$*-tractable for some* $f \in o(t)$ *if and only if the metric completion of* $(X, \mathcal{D})$ *is a compact metric space.*

The main technical contribution in the above theorems is an interplay of online learning and point-set topology, which requires novel algorithmic and lower-bounding techniques. For the $\log(t)$ vs. $\sqrt{t}$ dichotomy result, we identify a simple topological property (existence of a topological well-ordering), which entails the algorithmic result, and another topological property (*perfectness*), which entails the lower bound. The equivalence of the first property to countability and the second to uncountability (for compact metric spaces) follows from classical theorems of Cantor-Bendixson [33] and Mazurkiewicz-Sierpinski [74].

1.5 Our Contributions: The Lipschitz Experts Problem

We turn our attention to the *Lipschitz experts problem*: the full-feedback version of the Lipschitz MAB problem. Formally, a problem instance is specified by a triple $(X, \mathcal{D}, \mathbb{P})$, where $(X, \mathcal{D})$ is a metric space and $\mathbb{P}$ is a probability measure with universe $[0,1]^X$, the set of all functions from X to $[0,1]$, such that the expected payoff function $\mu : x \mapsto \mathbb{E}_{f \in \mathbb{P}}[f(x)]$ is a Lipschitz function on $(X, \mathcal{D})$. The metric structure of $(X, \mathcal{D})$ is known to the algorithm, the measure $\mathbb{P}$ is not. We will refer to $\mathbb{P}$ as the problem instance when the metric space $(X, \mathcal{D})$ is clear from the context.

In each round t, the algorithm picks a strategy $x_t \in X$, then the environment chooses an independent sample $f_t : X \to [0,1]$ distributed according to the measure $\mathbb{P}$. The algorithm receives payoff $f_t(x_t)$, and also observes the entire payoff function f_t. More formally, the algorithm can query the value of f_t at an arbitrary finite number of points. Some of our upper bounds are for a (very) restricted version, called *double feedback*, where in each round the algorithm picks two arms (x, y), receives the payoff for x and also observes the payoff for y. By abuse of notation, we will treat the bandit setting as a special case of the experts setting.

Note that the payoffs for different arms in a given round are not necessarily independent. This is essential, because for any limit point x in the metric space one could use many independent samples from the vicinity of x to learn the expected payoff at x in a single round.

Regret dichotomies. We show that the Lipschitz experts problem exhibits a regret dichotomy similar to the one in Theorem 1.6. Since the optimal regret for a finite strategy set is *constant* [62], the dichotomy is between $O(1)$ and $\sqrt{t}$ regret.

Theorem 1.9. *The Lipschitz experts problem on metric space $(X, \mathcal{D})$ 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 case occurs if and only if the completion of X is a compact metric space with countably many points.*

Theorem 1.9 and its bandit counterpart (Theorem 1.6) are proved jointly, using essentially the same ideas. In both theorems, the regret dichotomy corresponds to the transition from countable to uncountable strategy set (assuming the metric space is compact and complete). Note that the upper bound in Theorem 1.9 only assumes double feedback, whereas the lower bound is for the unrestricted full feedback.

Next, we investigate for which metric spaces the Lipschitz experts problem is $o(t)$-tractable. We extend Theorem 1.8 for the Lipschitz MAB problem to another regret dichotomy where the upper bound is for the bandit setting, whereas the lower bound is for full feedback.

Theorem 1.10. *The Lipschitz experts problem on metric space $(X, \mathcal{D})$ 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.*

Polynomial regret in (very) high dimension. In view of the $\sqrt{t}$ lower bound from Theorems 1.9, we are interested in matching upper bounds. Gupta et al. [50] observed that such bounds hold for every metric space $(X, \mathcal{D})$ of finite covering dimension: namely, the Lipschitz experts problem on $(X, \mathcal{D})$ is $\sqrt{t}$-tractable. Therefore, it is natural to ask whether there exist metric spaces of *infinite* covering dimension with polynomial regret.

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. We define the *log-covering dimension* of $(X, \mathcal{D})$ as the smallest number $d \geq 0$ such that X can be covered by $O(2^{r^{-d}})$ sets of diameter r for all $r > 0$. More formally:

$$\mathtt{LCD}(X) = \limsup_{r \to 0} \frac{\log\log N_r(X)}{\log(1/r)}, \tag{2}$$

where $N_r(X)$ is the minimal size (cardinality) of a r-covering of X, i.e., the smallest number of sets of diameter at most r sufficient to cover X. Note that the number of sets allowed by this definition is exponentially larger than the one allowed by the covering dimension.

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 *ϵ-uniform tree metric* is a metric on the ends of an infinitely deep uniform tree, in which the distance between two ends is ϵ^{-i}, where i is the level of their least common ancestor. It is easy to see that an ϵ-uniform tree metric such that the branching factor at each level i is $\exp(\epsilon^{-id}(2^d - 1))$ has log-covering dimension d.

For another example, consider the set of all probability measures over $X = [0, 1]^d$ under the Wasserstein W_1 metric, a.k.a., the Earthmover distance.[8] We show that the log-covering dimension of this metric space is equal to the covering dimension of $(X, \mathcal{D})$. In fact, this example extends to any metric space X of finite diameter and covering dimension d; see Appendix D for the details.

Theorem 1.11. *Let $(X, \mathcal{D})$ be a metric space of log-covering dimension d. Then the Lipschitz experts problem is (t^γ)-tractable for any $\gamma > \frac{d+1}{d+2}$.*

[8] The Wasserstein W_1 metric is one of the standard ways to define a distance on probability measures. In particular, it is widely used in Computer Science literature to compare discrete distributions, e.g., in the context of image retrieval [85].

The algorithm in Theorem 1.11 is a version of `UniformMesh`. The same algorithm enjoys a better regret bound if each function $f \in \texttt{support}(\mathbb{P})$ is itself a Lipschitz function on $(X, \mathcal{D})$. We term this special case the *uniformly Lipschitz experts problem.*

Theorem 1.12. *Let $(X, \mathcal{D})$ be a metric space of log-covering dimension d. Then the uniformly Lipschitz experts problem is (t^γ)-tractable for any $\gamma > \frac{d-1}{d}$.*

The analysis is much more sophisticated compared to Theorem 1.11, using a chaining technique from empirical process theory (see Talagrand [95] for background).

Per-metric optimal regret bounds. We find that the log-covering dimension is not the right notion to characterize optimal regret for arbitrary metric spaces. Instead, we define the *max-min-log-covering dimension* (`MaxMinLCD`): essentially, we take the definition of `MaxMinCOV` and replace covering dimension with log-covering dimension.

$$\texttt{MaxMinLCD}(X) = \sup_{Y \subset X} \ \inf\{\, \texttt{LCD}(Z) : \text{ open non-empty } Z \subset Y\}. \tag{3}$$

Note that in general $\texttt{MaxMinLCD}(X) \leq \texttt{LCD}(X)$. Equality holds for "homogeneous" metric spaces such as ϵ-uniform tree metrics. We derive the regret characterization in terms `MaxMinLCD`; the characterization is tight for the uniformly Lipschitz experts problem.

Theorem 1.13. *Let $(X, \mathcal{D})$ be an uncountable metric space and $d = \texttt{MaxMinLCD} \geq 0$. Then:*

(a) the Lipschitz experts problem is (t^γ)-tractable for any $\gamma > \frac{d+1}{d+2}$,
(b) the uniformly Lipschitz experts problem is (t^γ)-tractable for any $\gamma > \max(\frac{d-1}{d}, \frac{1}{2})$,
(c) the uniformly Lipschitz experts problem is not (t^γ)-tractable for any $\gamma < \max(\frac{d-1}{d}, \frac{1}{2})$.

The algorithms in parts (a) and (b) use a generalization of the transfinite decomposition from the bandit per-metric optimal algorithm (Theorem 1.5). The lower bound in part (c) builds on the lower-bounding technique for the $\sqrt{t}$ lower bound on uncountable metric spaces.

Our results for Lipschitz experts amount to a nearly complete characterization of per-metric optimal regret bounds, analogous to that in Table 1 on page 7. This characterization is summarized in the table below. (The characterization falls short of being complete, because the upper and lower bounds for finite $\texttt{MaxMinLCD} = d \in [0, \infty)$ do not match.)

1.6 Discussion

Accessing the metric space. In stating the theorems above, we have been imprecise about specifying the model of computation. In particular, we have ignored the thorny issue of how to provide an algorithm with an input describing a metric space that may have an infinite number of points. The simplest way to interpret our theorems is to ignore implementation details and interpret an "algorithm" to mean an abstract decision rule, i.e., a (possibly randomized) Borel-measurable function mapping the history of past observations to an arm $x \in X$, which is played in the current period. All of our theorems are valid under this interpretation, but they can also be made into precise algorithmic results provided that the algorithm is given appropriate oracle access to the metric space.

The zooming algorithm requires only a *covering oracle*, which takes a finite collection of open balls and either declares that they cover X or outputs an uncovered point. The algorithm poses only one oracle query in each round t, for a collection of at most t balls. (For infinite metric spaces of interest that admit a finite description, e.g., rational convex polytopes in Euclidean space, it is generally easy to implement a covering oracle given a description of the metric space.) The per-metric optimal algorithm in Theorem 1.5 uses more complicated oracles, and we defer the definition of these oracles to Section 5.

Table 2. Per-metric Optimal Bounds for Lipschitz Experts

If the completion of $(X, \mathcal{D})$ is …	then regret can be …	but not …
compact and countable	$O(1)$	—
compact and uncountable		
finite covering dimension	$\tilde{O}\left(\sqrt{t}\right)$	$o(\sqrt{t})$
$\texttt{MaxMinLCD} = d \in [0, \infty)$	$\tilde{O}\left(t^\gamma\right), \gamma > \frac{d+1}{d+2}$	$o\left(t^\gamma\right), \gamma = \frac{1}{2}$ or $\gamma < \frac{d-1}{d}$
$\texttt{MaxMinLCD} = \infty$	$o(t)$	$O\left(t^\gamma\right), \gamma < 1$
non-compact	$O(t)$	$o(t)$

The $\omega(\log t)$-regret algorithms for countably infinite metric spaces (Theorems 1.6 and 1.9) require an oracle that represents the well-ordering of the metric space. We also provide an extension for compact metric spaces with a finite number of limit points for which a more intuitive oracle access suffices. In fact, this extension holds for a much wider family of metric spaces: those with a finite *Cantor-Bendixson rank*, a classic notion from point-set topology.

Further directions. While general, our model is idealized in several ways. Numerical similarity information, such as the distances and the Lipschitz constant, may be difficult to obtain in practice. The notion of similarity is "worst-case," so that the distances may need to be large to accommodate a few outliers. The reward distribution does not change over time. These issues gave rise to a line of follow-up work, detailed in Section 2.

Map of the article. We discuss related work in Section 2. In particular, a considerable amount of *follow-up work* is surveyed in Section 2. Preliminaries are presented in Section 3, including sufficient background on metric topology and dimensionality notions, and the proof of the initial observation (Theorem 1.2).

In the rest of the article, we present our technical results. Section 4 is on Lipschitz bandits with benign payoff functions; it presents the zooming algorithm and extensions thereof. Section 5 is on the per-metric optimal algorithms for Lipschitz bandits, focusing on polynomial regret. The next two sections concern both Lipschitz bandits and Lipschitz experts: Section 6 is on the dichotomy between (sub)logarithmic and $\sqrt{t}$ regret, and Section 7 studies for which metric spaces the Lipschitz bandits/experts problem is $o(t)$-tractable. Section 8 is on the polynomial-regret algorithms for Lipschitz experts. We conclude with directions for further work in Section 9.

To preserve the flow of the article, some material is deferred to appendices. In Appendix A, we present sufficient background on Kullback-Leibler divergence (KL-divergence) and the technical proofs that use the KL-divergence technique. In Appendix B, we reduce the Lipschitz bandits/experts problem to that on complete metric spaces. In Appendix C, we present a self-contained proof of a theorem from general topology, implicit in Cantor [33], Mazurkiewicz and Sierpinski [74], which ties together the upper and lower bounds of the regret dichotomy result. Finally, in Appendix D, we flesh out the Earthmover distance example from Section 1.5.

2 RELATED AND FOLLOW-UP WORK

Multi-armed bandits. MAB problems have a long history; a thorough survey is beyond the scope of this article. On a very high level, there is a crucial distinction between regret-minimizing formulations [27, 35] and Bayesian/MDP formulations [21, 49]. Among regret-minimizing formulations, an important distinction is between stochastic payoffs [11, 69] and adversarial payoffs [12].

This article is on regret minimization with stochastic payoffs. The basic setting here is $k < \infty$ arms with no additional structure. Then the optimal regret is $R(t) = O(\sqrt{kt})$, and $R(t) = O_{\mathcal{I}}(\log t)$ with an instance-dependent constant [11, 12, 69]. Note that the distinction between regret rates with and without instance-dependent constants is inherent even in this basic bandit setting. The UCB1 algorithm [11] achieves the $O_{\mathcal{I}}(\log t)$ bound and simultaneously matches the $O(\sqrt{kt})$ bound up to a logarithmic factor.

Our zooming algorithm relies on the "UCB index" technique from Auer et al. [11]. This is a simple but very powerful idea: arms are chosen according to a numerical score, called *index*, which is defined as an upper confidence bound (UCB) on the expected payoff of a given arm. Thus, the UCB index can be represented as a sample average plus a confidence term, which represent, respectively, exploitation and exploration, so that the sum represents a balance between the two. Several papers [8, 9, 13, 47, 57, 73] designed improved versions of the UCB index for the k-armed MAB problem with stochastic payoffs, achieving regret bounds, which are even closer to the lower bound. Moreover, the UCB index idea and various extensions thereof have been tremendously useful in many other settings with exploration-exploitation tradeoff, e.g., References [1, 10, 17, 28, 62, 91, 93, 101]. It is worth noting that the zooming algorithm, as originally published in Reference [64], was one of the first results in this line of work.

Many papers enriched the basic MAB setting by assuming some structure on arms, typically to handle settings where the number of arms is very large or infinite. Most relevant to this article is the work on *continuum-armed bandits* [4, 14, 60], a special case of Lipschitz MAB where the metric space is $([0,1], \ell_1)$. A closely related model posits that arms correspond to leaves on a tree, but no metric space is revealed to the algorithm [67, 80, 81, 90]. Another commonly assumed structure is linear or convex payoffs, e.g., References [2, 15, 40, 44, 52]. Linear/convex payoffs is a much stronger assumption than similarity, essentially because it allows to make strong inferences about far-away arms. Accordingly, it admits much stronger regret bounds, such as $\tilde{O}(d\sqrt{t})$ for arms in $\mathbb{R}^d$. Other structures in the literature include infinitely many i.i.d. arms [23, 101], Gaussian Process Bandits [42, 68, 94] and Functional bandits [7]; Gaussian Process MAB and Functional MAB are discussed in more detail in Section 2.

Closely related to continuum-armed bandits is the model of (regret-minimizing) *dynamic pricing* with unlimited supply [26, 66]. In this model, an algorithm is a seller with unlimited supply of identical items, such as a digital good (a movie, a song, or a program) that can be replicated at no cost. Customers arrive sequentially, and to each customer the algorithm offers one item at a non-negotiable price. Here prices correspond to arms, and accordingly the "arms" have a real-valued structure. Due to the discontinuous nature of demand (a customer who values the item at v will pay a price of $v - \epsilon$ but will pay nothing if offered a price of $v + \epsilon$) dynamic pricing is not a special case of Lipschitz MAB, but there is a close relationship between the techniques that have been used to solve both problems. Moreover, when the distribution of customer values has bounded support and bounded probability density, the expected revenue is a Lipschitz function of the offered price, so regret-minimizing dynamic pricing in this case reduces to the Lipschitz MAB problem.[9] One can also consider selling $d > 1$ products, offering a different price for each. When the expected revenue is a Lipschitz function of the offered price vector, this is a special case of Lipschitz MAB with arms in $\mathbb{R}^d$.

Interestingly, the dichotomy between (poly)logarithmic and $\sqrt{t}$ regret has appeared in four different MAB settings: Theorem 1.6 in this article, k-armed bandits with stochastic payoffs (as mentioned above), bandits with linear payoffs [41], and an extension of MAB to pay-per-click auctions [18, 19, 43]. These four dichotomy results have no obvious technical connection.

Metric spaces and dimensionality notions. Algorithmic problems on metric spaces have a long history in many different domains. These domains include: constructing space-efficient and/or

[9] Some of the work on dynamic pricing, e.g., References [24, 102], makes the Lipschitz assumption directly.

algorithmically tractable representations such as metric embeddings, distance labels, or distance oracles; problems with costs where costs have a metric structure, e.g., facility location and traveling salesman; offline and online optimization on a metric space; finding hidden structure in a metric space (classification and clustering).

Covering dimension is closely related to several other notions of dimensionality of a metric space, such as Haussdorff dimension, capacity dimension, box-counting dimension, and Minkowski-Bouligand Dimension. All these notions are used to characterize the covering properties of a metric space in fractal geometry; discussing fine distinctions between them is beyond our scope. A reader can refer to Schroeder [86] for background.

Covering numbers and covering dimension have been widely used in Machine Learning to characterize the complexity of the *hypothesis space*: a space of functions over X, the domain for which the learner needs to predict or classify, under functional ℓ_2 norm and some distribution over X. This is different from the way covering numbers and similar notions are used in the context of the Lipschitz MAB problem, and we are not aware of a clear technical connection.[10] Non-metric notions to characterize the complexity of function classes include VC-dimension, fat-shattering dimension, and Rademacher averages; see Shalev-Shwartz and Ben-David [87] for background.

Various notions of dimensionality of metric spaces have been studied in the theoretical computer science literature, with a goal to arrive at (more) algorithmically tractable problem instances. The most popular notions have been the ball-growth dimension, e.g., References [3, 55, 58, 89], and the doubling dimension, e.g., References [36, 51, 59, 75, 88, 96]. These notions have been useful in many different problems, including metric embeddings, other space-efficient representations such as distance labels and sparse spanners, network primitives such as routing schemes and distributed hash tables, and approximation algorithms for various optimization problems such as traveling salesman, k-median, and facility location.

Concurrent and Independent Work

Bubeck et al. [29, 30] obtain results similar to Theorems 1.3 and 1.4. They use similar, but technically different, notions of instance-dependent metric dimension and relaxed Lipschitzness. They also obtain stronger regret bounds for some special cases; these extensions are similar in spirit to the extended analysis of the zooming algorithm in this article (but technically different). Their results use a different algorithm and the proof techniques appear different.

While the publication of our conference version [64] predated the submission of theirs [29],[11] we believe the latter is concurrent and independent work.

Follow-up Work

Since the conference publication of Kleinberg et al. [64], there has been a considerable amount of follow-up work on Lipschitz MAB and various extensions thereof.

Lower bounds. While our lower bound for "benign" problem instances (in Theorem 1.3) comes from the worst-case scenario when the zooming dimension equals the covering dimension, Slivkins [91] and Magureanu et al. [71] provide more refined, *instance-dependent* lower bounds. Slivkins [91] proves that the upper bound in Theorem 4.4 is tight, up to $O(\log^2 t)$ factors, *for*

[10] In the Lipschitz MAB problem, one is interested in the family $\mathcal{F}$ of all Lipschitz-continuous functions on $(X, \mathcal{D})$, and therefore one could consider the covering numbers for $\mathcal{F}$, or use any other standard notions such as VC-dimension or fat-shattering dimension. However, we have not been able to reach useful results with this approach.

[11] Bubeck et al. [29] acknowledge Theorem 1.3 as prior work. It appears that the authors of Reference [29] have not been aware of other results in Reference [64] at the time (which were only briefly mentioned in the conference version, and fleshed out in the full version [65]).

every value of the said upper bound.[12] Magureanu et al. [71] focus on regret bounds of the form $C \cdot \log(t) + O(1)$, where C depends on the problem instance, but not on time. They derive a lower bound on the C, and provide an algorithm that comes arbitrarily close to this lower bound.

Contextual Lipschitz MAB and applications. Lu et al. [70] and Slivkins [91], simultaneous and independent w.r.t. one another,[13] extend Lipschitz MAB to the contextual bandit setting, where in each round the algorithm receives a context ("hint") h and picks an arm x, and the expected payoff is a function of both h and x. The motivational examples include placing ads on webpages (webpages and/or users are contexts, ads are arms), serving documents to users (users are contexts, documents are arms), and offering prices to customers (customers are contexts, prices are arms). The similarity information is expressed as a metric on contexts and a metric on arms, with the corresponding two Lipschitz conditions. Lu et al. [70] consider this setting and extend `UniformMesh` to obtain regret bounds in terms of the covering dimensions of the two metric spaces. Slivkins [91] extends the zooming algorithm to the contextual setting and obtains improved regret bounds in terms of a suitable "contextual" version of the zooming dimension.

The "contextual zooming algorithm" from Slivkins [91] works in a more general setting where similarity information is represented as a metric space on the set of "allowed" context-arm pairs, and the expected payoff function is Lipschitz with respect to this metric space. This is a very versatile setting: it can also encode *sleeping bandits* [25, 46, 62] (in each round, some arms are "asleep," i.e., not available) and slowly changing payoffs [93] (here in each round t the context is t itself, and the metric on contexts expresses the constraint how fast the expected payoffs can change). This setting showcases the full power of the adaptive refinement technique that underlies the zooming algorithm.

Further, Slivkins et al. [92] use the zooming algorithm from this article and its contextual version from Slivkins [91] in the context of *ranked bandits* [83]. Here in each round a bandit algorithm chooses an ordered list of k documents (from a much larger pool of available documents) and presents it to a user who scrolls the list top-down and clicks on the first document that she finds relevant. The user may leave after the first click; the goal is to minimize the number of users with no clicks. The contribution of Reference [92] is to combine ranked bandits with Lipschitz MAB; among other things, this requires a significantly extended model: If two documents are close in the metric space, then their click probabilities are similar even conditional on the event that some other documents are not clicked by the current user.

Partial similarity information. A number of papers tackle the issue that the numerical similarity information required for the Lipschitz MAB problem may be difficult to obtain in practice. These papers make various assumptions on what is and is not revealed to the algorithm, with a general goal to do (almost) as well as if the full metric space were known. Bubeck et al. [31] study a version with strategy set $[0,1]^d$ and Lipschitz constant that is not revealed, and match the optimal regret rate for algorithms that know the Lipschitz constant. Minsker [76] considers the same strategy set and distance function of the form $\|x-y\|_\infty^\beta$, where the smoothness parameter $\beta \in (0,1]$ is not known. References [32, 78, 90, 98] study a version in which the algorithm only inputs a "taxonomy" on arms (i.e., a tree whose leaves are arms), whereas the numerical similarity information is not revealed at all. This version features a *second* exploration-exploitation tradeoff:

[12] In fact, this lower bound extends to the contextual bandit setting.

[13] The initial version of Slivkins [91] has appeared on `arxiv.org` in 2009. It contained the main algorithm, the same as in the final version, but only derived results for the covering dimension. The conference version from *COLT 2011* contained essentially the same results as in the final journal version from 2014.

the tradeoff between learning more about the numerical similarity information (or some relevant portions thereof), and exploiting this knowledge to run a Lipschitz MAB algorithm.

The latter line of work proceeds as follows. Slivkins [90] considers the metric space implicitly defined by the taxonomy, where the distance between any two arms is the maximal difference in expected rewards in the least common subtree. He puts forward an extension of the zooming algorithm, which adaptively reconstructs the implicit metric space, and (under some additional assumptions) essentially matches the performance of the zooming algorithm on the same metric space. Munos [78] and Valko et al. [98] allow a more general relation between the implicit metric space and the taxonomy, and moreover relax the Lipschitz condition to only hold w.r.t. the maximum (as in Theorem 4.3). Munos [78] focuses on deterministic rewards, and essentially matches the regret bound in Corollary 4.16, whereas Valko et al. [98] study the general IID case. Finally, Bull [32] considers a somewhat more general setting with multiple taxonomies on arms (or with arms embedded in $[0,1]^d$, where the embedding is then used to define the taxonomies). The paper extends and refines the algorithm from Slivkins [90], and carefully traces out the conditions under which one can achieve $\tilde{O}(\sqrt{T})$ regret. Munos [79] surveys some of this work, with emphasis on the techniques from [30, 78, 98].

As a stepping stone to the result mentioned above, Munos [78] considers Lipschitz MAB with deterministic rewards and essentially matches our result for this setting (Corollary 4.16).[14]

Beyond IID rewards. Several papers [16, 72, 91] consider Lipschitz bandits/experts with non-IID rewards.[15] Azar et al. [16] consider a version of Lipschitz MAB in which the IID condition is replaced by more sophisticated ergodicity and mixing assumptions, and essentially recover the performance of the zooming algorithm. Maillard and Munos [72] consider Lipschitz experts in a Euclidean space $(\mathbb{R}^d, \ell_2)$ of constant dimension d. Assuming the Lipschitz condition on realized payoffs (rather than expected payoffs), they achieve a surprisingly strong regret of $O(\sqrt{t})$. Slivkins [91] considers contextual bandits with Lipschitz condition on expected payoffs, and provides a "meta-algorithm," which uses an off-the-shelf bandit algorithm such as `EXP3` [12] as a subroutine and adaptively refines the space of contexts. Also, as discussed above, the contextual zoooming algorithm from Slivkins [91] can handle Lipschitz MAB with slowly changing rewards.

Other structural models of MAB. One drawback of Lipschitz MAB as a model is that $\mathcal{D}(x,y)$ only gives a worst-case notion of similarity between arms x and y: a hard upper bound on $|\mu(x) - \mu(y)|$ rather than a typical or expected upper bound. In particular, the distances may need to be very large to accommodate a few outliers, which would make $\mathcal{D}$ less informative elsewhere.[16] With this criticism in mind, Srinivas et al. [94] define a probabilistic model, called *Gaussian Processes Bandits*, where the expected payoff function is distributed according to a suitable Gaussian Process on X, thus ensuring a notion of "probabilistic smoothness" with respect to X. Further work in this model includes Krause and Ong [68] and Desautels et al. [42].

Given the work on Lipschitz MAB (and other "structured" bandit models such as linear payoffs) it is tempting to consider MAB with *arbitrary* known structure on payoff functions. Amin et al. [7] initiate this direction: in their model, the structure is explicitly represented as the collection of all possible payoff functions. However, their results do not subsume any prior work on Lipschitz MAB or MAB with linear or convex payoffs.

[14] While our original publications [64, 65] predate Reference [78], the latter is independent work to the best of our understanding.

[15] The first result in this direction appeared in Kleinberg [60]. He considers Lipschitz MAB with adversarial rewards, and proposes a version of `UniformMesh` where an adversarial bandit algorithm is used instead of `UCB1`. This algorithm achieves the same worst-case regret as `UniformMesh` on IID rewards.

[16] This concern is partially addressed by Theorem 1.4.

Further applications of our techniques. Ho et al. [56] design a version of the zooming algorithm in the context of crowdsourcing markets. Here the algorithm is an employer who offers a quality-contingent contract to each arriving worker, and adjusts the contract over time. This is an MAB problem in which arms are contracts (essentially, vectors of prices), and a single round is modeled as a standard "principal-agent model" from contract theory. Ho et al. [56] do not assume a Lipschitz condition, or any other explicit guarantee on similarity between arms. Instead, their algorithm estimates the similarity information on the fly, taking advantage of the structure provided by the principal-agent model.[17]

On a final note, one of our minor results—the improved confidence radius from Section 4.2—may be of independent interest. In particular, this result is essential for some of the main results in [5, 6, 17, 20], in the context of dynamic pricing and other MAB problems with global supply/budget constraints.

3 PRELIMINARIES

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

Problem Formulation and Notation

In the Lipschitz MAB problem, the problem instance is a triple $(X, \mathcal{D}, \mu)$, where $(X, \mathcal{D})$ is a metric space and $\mu : X \to [0, 1]$ is a a Lipschitz function on $(X, \mathcal{D})$ with Lipschitz constant 1. (In other words, μ satisfies *Lipschitz condition* (1)). $(X, \mathcal{D})$ is revealed to an algorithm, whereas μ is not. In each round t the algorithm chooses a strategy $x = x_t \in X$ and receives payoff $f_t(x) \in [0, 1]$ chosen independently from some distribution $\mathbb{P}_x$ with expectation $\mu(x)$. Without loss of generality, the diameter of $(X, \mathcal{D})$ is at most 1. To simplify exposition, the parameterized family of reward distributions $\mathbb{P}_x$ is assumed to be fixed over time, and suppressed from the notation.

Throughout the article, $(X, \mathcal{D})$ and μ will denote, respectively, a metric space of diameter ≤ 1 and a Lipschitz function as above. We will say that X is the set of strategies ("arms"), $\mathcal{D}$ is the *similarity function*, and μ is the *payoff function*.

Performance of an algorithm is measured via *regret* with respect to the best fixed strategy:

$$R(t) = t \sup_{x \in X} \mu(x) - \mathbb{E}\left[\sum_{s=1}^{t} \mu(x_s)\right], \tag{4}$$

where $x_t \in X$ is the strategy chosen by the algorithm in round t. Note that when the supremum is attained, the first summand in Equation (4) is the expected reward of an algorithm that always plays the best strategy.

Throughout this article, the constants in the $O(\cdot)$ notation are absolute unless specified otherwise. The notation $O_{\text{subscript}}$ means that the constant in $O()$ can depend on the things listed in the subscript. Denote $\sup(\mu, X) = \sup_{x \in X} \mu(x)$ and similarly $\operatorname{argmax}(\mu, X) = \operatorname{argmax}_{x \in X} \mu(x)$.

Metric Topology and Set Theory

Let X be a set and let $(X, \mathcal{D})$ be a metric space. An open ball of radius r around point $x \in X$ is $B(x, r) = \{y \in X : \mathcal{D}(x, y) < r\}$. The diameter of a set is the maximal distance between any two points in this set.

[17] References [32, 90, 98] estimate the "hidden" similarity information for a general Lipschitz MAB setting (using some additional assumptions), but Ho et al. [56] use a different, problem-specific approach that side-steps some of the assumptions.

A *Cauchy sequence* in $(X, \mathcal{D})$ is a sequence such that for every $\delta > 0$, there is an open ball of radius δ 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* $d(\mathbf{x}, \mathbf{y}) = \lim_{i \to \infty} d(x_i, y_i)$ 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 $(X^*, \mathcal{D})$ called the *(metric) completion* of $(X, \mathcal{D})$. The subspace of all constant sequences is identified with $(X, \mathcal{D})$: formally, it is a dense subspace of $(X^*, \mathcal{D})$, which is isometric to $(X, \mathcal{D})$. A metric space $(X, \mathcal{D})$ is *compact* if every collection of open balls covering $(X, \mathcal{D})$ has a finite subcollection that also covers $(X, \mathcal{D})$. Every compact metric space is complete, but not vice-versa.

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 article, 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.

A *well-ordering* on a set X is a total order on X with the property that every non-empty subset of X has a least element in this order. In Section 8.3.2, we use *ordinals*, a.k.a., *ordinal numbers*, a class of well-ordered sets that, in some sense, extends natural numbers beyond infinity. Understanding this article 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*.

Dimensionality Notions

Throughout this article, we will use various notions of dimensionality of a metric space. The basic notion will be the *covering dimension*, which is a version of the *fractal dimension* that is based on covering numbers. We will also use several refinements of the covering dimension that are tuned to the Lipschitz MAB problem.

Definition 3.1. Let Y be a set of points in a metric space $(X, \mathcal{D})$. For each $r > 0$, an *r-covering* of Y is a collection of subsets of Y, each of diameter strictly less than r, that cover Y. The minimal number of subsets in an r-covering is called the *r-covering number* of Y and denoted $N_r(Y)$. The *covering dimension* of Y with multiplier c, denoted $\texttt{COV}_c(Y)$, is the infimum of all $d \geq 0$ such that $N_r(Y) \leq c\, r^{-d}$ for each $r > 0$.

This definition is *robust*: $N_r(Y') \leq N_r(Y)$ for any $Y' \subset Y$, and consequently $\texttt{COV}_c(Y') \leq \texttt{COV}_c(Y)$. While covering numbers are often defined via radius-r balls rather than diameter-r sets, the former alternative does not have this appealing "robustness" property.

Remark. Fractal dimensions of infinite spaces are often defined using lim sup as the distance scale tends to 0. The lim sup-version of the covering dimension would be

$$\begin{aligned} \texttt{COV}(Y) &\triangleq \limsup_{r \to 0} \frac{\log N_r(Y)}{\log 1/r} \qquad (5)\\ &= \inf\left\{ d \geq 0 :\ \exists c\ \forall r > 0 \quad N(r) \leq c r^{-d} \right\} \\ &= \lim_{c \to \infty} \texttt{COV}_c(Y). \end{aligned}$$

This definition is simpler in that it does not require an extra parameter c. However, it hides an arbitrarily large constant, and is uninformative for finite metric spaces. On the contrary, the version in Definition 3.1 makes the constant explicit (which allows for numerically sharper bounds), and is meaningful for both finite and infinite metric spaces.

Remark. Instead of the covering-based notions in Definition 3.1 one could define and use the corresponding packing-based notions. A subset $S \subset Y$ is an *r-packing* of Y if the distance between any two points in S is at least r. An *r-net* of Y is a set-wise maximal r-packing; equivalently, S is an r-net if and only if it is an r-packing and the balls $B(x,r)$, $x \in S$ cover Y. The maximal number of points of an r-packing is called the *r-packing number* of Y and denoted $N_r^{\text{pack}}(Y)$. The "packing dimension" can then be defined as in Definition 3.1. It is a well-known folklore result that the packing and covering notions are closely related:

Fact 3.2. $N_{2r}(Y) \leq N_r^{\text{pack}}(Y) \leq N_r(Y)$.

Proof. Suppose the maximal size of an r-packing is finite, and let S be an r-packing of this size. First, for any r-covering $\{Y_i\}$, each set Y_i can contain at most one point in S, and each point in S is contained in some Y_i. So the r-covering has size at least $|S|$. Thus, $N_r^{\text{pack}}(Y) \leq N_r(Y)$. Second, $\{B(x,r) : x \in S\}$ is a $2r$-covering: else there exists a point x_0 that is not covered, and $S \cup \{x_0\}$ is an r-packing of larger size. So $N_{2r}(Y) \leq N_r^{\text{pack}}(Y)$. It remains to consider the case when there exists an r-packing S of infinite size. Then using the same argument as above, we show that any r-covering consists of infinitely many sets. □

For any set of finite diameter, the covering dimension (with multiplier 1) is at most the doubling dimension, which in turn is at most d for any point set in $(\mathbb{R}^d, \ell_p)$. The doubling dimension [54] has been a standard notion in the theoretical computer science literature (e.g., References [37, 51, 59, 96]). For the sake of completeness, and because we use it in Section 4.3, let us give the definition: the *doubling dimension* of a subset $Y \subset X$ is the smallest (infimum) $d > 0$ such that any subset $S \subset Y$ whose diameter is r can be covered by 2^d sets of diameter at most $r/2$. The doubling dimension is much more restrictive that the covering dimension. For example, $Y = \{2^{-i} : i \in \mathbb{N}\}$ under the ℓ_1 metric has doubling dimension 1 and covering dimension 0.

Concentration Inequalities

We use an elementary concentration inequality known as *the Chernoff bounds*. Several formulations exist in the literature; the one we use is from Reference [77].

Theorem 3.3 (Chernoff Bounds [77]). *Consider i.i.d. random variables $Z_1, \ldots, Z_n$ with values in $[0,1]$. Let $Z = \frac{1}{n}\sum_{i=1}^n Z_i$ be their average, and let $\zeta = \mathbb{E}[Z]$. Then:*

(a) $\Pr[|Z - \zeta| > \delta\zeta] < 2\exp(-\zeta n \delta^2/3)$ *for any* $\delta \in (0,1)$.
(b) $\Pr[Z > a] < 2^{-an}$ *for any* $a > 6\zeta$.

Initial Observation: Proof of Theorem 1.2

We extend algorithm `UniformMesh` from metric space $([0,1], \ell_1^d)$ to an arbitrary metric space of covering dimension d. The algorithm is parameterized by d. It divides time into phases of exponentially increasing length. During each phase i, the algorithm picks some $\delta_i > 0$, chooses an arbitrary δ_i-net S_i for the metric space, and only plays arms in S_i throughout the phase. Specifically, it runs an $|S_i|$-armed bandit algorithm on the arms in S_i. For concreteness, let us say that we use `UCB1` (any other bandit algorithm with the same regret guarantee will suffice), and each phase i lasts 2^i rounds. The parameter δ_i is tuned optimally given d and the phase duration T; the optimal value turns out to be $\delta = \tilde{O}(T^{-1/(d+2)})$. The algorithm can be analyzed using the technique from Reference [60].

Theorem 3.4. *Consider the Lipschitz MAB problem on a metric space $(X, \mathcal{D})$. Let d be the covering dimension of $(X, \mathcal{D})$ with multiplier c. Then regret of* `UniformMesh`*, parameterized by d, satisfies*

$$R(t) = O\left((c \log t)^{1/(d+2)}\ t^{1-1/(d+2)}\right) \quad \textit{for every time } t. \tag{6}$$

Proof. Let us analyze a given phase i of the algorithm. Let $R_i(t)$ be the regret accumulated in rounds 1 to t in this phase. Let $\delta = \delta_i$ and let $K = |S_i|$ be the number of arms in this phase that are considered by the algorithm. The regret of `UCB1` in t rounds is $O(\sqrt{K\, t \log t})$ [11]. It follows that

$$R_i(t) \leq O\left(\sqrt{Kt \log t}\right) + t(\mu^* - \sup(\mu, S_i)), \text{where } \mu^* = \sup(\mu, X).$$

Note that $\sup(\mu, S_i) \geq \mu^* - \delta$. (Indeed, since μ is a Lipschitz-continuous function on a compact metric space, there exists an optimal arm $x^* \in X$ such that $\mu(x^*) = \mu^*$. Take an arm $x \in S_i$ such that $\mathcal{D}(x, x^*) < \delta$. Then $\mu(x) \geq \mu^* - \delta$.) Further, $K \leq c\delta^{-d}$, since S_i is a δ-net. We obtain

$$R_i(t) \leq O\left(\sqrt{c\, \delta^{-d}\, t \log t} + \delta t\right).$$

Substituting $t = 2^i$, $\delta = (ct\ \log t)^{-1/(d+2)}$, yields $R_i(t) = O((c \log t)^{1/(d+2)}\ t^{1-1/(d+2)})$.

We obtain Equation (6) by summing over all phases $i = 1, 2, \ldots, \lceil \log t \rceil$. For the last phase $i = \lceil \log t \rceil$ (which is possibly incomplete), the regret accumulated in this phase is at most $R_i(2^i)$. □

4 THE ZOOMING ALGORITHM FOR LIPSCHITZ MAB

This section is on the *zooming algorithm*, which uses adaptive refinement to take advantage of "benign" input instances. We state and anlyze the algorithm, and derive a number of extensions.

The zooming algorithm proceeds in phases $i = 1, 2, 3, \ldots$. Each phase i lasts 2^i rounds. Let us define the algorithm for a single phase i_{ph} of the algorithm. For each arm $x \in X$ and time t, let $n_t(x)$ be the number of times arm x has been played in this phase before time t, and let $\mu_t(x)$ be the corresponding average reward. Define $\mu_t(x) = 0$ if $n_t(x) = 0$. Note that at time t both quantities are known to the algorithm.

Define the *confidence radius* of arm x at time t as

$$r_t(x) := \sqrt{\frac{8\, i_{\text{ph}}}{1 + n_t(x)}}. \tag{7}$$

The meaning of the confidence radius is that with high probability (i.e., with probability tending to 1 exponentially fast as i_{ph} increases) it bounds from above the deviation of $\mu_t(x)$ from its expectation $\mu(x)$. That is[18]

$$\text{w.h.p.} \quad |\mu_t(x) - \mu(x)| \leq r_t(x) \quad \text{for all times } t \text{ and arms } x. \tag{8}$$

Our intuition is that the samples from arm x available at time t allow us to estimate $\mu(x)$ only up to ±$r_t(x)$. Thus, the available samples from x do not provide enough confidence to distinguish x from any other arm in the ball of radius $r_t(x)$ around x. Call $B(x,\ r_t(x))$ the *confidence ball* of arm x (at time t).

Throughout the execution of the algorithm, a finite number of arms are designated *active*, so that in each round the algorithm only selects among the active arms. In each round at most one additional arm is activated. Once an arm becomes active, it stays active until the end of the phase. It remains to specify two things: the *selection rule,* which decides which arm to play in a given round, and the *activation rule,* which decides whether and which arm to activate.

[18] Here and throughout this article, we use the abbreviation "w.h.p." to denote the phrase *with high probability.*

- *Selection rule.* Choose an active arm x with the maximal *index*, defined as
$$I_t(x) = \mu_t(x) + 2\, r_t(x). \tag{9}$$
This definition of the index is meaningful, because as long as Equation (8) holds, the index of x is an upper bound on the expected payoff of any arm in the confidence ball of x. (We will prove this later.) The factor 2 in Equation (9) is needed, because we "spend" one $+r_t(x)$ to take care of the sampling uncertainty, and another $+r_t(x)$ to generalize from x to the confidence ball of x. Note that the index in algorithm UCB1 [11] is essentially $\mu_t(x) + r_t(x)$.
- *Activation rule.* Say that an arm is *covered* at time t if it is contained in the confidence ball of some active arm. We maintain the invariant that at each time all arms are covered. The activation rule simply maintains this invariant: If there is an arm that is not covered, then pick any such arm and make it active. Note that the confidence radius of this newly activated arm is initially greater than 1, so all arms are trivially covered. In particular, it suffices to activate at most one arm per round. The activation rule is implemented using the *covering oracle*, as defined in Section 1.6.

The bare pseudocode of the algorithm is very simple; see Algorithm 1.

ALGORITHM 1: Zooming Algorithm

```
for phase $i = 1, 2, 3, \ldots$ do
    Initially, no arms are active.
    for round $t = 1, 2, 3, \ldots, 2^i$ do
        Activation rule: if some arm is not covered, pick any such arm and activate it.
        Selection rule: play any active arm with the maximal index (9).
```

To state the provable guarantees, we need the notion of *zooming dimension* of a problem instance. As discussed in Section 1.4, this notion bounds the covering number of near-optimal arms, thus sidestepping the worst-case lower-bound examples. Throughout this section, $\mu^* \triangleq \sup(\mu, X)$ denotes the maximal reward, and $\Delta(x) = \mu^* - \mu(x)$ is the "badness" of arm x.

Definition 4.1. Consider a problem instance $(\mathcal{D}, X, \mu)$. The set of near-optimal arms at scale $r \in (0, 1]$ is defined to be
$$X_{\mu,\, r} \triangleq \{x \in X : \frac{r}{2} < \Delta(x) \leq r\}.$$
The *zooming dimension* with multiplier $c > 0$ is the smallest $d \geq 0$ such that for every scale $r \in (0, 1]$ the set $X_{\mu,\, r}$ can be covered by $c\, r^{-d}$ sets of diameter strictly less than $r/8$.

Theorem 4.2. *Consider an instance of the Lipschitz MAB problem. Fix any $c > 0$ and let d be the zooming dimension with multiplier c. Then the regret $R(t)$ of the zooming algorithm satisfies*
$$R(t) \leq O(c \log t)^{\frac{1}{d+2}} \times t^{\frac{d+1}{d+2}} \quad \textit{for all times } t. \tag{10}$$

The zooming algorithm is *self-tuning* in that it does not input the zooming dimension d. Moreover, it is not parameterized by the multiplier c, and yet it satisfies the corresponding regret bound for any given $c > 0$. For sharper guarantees, c can be tuned to the specific problem instance and specific time t.

Note that the regret bound in Theorem 4.2 has the same "shape" as the worst-case result (Theorem 1.2), except that d now stands for the zooming dimension rather than the covering dimension. Thus, the zooming dimension is our way to quantify the benignness of a problem instance. (It is immediate from Definition 4.1 that the covering dimension with multiplier c is an upper bound on the zooming dimension with the same multiplier.) Let us flesh out (and generalize) two examples from Section 1 where the zooming dimension is small:

- all arms with $\Delta(v) < r$ lie in a low-dimensional region $S \subset X$, for some $r > 0$.
- $\mu(x) = \max(0,\ \mu^* - \mathcal{D}(x, S))$ for some $\mu^* \in (0, 1]$ and subset $S \subset X$.

In both examples, for a sufficiently large constant multiplier c, the zooming dimension is bounded from above by $\mathtt{COV}(S)$ (as opposed to $\mathtt{COV}(X)$). Note that in the second example a natural special case is when S is a finite point set, in which case $\mathtt{COV}(S) = 0$. The technical fine print is very mild: $(X, \mathcal{D})$ can be any compact metric space, and the second example requires some open neighborhood of S to have constant doubling dimension. The first example is immediate; the second example is analyzed in Section 4.3.

Our proof of Theorem 4.2 does not require all the assumptions in the Lipschitz MAB problem. It never uses the triangle inequality, and it only needs a relaxed version of the Lipschitz condition Equation (1). If there exists a unique best arm x^*, then the relaxed Lipschitz condition is Equation (1) with $y = x^*$. In a more efficient notation: $\Delta(x) \leq \mathcal{D}(x, x^*)$ for each arm x. This needs to hold for each best arm x^* if there is more than one. A more general version, not assuming that the optimal payoff $\sup(\mu, X)$ is attained by some arm, is as follows:

$$(\forall \epsilon > 0) \quad (\exists x^* \in X) \quad (\forall x \in X) \quad \Delta(x) \leq \mathcal{D}(x, x^*) + \epsilon. \tag{11}$$

Theorem 4.3. *The guarantees in Theorem 4.2 hold even if the similarity function $\mathcal{D}$ is not required to satisfy the triangle inequality,*[19] *and the Lipschitz condition Equation (1) is relaxed to Equation (11).*

Further, we obtain a regret bound in terms of the covering numbers.

Theorem 4.4. *Fix an instance of the Lipschitz MAB problem (relaxed as in Theorem 4.3). Then the regret $R(t)$ of the zooming algorithm satisfies*

$$R(t) \leq \min_{\rho > 0} \left(\rho t + O(\log^2 t) \sum_{r \in \mathcal{S}:\, r \geq \rho} \frac{1}{r}\, N_{r/8}(X_{\mu,r}) \right), \text{ where } \mathcal{S} = \{2^{-i} : i \in \mathbb{N}\}.$$

This regret bound takes advantage of problem instances for which $X_{\mu,r}$ is a much smaller set than X. It can be useful even if the benignness of the problem instance cannot be summarized via a non-trivial upper-bound on the zooming dimension.

The rest of this section is organized as follows. In Section 4.1, we prove the above theorems. In addition, we provide some extensions and applications.

- In Section 4.2, we derive a regret bound that matches Equation (10) and gets much smaller if the maximal payoff is close to 1. This result relies on an improved confidence radius, which may be of independent interest.
- In Section 4.3, we analyze the special case in which the expected payoff of a given arm is a function of the distance from this arm to the (unknown) "target set" $S \subset X$. This is a generalization of the $\mu(x) = \max(0,\ \mu^* - \mathcal{D}(x, S))$ example above.
- In Section 4.4, we prove improved regret bounds for several examples in which the payoff of each arm x is $\mu(x)$ plus i.i.d. noise of known and "benign" distribution. For these results, we replace $\mu_t(x)$, $r_t(x)$ with better estimates: $\hat{\mu}_t(x)$, $\hat{r}_t(x)$ such that $|\hat{\mu}_t(x) - \mu(x)| \leq \hat{r}_t(x) < r_t(x)$ with high probability.

[19] Formally, we require $\mathcal{D}$ to be a symmetric function $X \times X \to [0, \infty]$ such that $\mathcal{D}(x, x) = 0$ for all $x \in X$. We call such a function a *quasi-distance* on X.

4.1 Analysis of the Zooming Algorithm

First, we use Chernoff bounds to prove Equation (8). A given phase will be called *clean* if for each round t in this phase and each arm $x \in X$, we have $|\mu_t(x) - \mu(x)| \leq r_t(x)$.

Claim 4.5. *Each phase i_{ph} is clean with probability at least $1 - 4^{-i_{\text{ph}}}$.*

Proof. The only difficulty is to set up a suitable application of Chernoff bounds along with the union bound. Let $T = 2^{i_{\text{ph}}}$ be the duration of a given phase i_{ph}.

Fix some arm x. Recall that each time an algorithm plays arm x, the payoff is sampled i.i.d. from some distribution $\mathbb{P}_x$. Define random variables $Z_{x,s}$ for $1 \leq s \leq T$ as follows: for $s \leq n(x)$, $Z_{x,s}$ is the payoff from the sth time arm x is played, and for $s > n(x)$ it is an independent sample from $\mathbb{P}_x$. For each $k \leq T$, we can apply Chernoff bounds to $\{Z_{x,s} : 1 \leq s \leq k\}$ and obtain that

$$\Pr\left[\left|\mu(x) - \frac{1}{k}\sum_{s=1}^{k} Z_{x,s}\right| \leq \sqrt{\frac{8\, i_{\text{ph}}}{1+k}}\right] > 1 - T^{-4}. \tag{12}$$

Let N be the number of arms activated in phase i_{ph}; note that $N \leq T$. Define X-valued random variables $x_1, \ldots, x_T$ as follows: x_j is the $\min(j, N)$th arm activated in this phase. For any $x \in X$ and $j \leq T$, the event $\{x = x_j\}$ is independent of the random variables $\{Z_{x,s}\}$; the former event depends only on payoffs observed before x is activated, while the latter set of random variables has no dependence on payoffs of arms other than x. Therefore, Equation (12) remains valid if we replace the probability on the left side with conditional probability, conditioned on the event $\{x = x_j\}$. Taking the union bound over all $k \leq T$, and using the notation of $\mu_t(x)$ and $r_t(x)$, it follows that

$$\Pr[\forall t\ |\mu(x) - \mu_t(x)| \leq r_t(x)\ \mid\ x_j = x] > 1 - T^{-3},$$

where t ranges over all rounds in phase i_{ph}. Integrating over all arms x, we obtain

$$\Pr[\forall t\ |\mu(x_j) - \mu_t(x_j)| \leq r_t(x_j)] > 1 - T^{-3}.$$

Finally, we obtain the claim by taking the union bound over all $j \leq T$. □

Next, we present a crucial argument that connects the best arm and the arm played at a given round, which in turn allows us to bound the number of plays of a suboptimal arm in terms of its badness.

Lemma 4.6. *If phase i_{ph} is clean, then we have $\Delta(x) \leq 3\, r_t(x)$ for any time t and any arm x.*

Proof. Suppose arm x is played at time t in clean phase i_{ph}. First, we claim that $I_t(x) \geq \mu^*$. Indeed, fix $\epsilon > 0$. By definition of μ^* there exists a arm x^* such that $\Delta(x^*) < \epsilon$. Recall that all arms are covered at all times, so there exists an active arm x_t that covers x^* at time t, meaning that x^* is contained in the confidence ball of x_t. Since arm x was chosen over x_t, we have $I_t(x) \geq I_t(x_t)$. Since this is a clean phase, it follows that $I_t(x_t) \geq \mu(x_t) + r_t(x_t)$. By the Lipschitz property, we have $\mu(x_t) \geq \mu(x^*) - \mathcal{D}(x_t, x^*)$. Since x_t covers x^*, we have $\mathcal{D}(x_t, x^*) \leq r_t(x_t)$ Putting all these inequalities together, we have $I_t(x) \geq \mu(x^*) \geq \mu^* - \epsilon$. Since this inequality holds for an arbitrary $\epsilon > 0$, we in fact have $I_t(x) \geq \mu^*$. Claim proved.

Furthermore, note that by the definitions of "clean phase" and "index," we have

$$\mu^* \leq I_t(x) \leq \mu(x) + 3\, r_t(x),$$

and therefore $\Delta(x) \leq 3\, r_t(x)$.

Now suppose arm x is not played at time t. If it has never been played before time t in this phase, then $r_t(x) > 1$ and thus the lemma is trivial. Else, let s be the last time arm x has been played before time t. Then by definition of the confidence radius $r_t(x) = r_s(x) \geq \frac{1}{3}\,\Delta(x)$. □

Corollary 4.7. *If phase i_{ph} is clean, then each arm x is played at most $O(i_{\text{ph}})\ (\Delta(x))^{-2}$ times.*

Proof. This follows by plugging the definition of the confidence radius into Lemma 4.6. □

Corollary 4.8. *In a clean phase, for any active arms x, y, we have $\mathcal{D}(x,y) > \frac{1}{3}\min(\Delta(x), \Delta(y))$.*

Proof. Assume x has been activated before y. Let s be the time when y has been activated. Then, by the algorithm specification, we have $\mathcal{D}(x,y) > r_s(x)$. By Lemma 4.6 $r_s(x) \geq \frac{1}{3}\Delta(x)$. □

Consider round t, which belongs to a clean phase i_{ph}. Let S_t be the set of all arms that are active at time t, and let

$$A_{(i,t)} = \left\{ x \in S_t \,:\, 2^i \leq \frac{1}{\Delta(x)} < 2^{i+1} \right\}.$$

Recall that by Corollary 4.7 for each $x \in A_{(i,t)}$, we have $n_t(x) \leq O(\log t)\,(\Delta(x))^{-2}$. Therefore,

$$\sum_{x \in A_{(i,t)}} \Delta(x)\, n_t(x) \leq O(\log t) \sum_{x \in A_{(i,t)}} \frac{1}{\Delta(x)} \leq O(2^i\ \log t)\, |A_{(i,t)}|.$$

Letting $r = 2^{-i}$, note that by Corollary 4.8 any set of diameter less than $r/8$ contains at most one arm from $A_{(i,t)}$. It follows that $|A_{(i,t)}| \leq N_{r/8}(X_{\mu,r})$, the smallest number of sets of diameter less than $r/8$ sufficient to cover all arms x such that $\frac{r}{2} < \Delta(x) \leq r$. It follows that

$$\sum_{x \in A_{(i,t)}} \Delta(x)\, n_t(x) \leq O(\log t)\, \frac{1}{r}\, N_{r/8}(X_{\mu,r}).$$

Let $\mathcal{S} = \{2^{-i} \,:\, i \in \mathbb{N}\}$. For each $\rho \in (0,1)$, we have

$$\begin{aligned}\sum_{x \in S_t} \Delta(x)\, n_t(x) &\leq \sum_{x \in S_t:\, \Delta(x) \leq \rho} \Delta(x)\, n_t(x) \ + \sum_{i < \log(1/\rho)} \sum_{x \in A_{(i,t)}} \Delta(x)\, n_t(x) \\ &\leq \rho(t - 2^{i_{\text{ph}}-1}) + O(\log t) \sum_{r \in \mathcal{S}:\, r \geq \rho} \frac{1}{r}\, N_{r/8}(X_{\mu,r}). \end{aligned} \tag{13}$$

Here, $t - 2^{i_{\text{ph}}-1}$ is the number of rounds in phase i_{ph} before and including round t.

Let $R_{\text{ph}}(t)$ be the left-hand side of Equation (13). By Claim 4.5, the probability that phase i_{ph} is non-clean is negligible. Therefore, we obtain the following:

Claim 4.9. *Fix round t and let i_{ph} be the round to which t belongs. Then,*

$$\mathbb{E}[R_{\text{ph}}(t)] \leq \inf_{\rho > 0} \left(\rho(t - 2^{i_{\text{ph}}-1}) + O(\log t) \sum_{r \in \mathcal{S}:\, r \geq \rho} \frac{1}{r}\, N_{r/8}(X_{\mu,r}) \right). \tag{14}$$

We complete the proof as follows. Let t be the current round and let i_{ph} be the current phase. Let $t_i = 2^i$ (for $i < i_{\text{ph}}$) be the last round of each phase $i < i_{\text{ph}}$, and let $t_{i_{\text{ph}}} = t$. Note that regret up to time t can be expressed as

$$R(t) = \sum_{i \leq i_{\text{ph}}} \mathbb{E}\left[\, R_{\text{ph}}(t_i)\, \right].$$

Theorem 4.4 follows by summing up Equation (14) over all phases $i \leq i_{\text{ph}}$.

We derive Theorem 4.3 from Equation (14) as follows. Note that $N_{r/8}(X_{\mu,r}) \leq c\, r^{-d}$ by definition of the zooming dimension d with multiplier $c > 0$. For a given phase i, letting $t_0 = t_i - 2^{i-1}$ and choosing ρ such that $\rho\, t_0 = (\frac{1}{\rho})^{d+1}(c\ \log t)$, we obtain

$$\mathbb{E}[R_{\text{ph}}(t_i)] \leq O(c \log t)^{1/(d+2)} \times t_0^{(d+1)/(d+2)}.$$

We obtain Theorem 4.3 by summing this over all phases $i \leq i_{\text{ph}}$.

4.2 Extension: Maximal Expected Payoff Close to 1

We obtain a sharper regret bound, which matches Equation (10) and gets much smaller if the optimal reward $\mu^* = \sup(\mu, X)$ is close to 1. The key ingredient here is a more elaborate confidence radius:

$$\hat{r}_t(x) \triangleq \frac{\alpha}{1+n_t(x)} + \sqrt{\alpha \, \frac{1-\mu_t(x)}{1+n_t(x)}} \quad \text{for some } \alpha = \Theta(i_{\mathrm{ph}}). \tag{15}$$

The confidence radius in Equation (15) performs as well as $r_t(\cdot)$ (up to constant factors) in the worst case: $\hat{r}_t(x) \le \sqrt{\frac{O(i_{\mathrm{ph}})}{n_t(x)}}$, and gets much better when $\mu_t(x)$ is close to 1: $\hat{r}_t(x) \le \frac{O(i_{\mathrm{ph}})}{n_t(x)}$. Note that the right side of Equation (15) can be computed from the observable data; in particular, it does not require the knowledge of μ^*.

Theorem 4.10. *Consider an instance of the Lipschitz MAB problem, in the relaxed setting of Theorem 4.3. Fix any $c > 0$ and let d be the zooming dimension with multiplier c. Let $\mu^* = \sup(\mu, X)$ be the optimal reward. Then zooming algorithm with confidence radius Equation (15) satisfies, for all times t,*

$$R(t) \le O(c \log^2 t) + O(c \log t)^{\frac{1}{d+1}} \times \max\left(t^{1-\frac{1}{d+1}},\ (1-\mu^*)\, t^{1-\frac{1}{d+2}}\right).$$

Compared to the regret bound in Theorem 4.2, this result effectively reduces the zooming dimension by 1 if μ^* is close to 1 (and $d > 1$). Moreover, regret becomes *polylogarithmic* if $\mu^* = 1$ and $d = 0$.

We analyze the new confidence radius Equation (15) using the following corollary of Chernoff bounds, which, to the best of our knowledge, has not appeared in the literature, and may be of independent interest.

Theorem 4.11. *Consider n i.i.d. random variables $Z_1 \ldots Z_n$ on $[0,1]$. Let Z be their average, and let $\zeta = \mathbb{E}[Z]$. Then for any $\alpha > 0$, letting $r(\alpha, x) = \frac{\alpha}{n} + \sqrt{\frac{\alpha x}{n}}$, we have*

$$\Pr\left[\, |Z - \zeta| < r(\alpha, Z) < 3\, r(\alpha, \zeta)\,\right] > 1 - \left(2^{-\alpha} + 2\, e^{-\alpha/72}\right).$$

Proof. Suppose $\zeta < \frac{\alpha}{6n}$. Then using Chernoff bounds (Theorem 3.3(b)) with $a = \frac{\alpha}{n} > 6\zeta$, we obtain that with probability at least $1 - 2^{-\alpha}$, we have $Z < \frac{\alpha}{n}$, and therefore $|Z - \zeta| < \frac{\alpha}{n} < r(\alpha, Z)$ and

$$|Z - \zeta| < \frac{\alpha}{n} < r(\alpha, Z) < (1+\sqrt{2})\, \frac{\alpha}{n} < 3\, r(\alpha, \zeta).$$

Now, suppose $\zeta \ge \frac{\alpha}{6n}$. Apply Chernoff bounds (Theorem 3.3(a)) with $\delta = \frac{1}{2}\sqrt{\frac{\alpha}{6\zeta n}}$. Thus, with probability at least $1 - 2\, e^{-\alpha/72}$, we have $|Z - \zeta| < \delta\zeta \le \zeta/2$. Plugging in δ,

$$|Z - \zeta| < \frac{1}{2}\sqrt{\frac{\alpha\zeta}{n}} \le \sqrt{\frac{\alpha Z}{n}} \le r(\alpha, Z) < 1.5\, r(\alpha, \zeta). \qquad \square$$

Proof of Theorem 4.10. Let us fix an arm x and time t. Let us use Theorem 4.11 with $n = n_t(x)$ and $\alpha = \Theta(i_{\mathrm{ph}})$ as in (15), setting each random variable X_i equal to 1 minus the reward from the ith time arm x is played in the current phase. Then $\zeta = 1 - \mu(x)$ and $Z = 1 - \mu_t(x)$, so the theorem says that

$$\Pr\left[\, |\mu_t(x) - \mu(x)| < r_t(x) < 3\left(\frac{\alpha}{n_t(x)} + \sqrt{\frac{\alpha\,(1-\mu(x))}{n_t(x)}}\right)\right] > 1 - 2^{\Omega(\alpha)}. \tag{16}$$

We modify the analysis in Section 4.1 as follows. We redefine a "clean phase" to mean that the event in the left-hand side of Equation (16) holds for all rounds t and all arms x. We use Equation (16) instead of the standard Chernoff bound in the proof of Claim 4.5 to show that each phase i_{ph} is clean with probability at least $1 - 4^{i_{\text{ph}}}$. Then, we obtain Lemma 4.6 as is, for the new definition of $r_t(x)$. Then, we replace Corollary 4.7 with a more efficient corollary based on the new $r_t(x)$. More precisely, we derive two regret bounds: one assuming $r_t(x) = \frac{O(i_{\text{ph}})}{n_t(x)}$, and another assuming $r_t(x) = \sqrt{\frac{O(i_{\text{ph}})(1-\mu^*)}{n_t(x)}}$, and take the maximum of the two. We omit the easy details. □

4.3 Application: Lipschitz MAB with a "Target Set"

We consider a version of the Lipschitz MAB problem in which the expected reward of each arm x is determined by the distance between this arm and a fixed *target set* $S \subset X$, which is not revealed to the algorithm. Here, the distance is defined as $\mathcal{D}(x,S) \triangleq \inf_{y\in S} \mathcal{D}(x,y)$. The motivating example is $\mu(x) = \max(0,\, \mu^* - \mathcal{D}(x,S))$. More generally, we assume that $\mu(x) = f(\mathcal{D}(x,S))$ for each arm x, for some known non-increasing function $f : [0,1] \to [0,1]$. We call this version the *Target MAB problem* with target set S and shape function f.[20]

The key idea is to use the quasi-distance function $\mathcal{D}_f(x,y) = f(0) - f(\mathcal{D}(x,y))$. It is easy to see that $\mathcal{D}_f$ satisfies Equation (11). Indeed, fix any arm $x^* \in S$. Then, for each $x \in X$, we have

$$\Delta(x) = \mu(x^*) - \mu(x) = f(0) - f(\mathcal{D}(x,S)) = \mathcal{D}_f(x,S) \le \mathcal{D}_f(x,x^*).$$

Therefore, Theorem 4.3 applies: We can use the zooming algorithm in conjunction with $\mathcal{D}_f$ rather than $\mathcal{D}$. The performance of this algorithm depends on the zooming dimension of the problem instance $(X, \mathcal{D}_f, \mu)$.

Theorem 4.12. *Consider the Target MAB problem with target set $S \subset X$ and shape function f. For some fixed multiplier $c > 0$, let d be the zooming dimension of $(X, \mathcal{D}_f, \mu)$. Then the zooming algorithm on $(\mathcal{D}_f, X)$ has regret $R(t) \le (c\,\log t)^{\frac{1}{d+2}}\;\; t^{\frac{d+1}{d+2}}$ for all times t.*

Note that the zooming algorithm is self-tuning: it does not need to know the properties of S or f, and in fact it does not even need to know that it is presented with an instance of the Target MAB problem. We obtain a further improvement via Theorem 4.10 if $f(0)$ is close to 1.

Let us consider the main example $\mu(x) = \max(0,\, \mu^* - \mathcal{D}(x,S))$ and, more generally,

$$\mu(x) = \max(\mu_0,\, \mu^* - \mathcal{D}(x,S)^{1/\alpha}), \tag{17}$$

for some constant $\alpha > 0$ and $0 \le \mu_0 < \mu^* \le 1$. Here μ_0 and μ^* are, respectively, the minimal and maximal expected payoffs. Equation (17) corresponds to $f(z) = \max(\mu_0,\, \mu^* - z^{1/\alpha})$. Then,

$$\mathcal{D}_f(x,y) = \min(\mu^* - \mu_0,\; (\mathcal{D}(x,y))^{1/\alpha}).$$

We find that the zooming dimension of the problem instance $(X, \mathcal{D}_f, \mu)$ is, essentially, at most α times the covering dimension of S. (This result holds as long as $(X, \mathcal{D})$ has constant doubling dimension.) Intuitively, S is a low-dimensional subset of the metric space, in the sense that it has a (much) smaller covering dimension.

Lemma 4.13. *Consider the Target MAB problem with payoff function given by (17). Let d be the covering dimension of the target set S, for any fixed multiplier $c > 0$. Let d_{DBL} be the doubling dimension of $(X, \mathcal{D})$; assume it is finite. Then the zooming dimension of $(X, \mathcal{D}_f, \mu)$ is αd, with constant*

[20] Note that the payoff function μ does not necessarily satisfy the Lipschitz condition with respect to $\mathcal{D}$. However, if $f(z) = \mu^* - z$ then $\mu(x) = \mu^* - \mathcal{D}(x, S)$, and the Lipschitz condition is satisfied, because $\mathcal{D}(x, S) - \mathcal{D}(y, S) \le \mathcal{D}(x, y)$.

multiplier

$$c_{\mathsf{zoom}} = \left(\max\left(c\, 2^{4\alpha+2},\ \frac{2}{\mu^* - \mu_0}\right)\right)^{d_{\mathsf{DBL}}}.$$

Proof. For each $r > 0$, it suffices to cover the set $S_r = \{x \in X : \Delta(x) \leq r\}$ with $c_{\mathsf{zoom}}\, r^{-\alpha d}$ sets of $\mathcal{D}_f$-diameter at most $r/16$. Note that $\Delta(x) = \min(\mu^* - \mu_0,\ (\mathcal{D}(x,S))^{1/\alpha})$.

Assume $r < \mu^* - \mu_0$. Then for each $x \in S_r$, we have $\mathcal{D}(x,S) \leq r^\alpha$. By definition of the covering dimension, S can be covered with $c\, r^{-\alpha d}$ sets $\{\, C_i\,\}_i$ of $\mathcal{D}$-diameter at most r^α. It follows that S_r can be covered with $r^{-\alpha d}$ sets $\{\, B(C_i, r)\,\}_i$, where $B(C_i, r) \triangleq \cup_{u \in C_i} B(x, r)$. The $\mathcal{D}$-diameter of each such set is at most $3\, r^\alpha$. Since d_{DBL} is the doubling dimension of $(X, \mathcal{D})$, each $B(C_i, r)$ can be covered by with $2^{(4\alpha+2)\, d_{\mathsf{DBL}}}$ of sets of $\mathcal{D}$-diameter at most $(r/16)^\alpha$. So, S_r can be covered by $c\, 2^{(4\alpha+2)\, d_{\mathsf{DBL}}}\, r^{-\alpha d}$ sets whose $\mathcal{D}$-diameter is at most $(r/16)^\alpha$, so that their $\mathcal{D}_f$-diameter is at most $r/16$.

For $r \geq \mu^* - \mu_0$, we have $S_r = X$, and by definition of the doubling dimension X can be covered by $(\frac{2}{\mu^*-\mu_0})^{d_{\mathsf{DBL}}}$ sets of diameter at most $\mu^* - \mu_0$. □

The most striking (and very reasonable) special case is when S consists of finitely many points.

Corollary 4.14. *Consider the Target MAB problem with payoff function given by Equation (17). Suppose the target set S consists of finitely many points. Let c_{zoom} be from Lemma 4.13 with $c = |S|$. Then the zooming algorithm on $(\mathcal{D}_f, X)$ has regret $R(t) = O(\sqrt{c_{\mathsf{zoom}}\, t\, \log t})$ for all times t. Moreover, the regret is $R(t) = O(c_{\mathsf{zoom}} \log t)^2$ if $\mu^* = 1$.*

Proof. The covering dimension of S is 0 with multiplier $c = |S|$. Then by Lemma 4.13 the zooming dimension is 0, with multiplier c_{zoom}. We obtain the $\tilde{O}(\sqrt{c_{\mathsf{zoom}}\, t \log t})$ regret using Theorem 4.12, and the $O(c_{\mathsf{zoom}}\, \log t)^2$ regret result using Theorem 4.10. □

The proof of Lemma 4.13 easily extends to shape functions f such that

$$x^{1/\alpha} \leq f(0) - f(x) \leq x^{1/\alpha'} \quad \forall x \in (0,1],$$

for some constants $\alpha \geq \alpha' > 0$. Then, using the notation in Lemma 4.13, the zooming dimension of $(X, \mathcal{D}_f, \mu)$ is αd, with multiplier $c_{\mathsf{zoom}} = \max(c\, 2^{(4\alpha'+2)\, d_{\mathsf{DBL}}},\ (\frac{2}{\mu^*-\mu_0})^{d_{\mathsf{DBL}}})$.

4.4 Application: Mean-zero Noise with Known Shape

Improved regret bounds are possible if the reward from playing each arm x is $\mu(x)$ plus noise of known shape. More precisely, we assume that the reward from playing arm x is $\mu(x)$ plus an independent random sample from some fixed, mean-zero distribution $\mathcal{P}$, called the *noise distribution*, which is revealed to the algorithm. We call this version the *noisy Lipschitz MAB problem*. We present several examples in which we take advantage of a "benign" shape of $\mathcal{P}$. In these examples, the payoff distributions are not restricted to have bounded support.[21]

Normal distributions. We start with perhaps the most natural example when the noise distribution $\mathcal{P}$ is the zero-mean normal distribution. Then instead of the confidence radius r_t defined by Equation (7), we can use the confidence radius $\hat{r}_t(\cdot) = \sigma\, r_t(\cdot)$, where σ is the standard deviation of $\mathcal{P}$. Consequently, we obtain a regret bound Equation (10) with the right-hand side multiplied by σ.

In fact, this result can be generalized to all noise distributions $\mathcal{P}$ such that

$$\mathbb{E}_{Z \sim \mathcal{P}}[e^{rZ}] \leq e^{r^2\sigma^2/2} \quad \text{for all } r \in [-\rho, \rho]. \tag{18}$$

[21] Recall that throughout the article the payoff distribution of each arm x has support $\mathcal{S}(x) \subset [0,1]$. In this subsection, by a slight abuse of notation, we do not make this assumption.

The normal distribution with standard deviation σ satisfies Equation (18) for $\rho = \infty$. Any distribution with support $[-\sigma, \sigma]$ satisfies (18) for $\rho = 1$. The meaning of Equation (18) is that it is precisely the condition needed to establish an Azuma-type inequality: if $Z_1, \dots, Z_n$ are independent samples from $\mathcal{P}$ then $\sum_{i=1}^n Z_i \leq \tilde{O}(\sigma\sqrt{n})$ with high probability. More precisely,

$$\Pr\left[\sum_{i=1}^n Z_i > \lambda\sigma\sqrt{n}\right] \leq \exp(-\lambda^2/2) \quad \text{for any} \lambda \leq \frac{1}{2}\,\rho\,\sigma\sqrt{n}. \tag{19}$$

We can derive an analog of Claim 4.5 for the new confidence radius $\hat{r}_t(\cdot) = \sigma\, r_t(\cdot)$ by using Equation (19) instead of the standard Chernoff bound; we omit the easy details.

Tool: Generalized confidence radius. More generally, we may be able to use a different, smaller confidence radius $\hat{r}_t(\cdot)$ instead of $r_t(\cdot)$ from Equation (7), perhaps in conjunction with a different estimate $\hat{\mu}_t(\cdot)$ of $\mu(\cdot)$ instead of the sample average $\mu_t(\cdot)$. We will need the pair $(\hat{\mu}_t, \hat{r}_t)$ to satisfy an analog of Claim 4.5:

$$\Pr\left[\ |\hat{\mu}_t(x) - \mu(x)| \leq \hat{r}_t(x) \text{ for all times } t \text{ and arms } x\ \right] \geq 1 - 4^{-i_{\mathrm{ph}}}. \tag{20}$$

Further, we will need the confidence radius $\hat{r}_t$ to be small in the following sense:

$$\text{for each arm } x \text{ and any } r > 0, \text{ inequality } \hat{r}_t(x) \leq r \text{ implies } n_t(x) \leq c_0\, r^{-\beta} \log t, \tag{21}$$

for some constants c_0 and $\beta \geq 0$. Recall that $\hat{r}_t = r_t$ satisfies Equation (21) with $\beta = 2$ and $c_0 = O(1)$.

Lemma 4.15. *Consider the Lipschitz MAB problem (relaxed as in Theorem 4.3). Consider the zooming algorithm with estimator $\hat{\mu}_t$ and confidence radius $\hat{r}_t$, and consider a problem instance such that the pair $(\hat{\mu}_t, \hat{r}_t)$ satisfies Equation (20). Suppose $\hat{r}_t$ satisfies Equation (21). Let d be the zooming dimension of the problem instance, for any fixed multiplier $c > 0$. Then regret of the algorithm is*

$$R(t) \leq O(c\, c_0\, \log^2 t) + O(c\, c_0\, \log^2 t)^{1/(d+\beta)} \times t^{1-1/(d+\beta)} \textit{ for all times } t. \tag{22}$$

Lemma 4.15 is proved by plugging in the improved confidence radius into the analysis in Section 4.1; we omit the easy details. We obtain an improvement over Theorems 4.2 and 4.3 whenever $\beta < 2$. Below, we give some examples for which we can construct improved $(\hat{\mu}_t, \hat{r}_t)$.

Example: Deterministic rewards. For the important special case of deterministic rewards, we obtain regret bound (22) with $\beta = 0$. (The proof is a special case of the next example.)

Corollary 4.16. *Consider the Lipschitz MAB problem with deterministic rewards (relaxed as in Theorem 4.3). Then the zooming algorithm with suitably defined estimator $\hat{\mu}_t$ and confidence radius $\hat{r}_t$ achieve regret bound Equation (22) with $\beta = 0$.*

Example: Noise distribution with a point mass. Consider noise distributions $\mathcal{P}$ having at least one *point mass*: a point $z \in \mathbb{R}$ of positive probability mass: $\mathcal{P}(z) > 0$. (Deterministic rewards correspond to the special case $\mathcal{P}(0) = 1$).

Corollary 4.17. *Consider the Lipschitz MAB problem (relaxed as in Theorem 4.3). Assume mean-zero noise distribution with at least one point mass. Then the zooming algorithm with suitably defined estimator $\hat{\mu}_t$ and confidence radius $\hat{r}_t$ achieve regret bound Equation (22) with $\beta = 0$.*

Proof. We will show that we can use a confidence radius $\hat{r}_t(u) = r_t(u)\, \mathbf{1}_{\{n_t(u) \leq c_{\mathcal{P}} \log t\}}$, for some constant $c_{\mathcal{P}}$ that depends only on $\mathcal{P}$. This implies regret bound Equation (22) with $\beta = 0$.

Indeed, let $p = \max_{z \in \mathbb{R}} \mathcal{P}(z)$ be the largest point mass in distribution $\mathcal{P}$, and $q = \max_{z \in \mathbb{R}:\, \mathcal{P}(z) < p} \mathcal{P}(z)$ be the second largest point mass. Let $S = \{z \in \mathbb{R} : \mathcal{P}(z) = p\}$, and let $k = |S| + \frac{1}{q}$ if $q > 0$, or $k = |S|$ if $q = 0$. Then for some $c_{\mathcal{P}} = \Theta(\frac{\log(|S|+k)}{p-q})$, it suffices to have $n \geq c_{\mathcal{P}} \log t$

independent samples from $\mathcal{P}$ to ensure that with probability at least $1 - t^{-4}$ each number $z \in S$ is sampled at least $n(p+q)/2$ times, whereas any number $z \notin S$ is sampled less often.[22]

For a given arm x and time t, we define a new estimator $\hat{\mu}_t(x)$ as follows. Let $n = n_t(x)$ be the number of rewards from x so far. If $n < c_{\mathcal{P}} \log t$, then use the sample average: Let $\hat{\mu}_t(x) = \mu_t(x)$. Else, let R be the set of rewards that have appeared at least $n(p+q)/2$ times. Then $R = \mu(x) + S$ with probability at least $1 - t^{-4}$. In particular, $\max(R) = \mu(x) + \max(S)$. So, we can define $\hat{\mu}_t(x) = \max(R) - \max(S)$. □

Example: Noise distributions with a sharp peak. If the noise distribution $\mathcal{P}$ has a sharp peak around 0, then small regions around this peak can be identified more efficiently than using the standard confidence radius r_t.

More precisely, suppose $\mathcal{P}$ has a probability density function $f(z)$, which is symmetric around 0 and non-increasing for $z > 0$, and suppose $f(z)$ has a sharp peak: $f(z) = \Theta(|z|^{-\alpha})$ on some open neighborhood of 0, for some constant $\alpha \in (0,1)$. We will show that we can use a new confidence radius $\hat{r}_t(x) = C\,(i_{\text{ph}}/n_t(x))^{1/(1-\alpha)}$, for a sufficiently high constant C, which leads to regret bound Equation (22) with $\beta = 1 - \alpha$.

Fix arm x and time t. We define the estimator $\hat{\mu}_t(x)$ as follows. Let S be the multiset of rewards received from arm x so far. Let $r = \frac{1}{2}\,\hat{r}_t(x)$. Cover the $[0,1]$ interval with $\lceil 1/r \rceil$ subintervals $I_j = [jr,\ (j+1)r)$. Pick the subinterval that has most points from S (break ties arbitrarily), and define $\hat{\mu}_t(x)$ as some point in this subinterval.

Let us show that $|\mu(x) - \hat{\mu}_t(x)| \le \hat{r}_t(x)$ with high probability. Let I_j be the subinterval that contains $\mu(x)$. Let $n = n_t(x)$ be the number of times arm x has been played so far; note that $n > \Omega(C\,r^{\alpha-1} \log t)$. By Chernoff bounds, for a sufficiently high constant C, it holds that with probability at least $1 - t^{-4}$ subinterval I_j contains more points from S than any other subinterval I_ℓ such that $|j - \ell| \ge 2$. Conditional on this high-probability event, the estimate $\hat{\mu}_t(x)$ lies in subinterval I_ℓ such that $|j - \ell| \le 1$, which implies that $|\mu(x) - \hat{\mu}_t(x)| \le 2r$.

5 OPTIMAL PER-METRIC PERFORMANCE

This section is concerned with Question (Q1) raised in Section 1.3: What is the best possible algorithm for the Lipschitz MAB problem on a given metric space $(X, \mathcal{D})$. We consider the worst-case regret of a given algorithm over all possible problem instances on $(X, \mathcal{D})$.[23] We focus on minimizing the exponent γ such that for each payoff function μ the algorithm's regret is $R(t) \le t^\gamma$ for all $t \ge t_0(\mu)$. With Theorem 1.2 in mind, we will use a more focused notation: We define the *regret dimension* of an algorithm on $(X, \mathcal{D})$ as, essentially, the smallest $d \ge 0$ such that one can achieve the exponent $\gamma = \frac{d+1}{d+2}$.

Definition 5.1. Consider the Lipschitz MAB problem on a given metric space $(X, \mathcal{D})$. For algorithm $\mathcal{A}$ and payoff function μ, define the *instance-specific regret dimension* of $\mathcal{A}$ as

$$\begin{aligned}\mathtt{DIM}_\mu(\mathcal{A}) &= \inf\{d \ge 0 \quad | \quad \exists t_0 = t_0(\mu) \quad R_{\mathcal{A}}(t) \le t^{1-1/(d+2)} \quad \text{for all } t \ge t_0\} \\ &= \inf\{d \ge 0 \quad | \quad \exists C = C(\mu) \quad R_{\mathcal{A}}(t) \le C\,t^{1-1/(d+2)} \quad \text{for all } t\}.\end{aligned}$$

The *regret dimension* of $\mathcal{A}$ is $\mathtt{DIM}(\mathcal{A}) = \sup_\mu \mathtt{DIM}_\mu(\mathcal{A})$, where the supremum is over all payoff functions μ.

[22] To prove that each number $z \notin S$ is sampled less than $n(p+q)/2$ times when $q > 0$, we need to be somewhat careful in how we apply the Union Bound. It is possible to partition the set $\mathbb{R} \setminus S$ into at most $O(|S| + \frac{1}{q})$ measurable subsets, namely intervals or points, whose measure is at most q (and at least $q/2$). Apply Chernoff bound to each subset separately, then take the Union Bound.

[23] Formally, we can define the *per-metric performance* of an algorithm on a given metric space as the worst-case regret of this algorithm over all problem instances on this metric space.

Thus, according to Theorem 1.2, the regret dimension of `UniformMesh` is at most the covering dimension of the metric space. We ask: ***Is it possible to achieve a better regret dimension***, perhaps using a more sophisticated algorithm? We show that this is indeed the case. Moreover, we provide an algorithm such that for any given metric space its regret dimension is arbitrarily close to optimal. Our main result as follows:

Theorem 5.2. *Consider the Lipschitz MAB problem on a compact metric space* $(X, \mathcal{D})$. *Then for any* $d > \texttt{MaxMinCOV}(X)$ *then there exists a bandit algorithm* $\mathcal{A}$ *whose regret dimension is at most* d*; moreover, the instance-specific regret dimension of* $\mathcal{A}$ *is at most the zooming dimension. No algorithm can have regret dimension strictly less than* $\texttt{MaxMinCOV}(X)$.

Here $\texttt{MaxMinCOV}(X)$ is the *max-min-covering dimension,* which we defined in Section 1. We show that $\texttt{MaxMinCOV}(X)$ can be arbitrarily small compared to $\texttt{COV}(X)$.

The rest of this section is organized as follows. The first two subsections are concerned with the lower bound: In Section 5.1, we develop a lower bound on regret dimension that relies on a certain "tree of balls" structure, and in Section 5.2, we derive the existence of this structure from the max-min-covering dimension. A lengthy KL-divergence argument (which is similar to prior work) is deferred to Section A. The next two subsections deal with an instructive special case: In Section 5.3, we define a family of metric spaces for which $\texttt{MaxMinCOV}(X)$ can be arbitrarily small compared to $\texttt{COV}(X)$, and in Section 5.4, we design a version of the zooming algorithm tailored to such metric spaces. Finally, in Section 5.5, we design and analyze an algorithm whose regret dimension is arbitrarily close to $\texttt{MaxMinCOV}(X)$. We use the max-min-covering dimension to derive the existence of a certain decomposition of the metric space, which we then take advantage of algorithmically. Our per-metric optimal algorithm builds on the machinery developed for the special case. Collectively, these results amount to Theorem 5.2.

5.1 Lower Bound on Regret Dimension

It is known [12] that a worst-case instance of the K-armed bandit problem consists of $K - 1$ arms with identical payoff distributions, and one that is slightly better. We refer to this as a "needle-in-haystack" instance. Our lower bound relies on a *multi-scale* needle-in-haystack instance in which there are K disjoint open sets, and $K - 1$ of them consist of arms with identical payoff distributions, but in the remaining open set there are arms whose payoff is slightly better. Moreover, this special open set contains $K' \gg K$ disjoint subsets, only one of which contains arms superior to the others, and so on down through infinitely many levels of recursion.

In more precise terms, we require the existence of a certain structure: an infinitely deep rooted tree whose nodes correspond to balls in the metric space, so that for any parent ball B the children balls are disjoint subsets of B.

Definition 5.3 (Ball-tree). Fix a metric space $(X, \mathcal{D})$. Let an *extensive-form ball* be a pair $w = (x, r)$, where $x \in X$ is a "center" and $r \in (0, 1]$ is a "radius."[24] A *ball-tree* is an infinite rooted tree where each node corresponds to an extensive-form ball. The following properties are required:

- all children of the same parent have the same radius, which is at most a quarter of the parent's.
- if (x, r) is a parent of (x', r') then $\mathcal{D}(x, x') + r' < r/2$.
- if (x, r_x) and (y, r_y) are siblings, then $r_x + r_y < \mathcal{D}(x, y)$.

[24] Note that an open ball $B(x, r)$ denotes a subset of the metric space, so there can be distinct extensive-form balls (x, r) and (x', r') such that $B(x, r) = B(x', r')$. We use extensive-form balls to avoid this ambiguity.

The ball-tree has *strength* $d \geq 0$ if each tree node with children of radius r has at least $\max(2, r^{-d})$ children.

Once there exists a ball-tree of strength d, we can show that, essentially, regret $O(t^{1-1/(d+2)})$ is the best possible. More precisely, we construct a probability distribution over problem instances, which is hard for every given algorithm. Intuitively, this is the best possible "shape" of a regret bound, since, obviously, a single problem instance cannot be hard for every algorithm.

Lemma 5.4. *Consider the Lipschitz MAB problem on a metric space $(X, \mathcal{D})$ such that there exists a ball-tree of strength $d \geq 0$. Assume 0-1 payoffs (i.e., the payoff of each arm is either 1 or 0). Then there exist a distribution $\mathcal{P}$ over problem instances μ and an absolute constant $C > 0$ such that for any bandit algorithm $\mathcal{A}$ the following holds:*

$$\Pr_{\mu \in \mathcal{P}} \left[R_{(\mathcal{A},\, \mu)}(t) \geq C\, t^{1-1/(d+2)} \text{ for infinitely many } t \right] = 1. \tag{23}$$

It follows that the regret dimension of any algorithm is at least d.

For our purposes, a weaker version of Equation (23) suffices: for any algorithm $\mathcal{A}$ there exists a payoff function μ such that the event in Equation (23) holds (which implies $\mathtt{DIM}(\mathcal{A}) \geq d$). In Section 6, we will also use this lower bound for $d = 0$. In the rest of this subsection, we prove Lemma 5.4.

Randomized problem instance. Given a metric space $(X, \mathcal{D})$ with a ball-tree, we construct a distribution $\mathcal{P}$ over payoff functions as follows. For each tree node $w = (x_0, r_0)$ define the *bump function* $F_w : X \to [0, 1]$ by

$$F_w(x) = \begin{cases} \min\{r_0 - \mathcal{D}(x, x_0),\ r_0/2\} & \text{if } x \in B(x_0, r_0), \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

This function constitutes a "bump" supported on $B(x_0, r_0)$.

An *end* in a ball-tree is an infinite path from the root: $\mathbf{w} = (w_0, w_1, w_2, \ldots)$. Let us define the payoff function induced by each node $w = w_j$ as

$$\mu_w = \frac{1}{3} + \frac{1}{3} \sum_{i=1}^{j} F_{w_i},$$

and the payoff function induced by the end $\mathbf{w}$ as

$$\mu_{\mathbf{w}} := \lim_{j \to \infty} \mu_{w_j} = \frac{1}{3} + \frac{1}{3} \sum_{i=1}^{\infty} F_{w_i}.$$

Let $\mathcal{P}$ be the distribution over payoff functions $\mu_{\mathbf{w}}$ in which the end $\mathbf{w}$ is sampled uniformly at random from the ball-tree (that is, w_0 is the root, and each subsequent node w_{i+1} is sampled independently and uniformly at random among the children of w_i).

Let us show that $\mu_{\mathbf{w}}$ is a valid payoff function for the Lipschitz MAB problem. First, $\mu_{\mathbf{w}}(x) \in [0, 1]$ for each arm $x \in X$, and the sum in the definition of $\mu_{\mathbf{w}}$ converges, because for each $i \geq 1$, letting r_i be the radius of w_i, we have $r_i \leq r_1/4^i$ and $F_{w_i}(x) \in [0,\ r_i]$. In fact, it is easy to see that the payoff function induced by any node or end in the ball-tree is bounded on $[\frac{1}{3}, \frac{2}{3}]$. Second, $\mu_{\mathbf{w}}$ is Lipschitz on $(X, \mathcal{D})$ due to Lemma 5.8, which we state and prove in Section 5.1.1 so as not to break the flow.

The salient property of our construction is as follows.

Lemma 5.5. *Consider a tree node u in a ball-tree. Let $u_1, \ldots, u_k$ be the children of u, and let r be their radius. Let $B_1, \ldots, B_k$ be the corresponding balls. Fix an arbitrary child u_i, and let $\mathbf{w}$ be an arbitrary end in the ball-tree such that $u_i \in \mathbf{w}$. Then:*

(i) $\mu_{\mathbf{w}}$ *coincides with* μ_u *on all* B_ℓ, $\ell \neq i$.
(ii) $\sup(\mu_{\mathbf{w}}, B_i) - \sup(\mu_u, X) \geq r/6$.
(iii) $0 \leq \mu_i - \mu_u \leq r/3$.

Proof. Let $\mathbf{w} = (w_0, w_1, \ldots)$, and let j_0 be the depth of u_i in the ball-tree. Then $u = w_{j_0-1}$ and $u_i = w_{j_0}$, and $\mu_{\mathbf{w}} = \mu_u + \frac{1}{3} F_{u_i} + \frac{1}{3} \sum_{j > j_0} F_{w_j}$.

Let B be the ball corresponding to u. Observe that the balls corresponding to tree nodes w_j, $j > j_0$ are contained in B_i. It follows that on $B \setminus B_i$ all functions F_{w_j}, $j \geq j_0$ are identically 0, and consequently $\mu_{\mathbf{w}} = \mu_u$. Since the balls $B_1, \ldots, B_k$ are pairwise disjoint, this implies part (i).

For part (ii), let (x^*, r^*) be the extensive-form ball corresponding to tree node u. Note that μ_u attains its supremum on $B(x^*, r^*/2)$, and in fact is a constant on that set. Also, recall that $B_i \subset B(x^*, r^*/2)$ by definition of the ball-tree. Let x_i be the center of u_i. Observe that on $B(x_i, r/2)$, we have $F_{u_i} = r/2$, and therefore $\mu_{\mathbf{w}} \geq \mu_u + r/6$.

For part (iii), note that $\mu_i - \mu_0 = \sum_{j \geq j_0} F_{w_j}$, and the latter is at most $\sum_{j \geq j_0} \frac{1}{2}\, r/4^{j-j_0} \leq r/3$. □

Below, we use Lemma 5.5 to derive a lower bound on regret.

Regret lower bounds via (ϵ, k)-ensembles. We make use of the lower-bounding technique from Auer et al. [12] for the basic k-armed bandit problem. For a cleaner exposition, we encapsulate the usage of this technique in a theorem. This theorem is considerably more general than the original lower bound in Reference [12], but the underlying idea and the proof are very similar. The theorem formulation (mainly, Definition 5.6 below) is new and may be of independent interest.

We use a very general MAB setting where the algorithm is given a strategy set X and a collection $\mathcal{F}$ of feasible payoff functions; we call it the *feasible MAB problem* on $(X, \mathcal{F})$. In our construction, $\mathcal{F}$ consists of all functions $\mu : X \to [0, 1]$ that are Lipschitz with respect to the metric space. The lower bound relies on the existence of a collection of subsets of $\mathcal{F}$ with certain properties, as defined below. These subsets correspond to children of a given tree node in the ball-tree (we give a precise connection after we state the definition).

Definition 5.6. Let X be the strategy set and $\mathcal{F}$ be the set of all feasible payoff functions. An (ϵ, k)*-ensemble* is a collection of subsets $\mathcal{F}_1, \ldots, \mathcal{F}_k \subset \mathcal{F}$ such that there exist mutually disjoint subsets $S_1, \ldots, S_k \subset X$ and a function $\mu_0 : X \to [\frac{1}{3}, \frac{2}{3}]$ such that for each $i = 1 \ldots k$ and each function $\mu_i \in \mathcal{F}_i$ the following holds: (i) $\mu_i \equiv \mu_0$ on each S_ℓ, $\ell \neq i$, and (ii) $\sup(\mu_i, S_i) - \sup(\mu_0, X) \geq \epsilon$, and (iii) $0 \leq \mu_i - \mu_0 \leq 2\epsilon$ on S_i.

For each tree node u, let $\mathcal{F}(u) = \{\mu_{\mathbf{w}} : u \in \mathbf{w}\}$ be the set of all payoff functions induced by ends $\mathbf{w}$ that contain u. By Lemma 5.5, if $u_1, \ldots, u_k$ are siblings whose radius is r, then $(\mathcal{F}(u_1), \ldots, \mathcal{F}(u_k))$ form an $(\frac{r}{6}, k)$-ensemble.

Theorem 5.7. *Consider the feasible MAB problem with 0-1 payoffs. Let* $\mathcal{F}_1, \ldots, \mathcal{F}_k$ *be an* (ϵ, k)*-ensemble, where* $k \geq 2$ *and* $\epsilon \in (0, \frac{1}{12})$. *Then for any* $t \leq \frac{1}{128}\, k\, \epsilon^{-2}$ *and any bandit algorithm there exist at least* $k/2$ *distinct* i*'s such that the regret of this algorithm on any payoff function from* $\mathcal{F}_i$ *is at least* $\frac{1}{60}\, \epsilon t$.

The idea is that if the payoff function μ lies in $\cup_i \mathcal{F}_i$, an algorithm needs to play arms in S_i for at least $\Omega(\epsilon^{-2})$ rounds to determine whether μ lies in a given $\mathcal{F}_i$, and each such round incurs regret at ϵ (or more) if $\mu \notin \mathcal{F}_i$.

Auer et al. [12] analyzed a special case in which there are k arms $x_1, \ldots, x_k$, and each $\mathcal{F}_i$ consists of a single payoff function that assigns expected payoff $\frac{1}{2} + \epsilon$ to arm x_i, and $\frac{1}{2}$ to all other arms. To preserve the flow of the article, the proof of Theorem 5.7 is presented in Appendix A.

Regret analysis. Let us fix a bandit algorithm $\mathcal{A}$, and let T the ball-tree of strength $d \geq 0$. Without loss of generality, let us assume that each tree node in T has finitely many children. Recall that each end of T induces a payoff function. Let $\mathcal{F}_T$ be the set of all payoff functions induced by the ends of T. Throughout, the constants in $\Omega(\cdot)$ are absolute.

Consider a level-j tree node u in T. Let $u_1, \ldots, u_k$ be the children of u, and let r be their radius. Recall that $k \geq \max(2, r^{-d})$. Then $\{\mathcal{F}(u_i) : 1 \leq i \leq k\}$ is a $(\frac{r}{6}, k)$-ensemble. By Theorem 5.7 there exist a subset $I_u \subset \{1, \ldots, k\}$ and a time $t > \Omega(k\, r^{-2})$ such that for any payoff function $\mu \in \mathcal{F}(u_i)$, $i \in I_u$, we have $R_{(\mathcal{A},\, \mu)}(t) \geq \Omega(rt)$. Plugging in $k \geq r^{-d}$ and $r \leq 4^{-j}$, we see that there exists a time $t \geq 2^{\Omega(j)}$ such that for each payoff function $\mu \in \mathcal{F}(u_i)$, $i \in I_u$, we have $R_{(\mathcal{A},\, \mu)}(t) \geq \Omega(t^{1-1/(d+2)})$.

Consider the distribution $\mathcal{P}$ over payoff functions $\mathcal{F}_T$ from our construction. Let $\mathcal{E}_j$ be the event that $\mu \in \mathcal{F}(u_i), i \in I_u$ for some level-j tree node u. If $\mu \in \mathcal{E}_j$ for infinitely many j's, then $R_{(\mathcal{A},\, \mu)}(t) \geq \Omega(t^{1-1/(d+2)})$ for infinitely many times t. We complete the proof of Lemma 5.4 by showing that

$$\Pr_{\mu \sim \mathcal{P}} [\mu \in \mathcal{E}_j \text{ for infinitely many } j] = 1. \tag{25}$$

The proof of Equation (25) is similar to the proof of the Borel-Cantelli Lemma. If $\mu \in \mathcal{E}_j$ only for finitely many j's, then $\mu \in \cap_{j \geq j_0} \neg\mathcal{E}_j$ for some j_0. Fix some $j_0 \in \mathbb{N}$, and let us show that $\Pr[\cap_{j \geq j_0} \neg\mathcal{E}_j] = 0$. For each tree node u at level j, we have $\Pr[\mathcal{E}_j \mid \mu \in \mathcal{F}(u)] \geq \frac{1}{2}$. It follows that for any $j > j_0$

$$\Pr[\neg\mathcal{E}_j \mid \neg\mathcal{E}_{j_0}, \ldots, \neg\mathcal{E}_{j-1}] \leq \frac{1}{2}.$$

Therefore, $\Pr[\cap_{j \geq j_0} \neg\mathcal{E}_j] = \Pr[\neg\mathcal{E}_{j_0}] \times \prod_{j > j_0} \Pr[\neg\mathcal{E}_j \mid \neg\mathcal{E}_{j_0}, \ldots, \neg\mathcal{E}_{j-1}] = 0$; claim proved.

5.1.1 *Lipschitz-continuity of the Lower-bounding Construction.*

In this subsection, we prove that $\mu_{\mathbf{w}}$ is Lipschitz function on $(X, \mathcal{D})$, for any end $\mathbf{w}$ of the ball-tree. In fact, we state and prove a more general lemma in which the bump functions are summed over all tree nodes with arbitrary weights in $[-1, 1]$. This lemma will also be used for the lower-bounding constructions in Section 6.2 and Section 8.3.1.

Lemma 5.8. *Consider a ball-tree on a metric space $(X, \mathcal{D})$. Let V be the set of all tree nodes. For any given weight vector $\sigma : V \to [-1, 1]$ and an absolute constant $c_0 \in [0, \frac{1}{2}]$ define the payoff function*

$$\mu_\sigma = c_0 + \frac{1}{3} \sum_{w \in V} \sigma(w) \cdot F_w,$$

where F_w is the bump function from Equation (24). Then μ_σ is Lipschitz on $(X, \mathcal{D})$.

In the rest of this subsection, we prove Lemma 5.8. (The proof for the special case $\mu_\sigma = \mu_{\mathbf{w}}$ uses essentially the same ideas and does not get much simpler.)

Let us specify some notation. Throughout, u, v, w denote tree nodes. Write $u \vdash w$ if u is a parent of w, and $u \succ w$ if u is an ancestor of w. (Accordingly, define $\dashv$ and $\prec$ relations.) Generally our convention will be that $u \succ w \succ v$. Let x_w and r_w be, respectively, the center and the radius of w, and let $B_w = B(x_w, r_w)$ denote the corresponding ball. Fix the weight vector $\sigma : V \to [-1, 1]$ and arms $x, y \in X$. Write $\mu = \mu_\sigma$ for brevity. We need to prove that $|\mu(x) - \mu(y)| \leq \mathcal{D}(x, y)$.

We start with some observations about the bump functions. First, $F_w(y) \leq \mathcal{D}(x, y)$ under appropriate conditions.

Claim 5.9. *If $y \in B_w$ and $x \notin B_w$, then $F_w(y) \leq \mathcal{D}(x, y)$.*

Proof. Observe that

$$\begin{aligned} F_w(y) &\le r_w - \mathcal{D}(x_w, y) && \text{(because } y \in B_w\text{)} \\ &\le \mathcal{D}(x_w, x) - \mathcal{D}(x_w, y) && \text{(because } x \notin B_w\text{)} \\ &\le \mathcal{D}(x, y) && \text{(by triangle inequality).} \end{aligned}$$ □

Second, each bump function F_w is Lipschitz.

Claim 5.10. $|F_w(x) - F_w(y)| \le \mathcal{D}(x, y)$.

Proof. If $x, y \notin B_w$, then $F_w(x) = F_w(y) = 0$, and we are done. If $x \notin B_w$, but $y \in B_w$, then $F_w(x) = 0$ and $F_w(y) \le \mathcal{D}(x, y)$ by Claim 5.9, and we are done. In what follows, assume $x, y \in B_w$.

We consider four cases, depending on whether $\mathcal{D}(x, x_w)$ and $\mathcal{D}(y, x_w)$ are larger than $r_w/2$. If both $\mathcal{D}(x, x_w)$ and $\mathcal{D}(y, x_w)$ are at least $r_w/2$, then $F_w(x) = F_w(y) = r_w/2$, and we are done. If both $\mathcal{D}(x, x_w)$ and $\mathcal{D}(y, x_w)$ are at most $r_w/2$, then $F_w(x) - F_w(y) = \mathcal{D}(y, x_w) - \mathcal{D}(x, x_w)$, which is at most $\mathcal{D}(x, y)$ by triangle inequality, and we are done. If $\mathcal{D}(x, x_w) \le r_w/2 \le \mathcal{D}(y, x_w)$, then

$$\begin{aligned} F_w(x) - F_w(y) = r_w/2 - (r_w - \mathcal{D}(y, x_w)) &= \mathcal{D}(y, x_w) - r_w/2 \\ &\le \mathcal{D}(y, x_w) - \mathcal{D}(x, x_w) \le \mathcal{D}(x, y). \end{aligned}$$

The fourth case is treated similarly. □

Third, we give a convenient upper bound for $F_w(y) - F_w(x)$, assuming $x \in B_w$.

Claim 5.11. *Assume $x \in B_w$. Then $F_w(y) - F_w(x) \le \max(0,\ \mathcal{D}(x, x_w) - r_w/2)$.*

Proof. This is because $F_w(y) \le r_w/2$ and $F_w(x) = \min(r_w/2, r_w - \mathcal{D}(x, x_w))$. □

Some of the key arguments are encapsulated below. First, F_u is constant on B_w, $u \succ w$.

Claim 5.12. *$F_u(x) = r_u/2$ whenever $x \in B_w$ and $u \succ w$.*

Proof. Since $B_w \supset B_v$ whenever $w \succ v$, it suffices to assume that u is a parent of w. Then

$$\begin{aligned} \mathcal{D}(x, x_u) &\le \mathcal{D}(x_w, x_u) + \mathcal{D}(x, x_w) && \text{(by triangle inequality)} \\ &< \mathcal{D}(x_w, x_u) + r_w && \text{(since } x \in B_w\text{)} \\ &< r_u/2 && \text{(by definition of ball-tree).} \end{aligned}$$ □

Second, suppose B_w separates x and y (in the sense that B_w contains y but not x), and $\mathcal{D}(x, y)$ is small compared to r_w. We show that $y \notin B_v$, $w \vdash v$.

Claim 5.13. *Assume $y \in B_w$ and $x \notin B_w$. Then $y \notin B_v$, whenever $w \vdash v$ and $\mathcal{D}(x, y) \le r_w/2$.*

Proof. Observe that

$$\begin{aligned} \mathcal{D}(x_v, y) &\ge \mathcal{D}(x_w, y) - \mathcal{D}(x_w, x_v) && \text{(by triangle inequality)} \\ &\ge \mathcal{D}(x_w, x) - \mathcal{D}(x, y) - \mathcal{D}(x_w, x_v) && \text{(by triangle inequality)} \\ &\ge r_w - \mathcal{D}(x, y) - \mathcal{D}(x_w, x_v) && \text{(because } x \notin B_w\text{)} \\ &\ge r_v + r_w/2 - \mathcal{D}(x, y). \end{aligned}$$

The last inequality follows, because $r_w/2 - \mathcal{D}(x_w, x_v) \ge r_v$ by definition of ball-tree. It follows that $\mathcal{D}(x_v, y) \ge r_v$ whenever $\mathcal{D}(x, y) \le r_w/2$. □

Claim 5.14. *Assume $w \vdash v$ and $x \in B_w \setminus B_v$ and $y \in B_v$. Then $F_w(y) - F_w(x) + F_v(y) \le \mathcal{D}(x, y)$.*

Proof. If $F_w(y) \leq F_w(x)$, then it suffices to observe that $F_v(y) \leq \mathcal{D}(x,y)$ by Claim 5.9. From here on, assume $F_w(y) > F_w(x) \geq 0$. Observe that

$$\begin{aligned} F_w(y) - F_w(x) &\leq \mathcal{D}(x, x_w) - r_w/2 && \text{(by Claim 5.11)} \\ F_v(y) &\leq r_v - \mathcal{D}(x_v, y) && \text{(by definition of } F_v\text{)} \\ 0 &\leq r_w/2 - r_v - \mathcal{D}(x_w, x_v) && \text{(by definition of ball-tree).} \end{aligned}$$

Summing this up,

$$\begin{aligned} F_w(y) - F_w(x) + F_v(y) &\leq \mathcal{D}(x, x_w) - \mathcal{D}(x_w, x_v) - \mathcal{D}(x_v, y) \\ &\leq \mathcal{D}(x, x_w) - \mathcal{D}(x_w, y) && \text{(by triangle inequality)} \\ &\leq \mathcal{D}(x, y) && \text{(by triangle inequality)} \quad \square \end{aligned}$$

Now, we are ready to put the pieces together. Let w be the least common ancestor of x and y in the ball-tree, i.e., the smallest tree node w such that $x, y \in B_w$. (Such w exists because the radii of the tree nodes go to zero along any end of the ball-tree.) Observe that:

- $F_u(x) = F_u(y) = r_u/2$ for all $u \succ w$ (by Claim 5.12).
- $F_{w'}(x) = F_{w'}(y) = 0$ for all tree nodes w' incomparable with w, because $x, y \notin B_{w'}$.

Therefore,

$$\mu(y) - \mu(x) = \frac{1}{3}\left(\sum_{v \preceq w} \sigma(v)\, F_v(x)\right). \tag{26}$$

Let w_x (respectively, w_y) be the unique child containing x (respectively, y) if such child exists, and an arbitrary child of w otherwise. By minimality of w and the fact that w has at least two children, we can pick w_x and w_y so that they are distinct. Then:

- $F_v(x) = 0$ for all tree nodes $v \preceq w_y$, because $x \notin B_v$.
- $F_v(y) = 0$ for all tree nodes $v \preceq w_x$, because $y \notin B_v$.

Plugging these observations into Equation (26), we obtain

$$\mu(y) - \mu(x) = \frac{1}{3}\left(\sigma(w)\,(F_w(y) - F_w(x)) + \sum_{v \preceq w_x} \sigma(v)\, F_v(x) + \sum_{v \preceq w_y} \sigma(v)\, F_v(y)\right).$$

$$|\mu(y) - \mu(x)| \leq \frac{1}{3}\left(|F_w(y) - F_w(x)| + \sum_{v \preceq w_x} F_v(x) + \sum_{v \preceq w_y} F_v(y).\right). \tag{27}$$

Note that Equation (27) no longer depends on the weight vector σ.

To complete the proof, it suffices to show the following:

$$|F_w(y) - F_w(x)| + \sum_{v \preceq w_x} F_v(x) \leq \frac{4}{3}\,\mathcal{D}(x,y), \tag{28}$$

$$|F_w(y) - F_w(x)| + \sum_{v \preceq w_y} F_v(y) \leq \frac{4}{3}\,\mathcal{D}(x,y). \tag{29}$$

In the remainder of the proof, we show Equation (29) (and Equation (28) follows similarly). Let Γ denote the left-hand side of Equation (29). If $y \notin B_{w_y}$, then $F_v(y) = 0$ for all $v \preceq w_y$, and $\Gamma \leq \mathcal{D}(x,y)$ by Claim 5.10. From here on, assume $y \in B_{w_y}$. Then by Claim 5.12, we have $F_w(y) = r_w/2$. It follows that

$$\Gamma = F_w(y) - F_w(x) + \sum_{v \preceq w_y} F_v(y) \leq \mathcal{D}(x,y) + \sum_{v \prec w_y} F_v(y), \tag{30}$$

where the last inequality follows from Claim 5.14.

Now consider two cases, depending on whether $\mathcal{D}(x,y) \leq r_{w_y}/2$. If so, then $y \notin B_v$ for any $v \prec w_y$ by Claim 5.13, and therefore $F_v(y) = 0$ for all such v, and we are done.

The remaining case is that $\mathcal{D}(x,y) > r_{w_y}/2$. Define the sequence of tree nodes $(v_j : j \in \mathbb{N})$ inductively by $v_0 = w_y$ and for each $j \in \mathbb{N}$ letting v_{j+1} be the child of v_j that contains y, if such child exists, and any child of v_j otherwise. Then

$$F_{v_j}(y) \leq r_{v_j}/2 \leq 4^{-j}\, r_{v_0}/2 < 4^{-j}\, \mathcal{D}(x,y), \qquad \forall j \geq 1,$$

$$\sum_{v \prec w_y} F_v(y) = \sum_{j=1}^{\infty} F_{v_j}(y) \leq \sum_{j=1}^{\infty} 4^{-j}\, \mathcal{D}(x,y) = \mathcal{D}(x,y)/3.$$

Plugging this into Equation (30) completes the proof of Equation (29), which in turn completes the proof of Lemma 5.8.

5.2 The Max-min-covering Dimension

We would like to derive the existence of a strength-d ball-tree using a covering property similar to the covering dimension. We need a more nuanced notion, which we call the max-min-covering dimension, to ensure that *each* of the open sets arising in the construction of the ball-tree has sufficiently many disjoint subsets to continue to the next level of recursion.[25] Further, this new notion is an intermediary that connects our lower bound with the upper bound that we develop in the forthcoming subsections.

Definition 5.15. For a metric space $(X, \mathcal{D})$ and subsets $Y \subseteq X$, we define

$$\begin{aligned} \mathtt{MinCOV}(Y) &= \inf\{\mathtt{COV}(U) : U \subset Y \text{ is non-empty and open in } (Y, \mathcal{D})\}, \\ \mathtt{MaxMinCOV}(X) &= \sup\{\mathtt{MinCOV}(Y) \ : \ Y \subseteq X\}. \end{aligned}$$

We call them the *min-covering dimension* and the *max-min-covering dimension* of X, respectively.

The infimum over open $U \subseteq Y$ in the definition of min-covering dimension ensures that every open set that may arise in the needle-in-haystack construction described above will contain $\Omega(\delta^{\epsilon-d})$ disjoint δ-balls for some sufficiently small positive δ, ϵ. Constructing lower bounds for Lipschitz MAB algorithms in a metric space X only requires that X should have *subsets* with large min-covering dimension, which explains the supremum over subsets in the definition of max-min-covering dimension. Note that for every subset $Y \subseteq X$, we have $\mathtt{MinCOV}(Y) \leq \mathtt{COV}(Y) \leq \mathtt{COV}(X)$, which implies $\mathtt{MaxMinCOV}(X) \leq \mathtt{COV}(X)$.

Lemma 5.16. *Consider the Lipschitz MAB problem on a metric space* $(X, \mathcal{D})$ *and* $\mathtt{MaxMinCOV}(X) > 0$. *Then for any* $d \in (0, \mathtt{MaxMinCOV}(X))$ *there exists a ball-tree of strength* d. *It follows (using Lemma 5.4) that the regret dimension of any algorithm is at least* $\mathtt{MaxMinCOV}(X)$.

Recall from Section 3 that an *r-packing* of a metric space $(X, \mathcal{D})$ be a subset $S \subset X$ such that any two points in S are at distance at least r from one another. The proof will use the following simple packing lemma.

[25] We have defined this notion while stating our results in Section 1.4. Here, we restate if for the sake of convenience.

Lemma 5.17 (Folklore). *Suppose $(X, \mathcal{D})$ is a metric space of covering dimension d. Then for any $b < d$, $r_0 > 0$ and $C > 0$ there exists $r \in (0, r_0)$ such that X contains an r-packing of size at least $C\,r^{-b}$.*

Proof. Let $r < r_0$ be a positive number such that every covering of $(X, \mathcal{D})$ with radius-r balls requires more than $C\,r^{-b}$ balls. Such an r exists, because the covering dimension of $(X, \mathcal{D})$ is strictly greater than b.

Now let S be any maximal r-packing in $(X, \mathcal{D})$. For every $x \in X$ there must exist some point $y \in S$ such that $\mathcal{D}(x, y) < r$, as otherwise $S \cup \{x\}$ would be an r-packing, contradicting the maximality of S. Therefore, balls $B(x, r)$, $x \in S$ cover the metric space. It follows that $|S| \geq C\,r^{-b}$ as desired. □

Proof of Lemma 5.16. Pick $c \in (d, \mathtt{MaxMinCOV}(X))$. Choose $Y \subset X$ such that $\mathtt{MinCOV}(Y) \geq c$.

Let us recursively construct a ball-tree of strength d. Each tree node will correspond to an extensive-form ball centered in Y. Define the root to be some radius-1 extensive-form ball with center in Y.[26] Now, suppose we have defined a tree node w, which corresponds to an extensive-form ball with center $y \in Y$ and radius r. Let us consider the set $B = Y \cap B(y, \frac{r}{4})$. Then B is non-empty and open in the metric space $(Y, \mathcal{D})$. By definition of the min-covering dimension, we have $\mathtt{COV}(B) \geq c$. Now Lemma 5.17 guarantees the existence of a $(2r')$-packing $S \subset B$ such that $r' < r/4$ and $|S| \geq (r')^{-d}$. Let the children of w correspond to points in S, so that for each $x \in S$ there is a child with center x and radius r'. □

5.3 Special Case: Metric Space with a "Fat Subset"

To gain intuition on the max-min-covering dimension, let us present a family of metric spaces where for a given covering dimension the max-min-covering dimension can be arbitrarily small.

Let us start with two concrete examples. Both examples involve an infinite rooted tree where the out-degree is low for most nodes and very high for a few. On every level of the tree the high-degree nodes produce exponentially more children than the low-degree nodes. For concreteness, let us say that all low-degree nodes have degree 2, all high-degree nodes on a given level of the tree have the same degree, and this degree is such that the tree contains 4^i nodes on every level i. The two examples are as follows:

- one high-degree node on every level; the high-degree nodes form a path, called the *fat end.*
- 2^i high-degree nodes on every level i; the high-degree nodes form a binary tree, called the *fat subtree.*

We assign a *width* of $2^{-i/d}$, for some constant $d > 0$, to each level-i node; this is the diameter of the set of points contained in the corresponding subtree. The tree induces a metric space $(X, \mathcal{D})$ where X is the set of all ends[27], and for $x, y \in X$, we define $\mathcal{D}(x, y)$ to be the width of the least common ancestor of ends x and y.

In both examples, the covering dimension of the entire metric space is $2d$, whereas there exists a low-dimensional "fat subset"—the fat end or the fat subtree—which is, in some sense, responsible for the high covering dimension of X. Specifically, for any subtree U containing the fat end (which is just a point in the metric space) it holds that $\mathtt{COV}(U) = 2d$ but $\mathtt{COV}(X \setminus U) = d$. Similarly, if S is the fat subtree and U is a union of subtrees that cover S then $\mathtt{COV}(U) = 2d$ but $\mathtt{COV}(S \cup (X \setminus U)) = d$.

It is easy to generalize the notion of "fat subtree" to an arbitrary metric space:

Definition 5.18. Given a metric space $(X, \mathcal{D})$, a closed subset $S \subset X$ is called *d-fat*, $d < \mathtt{COV}(X)$, if $\mathtt{COV}(S) \leq d$ and $\mathtt{COV}(X \setminus U) \leq d$ for any open neighborhood U of S.

[26] Recall from Definition 5.3 that extensive-form ball is a pair (x, r) where $x \in X$ is the "center" and $r \in (0, 1]$ is the "radius."

[27] Recall from Section 5.1 that an end of an infinite rooted tree is an infinite path starting at the root.

In both examples above, the max-min-covering dimension is d. This is because every point outside the fat subset has an open neighborhood whose covering dimension is at most d (and the covering dimension of the fat subset itself is at most d, too). We formalize this argument as follows:

Claim 5.19. *Suppose metric space $(X, \mathcal{D})$ contains a subset $S \subset X$ of covering dimension at most d such that every $x \in X \setminus S$ has an open neighborhood N_x of covering dimension at most d. Then* $\texttt{MaxMinCOV}(X) \leq d$.

Proof. Equivalently, we need to show that $\texttt{MinCOV}(Y) \leq d$ for any subset $Y \subset X$.

Fix $Y \subset X$. For each $\epsilon > 0$, we need to produce a non-empty subset $U \subset Y$ such that U is open in $(Y, \mathcal{D})$ and its covering dimension is at most $d + \epsilon$. If $Y \subset S$, then we can simply take $U = Y$, because $\texttt{COV}(Y) \leq \texttt{COV}(S) \leq d$. Now suppose there exists a point $x \in Y \setminus S$. Then $U = N_x \cap Y$ is non-empty and open in the metric space restricted to Y, and $\texttt{COV}(U) \leq \texttt{COV}(N_x) \leq d$. □

In fact, this property applies to any d-fat subset in a compact metric space.

Lemma 5.20. *Suppose a compact metric space $(X, \mathcal{D})$ contains a d-fat subset $S \subset X$. Then*

(a) every $x \in X \setminus S$ has an open neighborhood N_x of covering dimension at most d.
(b) $\texttt{MaxMinCOV}(X) \leq d$.

Proof. To prove part (a), consider some point $x \in X \setminus S$. Since S is closed, $\mathcal{D}(x, S) > 0$. Denoting $r = \frac{1}{4}\,\mathcal{D}(x, S)$, let U be the union of all radius-r open balls centered in S. Then U is an open set containing S, so $\texttt{COV}(X \setminus U) \leq d$. Since $B(x, r) \subset X \setminus U$, its covering dimension is at most d, too.

Part (b) follows from part(a) by Claim 5.19, using the fact that $\texttt{COV}(S) \leq d$. □

5.4 Warm-up: Taking Advantage of Fat Subsets

As a warm-up for our general algorithmic result, let us consider metric spaces with d^*-fat subsets, and design a modification of the zooming algorithm whose regret dimension can be arbitrarily close to d^* for such metric spaces. In particular, we establish that $\texttt{COV}(X)$ is, in general, not an optimal regret dimension. Further, our algorithm essentially retains the instance-specific guarantee with respect to the zooming dimension. As a by-product, we develop much of the technology needed for the general result in the next subsection.

The zooming algorithm from Section 4 may perform poorly on metric spaces with a fat subset S if the optimal arm x^* is located inside S. This is because as the confidence ball containing x^* shrinks, it may be too burdensome to keep covering[28] the profusion of arms located near x^*, in the sense that it may require activating too many arms. We fix this problem by imposing *quotas* on the number of active arms. Thus, some arms may not be covered. However, we show that (for a sufficiently long phase, with very high probability) there exists an optimal arm that is covered, which suffices for the technique in Section 4.1 to produce the desired regret bound.

We define the quotas as follows. For a phase of duration T and a fixed $d > d^*$, the quotas are

$$\forall\, Y \in \{X \setminus S,\ S\}: \quad |\{\textit{active arms}\ x \in Y : r_t(x) \geq \rho\}| \leq \rho^{-d}, \quad \rho = T^{-1/(d+2)}.$$

We use a generic modification of the zooming algorithm where the activation rule only considers arms that, if activated, do not violate any of the given quotas.

To pave the way for a generalization, consider a sequence of sets $(S_0, S_1, S_2) = (X, S, \emptyset)$, and let $k = 1$ be the number of non-trivial sets in this sequence. Our algorithm and analysis easily generalize to a sequence of closed subsets $S_0, \ldots, S_{k+1} \subset X$, $k \geq 1$, which satisfies the following properties:

[28] Recall that an arm x is called *covered* at time t if for some active arm y we have $\mathcal{D}(x, y) \leq r_t(y)$.

ALGORITHM 2: (zooming algorithm with quotas)

```
for phase $i = 1, 2, 3, \ldots$ do
    Initially, no arms are active.
    for round $t = 1, 2, 3, \ldots, 2^i$ do
        EligibleArms = {arms $x \in X$: activating $x$ does not violate any quotas}.
        Activation rule: if some arm $x \in$ EligibleArms is not covered,
            pick any such arm and activate it.
        Selection rule: play any active arm with the maximal index (9).
```

- $X = S_0 \supset S_1 \supset \cdots \supset S_k \supset S_{k+1} = \emptyset$; the sequence is strictly decreasing.
- for every $i \in \{0, \ldots, k\}$ and any open subset $U \subset X$ that contains S_{i+1}, it holds that $\mathtt{COV}(S_i \setminus U) \leq d$.[29]

We call such sequence a *d-fatness decomposition* of length k.

We generalize the quotas in an obvious way: for a phase of duration T and a fixed $d > d^*$,

$$\forall i \in \{0, \ldots, k\} :$$
$$|\{\textit{active arms } x \in S_i \setminus S_{i+1} : r_t(x) \geq \rho\}| \leq \rho^{-d}, \quad \rho = T^{-1/(d+2)}. \tag{31}$$

This completes the specification of the algorithm. The invariant Equation (31) holds for each round; this is because during a given phase the confidence radius of every active arm does not increase over time.

Remark 5.21. Essentially, the algorithm "knows" the decomposition $S_0, \ldots, S_{k+1}$ and the parameter d. To implement the algorithm, it suffices to use the decomposition via a covering oracle for each subset $S_{i+1} \setminus S_i$, $i \in \{0, \ldots, k\}$. Here a covering oracle for subset $Y \subset X$ takes a finite collection of open balls, where each ball is represented as a (center, radius) pair, and either declares that these balls cover Y or outputs an uncovered point.

Theorem 5.22. *Consider the Lipschitz MAB problem on a compact metric space $(X, \mathcal{D})$, which contains a d^*-fatness decomposition $(S_0, \ldots, S_{k+1})$ of finite length $k \geq 1$. Let $\mathcal{A}$ be the zooming algorithm (Algorithm 2) with quotas Equation (31), for some known parameter $d > d^*$. Then the regret dimension of $\mathcal{A}$ is at most d. Moreover, the instance-specific regret dimension of $\mathcal{A}$ is bounded from above by the zooming dimension.*

The remainder of this section presents the full proof of Theorem 5.22. The following rough outline of the proof may serve as a useful guide.

Proof Outline. For simplicity, assume there is a unique optimal arm, and call it x^*. The desired regret bounds follow from the analysis in Section 4.1 as long as x^* is covered w.h.p. throughout any sufficiently long phase. All arms in $S = S_k$ are covered eventually, because $\mathtt{COV}(S) < d$, so if $x^* \in S$, then we are done. If $x^* \notin S$, then pick the largest ℓ such that $x^* \in S_\ell \setminus S_{\ell+1}$. Then there is some $\epsilon > 0$ such that all arms in $S_{\ell+1}$ are suboptimal by at least ϵ. Letting U be an $\frac{\epsilon}{2}$-neighborhood of $S_{\ell+1}$, note that each arm in U is suboptimal by at least $\frac{\epsilon}{2}$. It follows that (w.h.p.) the algorithm cannot activate too many arms in U. However, $S_\ell \setminus U$ has a low covering dimension, so (w.h.p.) the algorithm cannot activate too many arms in $S_\ell \setminus U$, either. It follows that (for a sufficiently long phase, w.h.p.) the algorithm stays within the quota, in which case x^* is covered. □

To prove Theorem 5.22, we incorporate the analysis from Section 4.1 as follows. We state a general lemma that applies to the zooming algorithm with a modified activation rule (and no

[29] For $i = k$, this condition is equivalent to $\mathtt{COV}(S_k) \leq d$.

other changes). We assume that the new activation rule is at least as selective as the original one: an arm is activated only if it is not covered, and at most one arm is activated in a given round. We call such algorithms *zooming-compatible.*

A phase of a zooming-compatible algorithm is called *clean* if the property in Claim 4.5 holds for each round in this phase. Claim 4.5 carries over: each phase i is clean with probability at least $1 - 4^{-i}$. Let us say that a given round in the execution of the algorithm is *well-covered* if after the activation step in this round some optimal arm is covered. (We focus on compact metric spaces, so the supremum $\mu^* = \sup(\mu, X)$ is achieved by some arm; we will call such arm *optimal.*) A phase of the algorithm is called well-covered if all rounds in this phase are well-covered.

An algorithm is called (k, d)*-constrained* if in every round t it holds that

$$|\{\textit{active arms}\ x \in X : r_t(x) \geq \rho\}| \leq (k+1)\,\rho^{-d}, \quad \rho = T^{-1/(d+2)},$$

where T is the duration of the current phase. Note that the zooming algorithm with quotas Equation (31) is (k, d)-constrained by design.

Lemma 5.23 (Immediate from Section 4.1). *Consider the Lipschitz MAB problem on a compact metric space. Let $\mathcal{A}_T$ be one phase of a zooming-compatible algorithm, where T is the duration of the phase. Consider a clean run of $\mathcal{A}_T$.*

(a) Consider some round $t \leq T$ such that all previous rounds are well-covered. Then,

$$\mathcal{D}(x,y) > \min(r_t(x), r_t(y)) \geq \frac{1}{3}\min(\Delta(x), \Delta(y)), \tag{32}$$

for any two distinct active arms $x, y \in X$.

(b) Suppose the phase is well-covered. If $\mathcal{A}_T$ is (c, d)-constrained, $d \geq 0$, then it has regret

$$R(T) \leq O(c \log T)^{\frac{1}{d+2}} \times T^{\frac{d+1}{d+2}}.$$

This regret bound also holds if d is the zooming dimension with multiplier $c > 0$.

An algorithm is called *eventually well-covered* if for every problem instance $(X, \mathcal{D}, \mu)$ there is a constant i_0 such that every clean phase $i > i_0$ is guaranteed to be well-covered, as long as the preceding phase $i - 1$ is also clean.[30]

Corollary 5.24 (Immediate from Section 4.1). *Consider the Lipschitz MAB problem on a compact metric space. Let $\mathcal{A}$ be a zooming-compatible algorithm. Assume $\mathcal{A}$ is eventually well-covered. Then its (instance-specific) regret dimension is at most the zooming dimension. Further, if $\mathcal{A}$ is (k, d)-constrained, for some constant $k > 0$, then the regret dimension of $\mathcal{A}$ is at most d.*

Since our algorithm is (k, d)-constrained by design, to complete the proof of Theorem 5.22 it suffices to prove that the algorithm is eventually well-covered. This is the part of the analysis that is new, compared to Section 4.1. The crux of the argument is encapsulated in the following claim.

Claim 5.25. *Consider the Lipschitz MAB problem on a compact metric space. Fix $d > 0$. Let $S' \subset S \subseteq X$ (where S' can be empty) be closed subsets such that $\texttt{COV}(S \setminus U) < d$ for any open neighborhood U of S'. Further, suppose S contains some optimal arm and S' does not. Let $\mathcal{A}_T$ be a clean phase of a zooming-compatible algorithm, where T is the duration of the phase. Suppose $\mathcal{A}_T$ activates an arm whenever some arm in $S \setminus S'$ is not covered and, for some $\rho_T > 0$,*

$$|\{\text{active arms}\ x \in S \setminus S' : r_t(x) \geq \rho_T\}| < \rho_T^{-d}.$$

[30] The last clause ("as long as the preceding phase is also clean") is not needed for this subsection; it is added for compatibility with the analysis of the per-metric optimal algorithm in Section 5.5.

Here ρ_T depends on T so that $\rho_T \to 0$ as $T \to \infty$. Then the phase is well-covered whenever $T \geq T_0$, for some finite T_0, which may depend on the problem instance.

Proof. Recall that a ρ-packing is a set $P \subset X$ such that any two points in this set are at distance at least ρ. For any $C > 0$ and any subset $Y \subset X$ with $\texttt{COV}(Y) < d$, there exists ρ_0 such that for any $\rho \leq \rho_0$ any ρ-packing of Y consists of at most $C\rho^{-d}$ points.

We pick $\rho_0 > 0$ as follows.

- If S' is empty, then we pick $\rho_0 > 0$ such that any ρ-packing of S, $\rho \leq \rho_0$, consists of at most ρ^{-d} points. Such ρ_0 exists because $\texttt{COV}(S) < d$.
- Now suppose S' is not empty. Since the metric space is compact and S' is closed, μ attains its supremum on S'. Since S' does not contain any optimal arm, it follows that $\sup(\mu, S') < \mu^* - \epsilon$ for some $\epsilon > 0$. Let $U = \cup_{x \in S'} B(x, \frac{\epsilon}{2})$ be the $\frac{\epsilon}{2}$-neighborhood of S'. Then, for each arm $x \in U$, we have $\Delta(x) > \frac{\epsilon}{2}$. Since the metric space is compact, there is $c_0 < \infty$ such that any $\frac{\epsilon}{6}$-packing of U consists of at most c_0 points. Moreover, we are given that $\texttt{COV}(S \setminus U) < d$. Pick $\rho_0 > 0$ such that any $\frac{\rho}{4}$-packing of $S \setminus U$ consists of at most $\rho^{-d} - c_0$ points, for any $\rho \leq \rho_0$.

Supose T is such that $\rho_T \leq \rho_0$; denote $\rho = \rho_T$. Let us prove that all rounds in this phase are well-covered. Let us use induction on round t. The first round of the phase is well-covered by design, because in this round some arm is activated, and the corresponding confidence ball covers the entire metric space. Now assume that for some round t, all rounds before t are well-covered. Let P be the set of all arms $x \in S$ that are active at time t with $r_t(x) \geq \rho$. We claim that $|P| < \rho^{-d}$. Again, we consider two cases depending on whether S' is empty.

- Suppose S' is empty. By Lemma 5.23(a), P is an ρ-packing, so $|P| < \rho^{-d}$ by our choice of ρ_0.
- Suppose S' is not empty. For any active arm $x \in U$ it holds that $\Delta(x) \geq \frac{\epsilon}{2}$. Then by Lemma 5.23(a) the active arms in U form an $\frac{\epsilon}{6}$-packing of U. So U contains at most $c_0 < \infty$ active arms.

  Further, let P' be the set of all arms in $S \setminus U$ that are active at round t with $r_t(x) \geq \rho$. By Lemma 5.23(a), P is a ρ-packing, so $|P'| < \rho^{-d} - c_0$ by our choice of ρ_0. Again, it follows that $|P| < \rho^{-d}$.

Therefore, by our assumption, the algorithm activates an arm whenever some arm in $S \setminus S'$ is not covered. It follows that $S \setminus S'$ is covered after the activation step, so in particular some optimal arm is covered. □

Corollary 5.26. *In the setting of Theorem 5.22, any clean phase of algorithm $\mathcal{A}$ of duration $T \geq T_0$ is well-covered, for some finite T_0, which can depend on the problem instance.*

Proof. Pick the largest $\ell \in \{0, \ldots, k\}$ such that $S_\ell \setminus S_{\ell+1}$ contains some optimal arm. Then Claim 5.25 applies with $S = S_\ell$ and $S' = S_{\ell+1}$. □

In passing, let us give an example of a fatness decomposition of length > 1. Start with a metric space $(X, \mathcal{D})$ with a d-fat subset S. Consider the product metric space $(X \times X, \mathcal{D}^*)$ defined by

$$\mathcal{D}^*((x_1, x_2), (y_1, y_2)) = \mathcal{D}(x_1, y_1) + \mathcal{D}(x_2, y_2).$$

This metric space admits a $2d$-fatness decomposition

$$(S_0, S_1, S_2, S_3) = (X \times X,\ (S \times X) \cup (X \times S),\ S \times S,\ \emptyset).$$

5.5 Transfinite Fatness Decomposition

The fact that $d = \texttt{MaxMinCOV}(X) < \texttt{COV}(X)$ does not appear to imply the existence of a d-fatness decomposition of any finite length. Instead, we prove the existence of a much more general structure, which we then use to design the per-metric optimal algorithm. This structure is a *transfinite* sequence of subsets of X, i.e., a sequence indexed by ordinal numbers rather than integers.[31]

Definition 5.27. Fix a metric space $(X, \mathcal{D})$. A *transfinite d-fatness decomposition* of length β, where β is an ordinal, is a transfinite sequence $\{S_\lambda\}_{0 \le \lambda \le \beta+1}$ of closed subsets of X such that:

(a) $S_0 = X$, $S_{\beta+1} = \emptyset$, and $S_\nu \supseteq S_\lambda$ whenever $\nu < \lambda$.
(b) for any ordinal $\lambda \le \beta$ and any open set $U \subset X$ containing $S_{\lambda+1}$ it holds that $\texttt{COV}(S_\lambda \setminus U) \le d$.[32]
(c) If λ is a limit ordinal, then $S_\lambda = \bigcap_{\nu<\lambda} S_\nu$.

For finite length β this is the same as (non-transfinite) d-fatness decomposition. The smallest infinite length β is a countable infinity $\beta = \omega$. Then the transfinite sequence $\{S_\lambda\}_{0 \le \lambda \le \beta+1}$ consists of subsets $\{S_i\}_{i \in \mathbb{N}}$ followed by $S_\omega = \cap_{i \in \mathbb{N}} S_i$ and $S_{\omega+1} = \emptyset$.

Proposition 5.28. *For every compact metric space $(X, \mathcal{D})$, the max-min-covering dimension is equal to the infimum of all d such that $(X, \mathcal{D})$ has a transfinite d-fatness decomposition.*

Proof. Assume there exists a transfinite d-fatness decomposition $\{S_\lambda\}_{0 \le \lambda \le \beta+1}$, for some ordinal β. Let us show that $\texttt{MaxMinCOV}(X) \le d$. Suppose not, then there exists a non-empty subset $Y \subseteq X$ with $\texttt{MinCOV}(Y) > d$. Let us use transfinite induction on λ to prove that $Y \subseteq S_\lambda$ for all $\lambda \le \beta$. This would imply $Y \subseteq S_\beta$ and consequently $\texttt{COV}(S_\beta) > d$, contradiction.

The transfinite induction consists of three cases: "zero case," "limit case," and "successor case." The zero case is $Y \subseteq S_0 = X$. The limit case is easy: If $\lambda \le \beta$ is a limit ordinal and $Y \subseteq S_\nu$ for every $\nu < \lambda$, then $Y \subseteq S_\lambda = \cap_{\nu<\lambda} S_\nu$. For the successor case, we assume $Y \subseteq S_\lambda$, $\lambda + 1 \le \beta$, and we need to show that $Y \subseteq S_{\lambda+1}$. Suppose not, and pick some $x \in Y \cap (S_\lambda \setminus S_{\lambda+1})$. Since $S_{\lambda+1}$ is closed, x is at some positive distance 2ϵ from $S_{\lambda+1}$. Then an ϵ-neighborhood U of $S_{\lambda+1}$ is disjoint with a ball $B = B(x, \epsilon)$. So $B \subseteq S_\lambda \setminus U$, which implies $\texttt{COV}(B) \le d$ by definition of transfinite d-fatness decomposition. However, since $B \cap Y$ is open in the metric topology induced by Y, by definition of the min-covering dimension, we have $\texttt{COV}(B) > d$. We obtain a contradiction, which completes the successor case.

Now given any $d > \texttt{MaxMinCOV}(X)$, let us construct a transfinite d-fatness decomposition of length β, where β is any ordinal whose cardinality exceeds that of X. For a metric space $(Y, \mathcal{D})$, a point is called d-thin if it is contained in some open $U \subset Y$ such that $\texttt{COV}(Y) < d$, and d-thick otherwise. Let $\texttt{Fat}(Y, d)$ be the set of all d-thick points; note that $\texttt{Fat}(Y, d)$ is a closed subset of Y. For every ordinal $\lambda \le \beta + 1$, we define a set $S_\lambda \subset X$ using transfinite induction as follows:

1. $S_0 = X$ and $S_{\lambda+1} = \texttt{Fat}(S_\lambda, d)$ for each ordinal λ.
2. If λ is a limit ordinal, then $S_\lambda = \bigcap_{\nu<\lambda} S_\nu$.

This completes the construction of a sequence $\{S_\lambda\}_{\lambda \le \beta+1}$.

Note that each S_λ is closed, by transfinite induction. It remains to show that the sequence satisfies the properties (a)–(c) in Definition 5.27. It follows immediately from the construction that $S_0 = X$ and $S_\nu \supseteq S_\lambda$ when $\nu < \lambda$. To prove that $S_\beta = \emptyset$, observe first that the sets $S_\lambda \setminus S_{\lambda+1}$ (for

[31] Formally, a transfinite sequence of length β (where β is an ordinal) is a mapping from {ordinals $\lambda : 0 \le \lambda \le \beta$} to the corresponding domain, in this case the power set of X.
[32] For $\lambda = \beta$, this is equivalent to $\texttt{COV}(S_\beta) \le d$.

$0 \le \lambda < \beta$) are disjoint subsets of X, and the number of such sets is greater than the cardinality of X, so at least one of them is empty. This means that $S_\lambda = S_{\lambda+1}$ for some $\lambda < \beta$. If $S_\lambda = \emptyset$, then $S_\beta = \emptyset$ as desired. Otherwise, the relation $\mathtt{Fat}(S_\lambda, d) = S_\lambda$ implies that the metric space $(S_\lambda, \mathcal{D})$ contains no open set $U \subset S_\lambda$ with $\mathtt{COV}(U) < d$. It follows that $\mathtt{MinCOV}(S_\lambda) \ge d$, contradicting the assumption that $\mathtt{MaxMinCOV}(X) < d$. This completes the proof of property (a). To prove property (b), note that if U is an open neighborhood of $S_{\lambda+1}$ then the set $T = S_\lambda \setminus U$ is closed (hence compact) and is contained in $\mathtt{Thin}(S_\lambda, d)$. Consequently, T can be covered by open sets V satisfying $\mathtt{COV}(V) < d$. By compactness of T, this covering has a finite subcover $V_1, \dots, V_m$, and consequently $\mathtt{COV}(T) = \max_{1 \le i \le m} \mathtt{COV}(V_i) < d$. Finally, property (c) holds by design. □

Theorem 5.29. *Consider the Lipschitz MAB problem on a compact metric space $(X, \mathcal{D})$ with a transfinite d^*-fatness decomposition, $d^* \ge 0$. Then for each $d > d^*$ there exists an algorithm $\mathcal{A}$ (parameterized by d) such that $\mathtt{DIM}(\mathcal{A}) \le d$. Moreover, the instance-specific regret dimension of $\mathcal{A}$ is bounded from above by the zooming dimension.*

In the rest of this section, we design and analyze an algorithm for Theorem 5.29. The algorithm from the previous subsection has regret proportional to the length of the fatness decomposition, so it does not suffice even if the fatness decomposition has countably infinite length. As it turns out, the main algorithmic challenge in dealing with fatness decompositions of transfinite length is to handle the special case of *finite* length k so that the regret bound does not depend on k.

In what follows, let $\{S_\lambda\}_{0 \le \lambda \le \beta+1}$, be a transfinite d^*-fatness decomposition of length β, for some ordinal β and $d^* \ge 0$. Fix some $d > 0$.

Proposition 5.30. *For any closed $V \subset X$, there is a maximal ordinal λ such that V intersects S_λ.*

Proof. Let $\Omega = \{\text{ordinals } \nu \le \beta\colon\ V \text{ intersects } S_\nu\}$, and let $\nu = \sup(\Omega)$. Then,

$$S_\nu \cap V = \bigcap_{\lambda \in \Omega} (S_\lambda \cap V),$$

and this set is nonempty, because X is compact and the closed sets $\{S_\lambda \cap V \,:\, \lambda \in \Omega\}$ have the finite intersection property. (To derive the latter, consider a finite subset $\Omega' \subset \Omega$ and let $\nu' = \max(\Omega') \in \Omega$. Then $\bigcap_{\lambda \in \Omega'} (S_\lambda \cap V) = S_{\nu'} \cap V$, which is not empty by definition of Ω.) □

Recall that the supremum $\mu^* = \sup(\mu, X)$ is attained, because the metric space is compact. Further, recall that the arms x such that $\mu(x) = \mu$ are called optimal. Let $\lambda_{\max}$ be the maximal λ such that S_λ contains an optimal arm. Such $\lambda_{\max}$ exists by Proposition 5.30, because the set $V = \mu^{-1}(\mu^*)$ is non-empty and closed. Note that $S_{\lambda_{\max}}$ contains an optimal arm, whereas $S_{\lambda_{\max}+1}$ does not.

Our algorithm is a version of Algorithm 2 from the previous subsection, with a different "eligibility rule"—the definition of $\mathtt{EligibleArms}$. For phase duration T and an ordinal $\lambda \le \beta$, define the quota as the following condition:

$$\mathsf{Q}_\lambda \triangleq [|\{\textit{active arms } x \in S_\lambda \,:\, r_t(x) \ge \rho\}| < \rho^{-d}], \quad \rho = T^{-1/(d+2)}.$$

The algorithm maintains the *target ordinal* λ^*, recomputed after each phase, so that some arm in S_{λ^*} is activated as long as the quota Q_{λ^*} is satisfied. Further, there is a subset $\mathcal{N}$ of cardinality at most $T^{d/(d+2)}$, chosen in the beginning of each phase, such that all arms in $\mathcal{N}$ are always eligible and all arms not in $S_{\lambda^*} \cup \mathcal{N}$ are never eligible.

Note that such algorithm is $(1, d)$-constrained by design, because in any round t there can be at most ρ^{-d} active arms in $S_{\lambda^*} \setminus \mathcal{N}$ with confidence radius less than $\rho = T^{-1/(d+2)}$.

The analysis hinges on proving that after any sufficiently long clean phase the target ordinal is $\lambda_{\max}$, and then the subsequent phase (assuming it is also clean) is well-covered, and then the desired regret bounds follow from Corollary 5.24. Any sufficiently long clean phase with target

ALGORITHM 3: (the per-metric optimal algorithm)

```
Target ordinal $\lambda^* \leftarrow 0$.
for phase $i = 1, 2, 3, \ldots$ do
  {Phase duration is $T = 2^i$}
  Compute an $\epsilon_0 > 0$ and an $\epsilon_0$-net $\mathcal{N}$ of $X$ such that $|\mathcal{N}| < T^{d/(d+2)}$.
     {use greedy heuristic}
  Initially, no arms are active.
  for round $t = 1, 2, 3, \ldots, T$ do
    $\texttt{EligibleArms} = \begin{cases} \mathcal{N} \cup S_{\lambda^*} & \text{if constraint } \mathtt{Q}_{\lambda^*} \text{ is satisfied,} \\ \mathcal{N} & \text{otherwise.} \end{cases}$
    Activation rule: if some arm $x \in \texttt{EligibleArms}$ is not covered,
       pick any such arm and activate it.
    Selection rule: play any active arm with the maximal index (33).
  {Recompute the target ordinal $\lambda^*$}
  $\epsilon^* = 6\ \max(\epsilon_0, 4T^{-1/(d+2)}\sqrt{\log T})$.
  $\lambda^* = \max\{\lambda :\ S_\lambda \text{ intersects } \bar{B}(A, \epsilon^*)\ \}$, where $A = \{\text{active arms } x\text{: } r_T(x) < \epsilon^*\ \}$.
```

ordinal $\lambda_{\mathtt{max}}$ is well-covered by Claim 5.25. So the only new thing to prove is that that after any sufficiently long clean phase the target ordinal is $\lambda_{\mathtt{max}}$.

We also change the definition of index to

$$I_t(x) = \mu_t(x) + 3\, r_t(x), \tag{33}$$

where, as before, $\mu_t(x)$ denotes the average payoff from arm x in rounds 1 to $t-1$ of the current phase, and $r_t(x)$ is the current confidence radius of this arm. It is easy to check that the analysis in Section 4.1, and therefore also Lemma 5.23 and Corollary 5.24, carry over to any index of the form $I_t(x) = \mu_t(x) + c_0\, r_t(x)$ for some absolute constant $c_0 \geq 2$ (the upper bound on regret increases by the factor of c_0).

The pseudocode is summarized as Algorithm 3. In the beginning of each phase, the subset $\mathcal{N} \subset X$ is defined as follows. We choose $\mathcal{N}$ to be an ϵ_0-net[33] of X, which consists of at most $T^{d/(d+2)}$ points, for (essentially) the smallest possible $\epsilon_0 > 0$. More precisely, we compute an $\epsilon_0 > 0$ and an ϵ_0-net $\mathcal{N}$ using a standard *greedy heuristic*. For a given $\epsilon > 0$, we construct an ϵ-net $S \subset X$ as follows: while there exists a point $x \in X$ such that $\mathcal{D}(S, x) \triangleq \inf_{y \in S} \mathcal{D}(x, y) < \epsilon$, add any such point to S, and abort if $|S| > T^{d/(d+2)}$. We consecutively try $\epsilon = 2^{-j}$ for each $j = 1, 2, 3, \ldots$, and pick the smallest ϵ, which results in an ϵ-net of at most $T^{d/(d+2)}$ points.

In the end of each phase, the new target ordinal λ^* is defined as follows. We pick an ϵ^* according to T and ϵ_0, and focus on arms whose confidence radius is less than ϵ^*. Let A be the set of all such arms. We define λ^* as the largest ordinal λ such that S_λ intersects $\bar{B}(A, \epsilon^*) \triangleq \{x \in X :\ \mathcal{D}(A, x) \leq \epsilon^*\}$, the the closed ϵ^*-neighborhood of A. Such ordinal exists by Proposition 5.30.

Implementation details. To implement Algorithm 3, it suffices to use the following oracles:

- For any finite set of open balls $B_1, \ldots, B_n$ (given via the centers and the radii) whose union is denoted by B, the *depth oracle* returns $\sup\{\lambda :\ S_\lambda \text{ intersects the closure of } B\}$.
- Given balls $B_1, \ldots, B_n$ as above, and an ordinal λ, the *enhanced covering oracle* either reports that B covers S_λ, or it returns an arm $x \in S_\lambda \setminus B$.

[33] Recall from Section 3 that an ϵ-net of a metric space $(X, \mathcal{D})$ is a subset $S \subset X$ such that any two points in S are at distance at least ϵ from one another, and any point in X is within distance ϵ from some point in S.

To avoid the question of how arbitrary ordinals are represented on the oracle's output tape, we can instead say that the depth oracle outputs a point $u \in S_\lambda \setminus S_{\lambda+1}$ instead of outputting λ. In this case, the definition of the covering should be modified so that it inputs a point $u \in S_\lambda \setminus S_{\lambda+1}$ rather than the ordinal λ itself.

Analysis. We bring in the machinery developed in the previous subsection. Note that Algorithm 3 is zooming-compatible and $(1, d)$-constrained by design. Therefore, by Corollary 5.24, we only need to prove that it is eventually well-covered. If in a given clean phase the target ordinal is $\lambda_{\mathtt{max}}$, then this phase satisfies the assumptions in Claim 5.25 for $S = S_{\lambda_{\mathtt{max}}}$ and $S' = S_{\lambda_{\mathtt{max}}+1}$. It follows that any sufficiently long clean phase with target ordinal $\lambda_{\mathtt{max}}$ is well-covered. Thus, it remains to show that after any sufficiently long clean phase of Algorithm 3 the target ordinal is $\lambda_{\mathtt{max}}$. (This is where we use the new definition of index.)

Claim 5.31. *After any sufficiently long clean phase of Algorithm 3 the target ordinal is* $\lambda_{\mathtt{max}}$.

To prove Claim 5.31, we need to "open up the hood" and analyze the internal workings of the algorithm. (We have been avoiding this so far by using Corollary 5.24.) Such analysis is encapsulated in the following claim. Note that we cannot assume that the phase is well-covered.

Claim 5.32. *Consider a clean phase of Algorithm 3 of duration T, with ϵ_0-net $\mathcal{N}$. Let y be an arm that has been played at least once in this phase. Then:*

(a) $\Delta(y) \le 4\, r_T(y) + \epsilon_0$.
(b) For any optimal arm x^*, there exists an active arm x such that

$$\min(\mathcal{D}(x, x^*),\ r_T(x)) \le 4\, r_T(y) + 2\,\epsilon_0.$$

Proof. Let $x_{\mathtt{net}} \in \mathcal{N}$ be such that $\mathcal{D}(x^*, x_{\mathtt{net}}) \le \epsilon_0$. Let t be the last time arm y is played in this phase. Let x be an arm that covers $x_{\mathtt{net}}$ at time t. (Since $\mathcal{N} \subset$ `EligibleArms`, all points in $\mathcal{N}$ are covered at all times.) Then:

$$\begin{aligned}
I_t(x) &\ge \mu(x) + 2r_t(x) && \text{by definition of index and confidence radius}\\
&\ge \mu(x_{\mathtt{net}}) + r_t(x) && \text{because } x \text{ covers } x_{\mathtt{net}} \text{ at time } t\\
&\ge \mu^* - \epsilon_0 + r_t(x) && \text{because } \mathcal{D}(x^*, x_{\mathtt{net}}) \le \epsilon_0\\
I_t(x) &\le I_t(y) && \text{because arm } y \text{ is played at time } t\\
&\le \mu(y) + 4\, r_t(y) && \text{by definition of index and confidence radius}\\
&= \mu^* - \Delta(y) + 4\, r_t(y).
\end{aligned}$$

Combining the two inequalities, we obtain

$$\mu^* - \Delta(y) + 4\, r_t(y) \ge I_t(x) \ge \mu^* - \epsilon_0 + r_t(x).$$

Noting that $r_T(y) = r_t(y)$, we obtain

$$\Delta(y) + r_t(x) \le 4\, r_T(y) + \epsilon_0.$$

This immediately implies part (a) of the claim. Part (b) follows by triangle inequality, because

$$\mathcal{D}(x, x^*) \le \mathcal{D}(x, x_{\mathtt{net}}) + \mathcal{D}(x_{\mathtt{net}}, x^*) \le r_t(x) + \epsilon_0. \qquad \square$$

We also need a simple and well-known fact about compact metric spaces.

Claim 5.33 (Folklore). *For any given $\delta > 0$ there exists $T_0 < \infty$ such that in any phase of Algorithm 3 of duration $T > T_0$, the algorithm computes an ϵ_0-net $\mathcal{N}$ such that $\epsilon_0 < \delta$.*

Proof. Fix $\delta > 0$. Since the metric space is compact, there exists a covering of X with finitely many subsets $S_1, \ldots, S_n \subset X$ of diameter less than $\frac{\delta}{2}$. Suppose T is large enough so that $n < T^{d/(d+2)}$. Suppose the algorithm computes an ϵ_0-net $\mathcal{N}$ such that $\epsilon_0 \geq \delta$. Then the following iteration of the greedy heuristic (if not aborted) would construct an $\epsilon_0/2$-net $\mathcal{N}'$ for X with more than $T^{d/(d+2)}$ points. However, any two points in $\mathcal{N}'$ lie at distance $\geq \delta/2$ from one another, so they cannot lie in the same set S_i. It follows that $|\mathcal{N}'| \leq n$, contradiction. □

Proof of Claim 5.31. Consider a clean phase of Algorithm 3 of duration T, with an ϵ_0-net $\mathcal{N}$. Let ϵ^* and A be defined as in Algorithm 3, so that $A = \{\text{active arms } x\colon r_t(x) < \epsilon^* \}$. We need to show that for any sufficiently large T two things happen: $\bar{B}(A, \epsilon^*)$ intersects $S_{\lambda_{\max}}$ and it does not intersect $S_{\lambda_{\max}+1}$.

Let $x_{\mathtt{freq}}$ be the most frequently played arm by the end of the phase. We claim that

$$r_T(x_{\mathtt{freq}}) < 4T^{-1/(d+2)} \sqrt{\log T}.$$

Suppose not. By our choice of $x_{\mathtt{freq}}$, at time T all arms have confidence radius at least $r_T(x_{\mathtt{freq}})$. Since the algorithm is $(1, d)$-constrained, it follows that at most $n = 2T^{d/(d+2)}$ arms are activated throughout the phase. So by the pigeonhole principle $n_T(x_{\mathtt{freq}}) \geq T/n = \frac{1}{2}T^{2/(d+2)}$, which implies the desired inequality. Claim proved.

Let $x^* \in S_{\lambda_{\max}}$ be some optimal arm. Taking $y = x_{\mathtt{freq}}$ in Claim 5.32(b) and noting that $4\, r_T(x_{\mathtt{freq}}) + 2\, \epsilon_0 \leq \epsilon^*$, we derive that there exists an active arm x such that $\mathcal{D}(x, x^*) \leq \epsilon^*$ and $r_T(x) \leq \epsilon^*$. It follows that $x \in A$ and $x^* \in \bar{B}(A, \epsilon^*)$. Therefore $\bar{B}(A, \epsilon^*)$ intersects $S_{\lambda_{\max}}$.

Since the metric space is compact and $S_{\lambda_{\max}+1}$ is a closed subset that does not contain an optimal arm, it follows that any arm in this subset has expected payoff at most $\mu^* - \epsilon$, for some $\epsilon > 0$. Assume T is sufficiently large so that $\epsilon^* < \epsilon/6$. (We can make sure that $\epsilon_0 < \epsilon/6$ by Claim 5.33).

To complete the proof, we need to show that $\bar{B}(A, \epsilon^*)$ does not intersect $S_{\lambda_{\max}+1}$. Suppose this is not the case. Then there exists $x \in S_{\lambda_{\max}+1}$ and active $y \in X$ such that $\mathcal{D}(x, y) \leq \epsilon^*$ and $r_T(y) \leq \epsilon^*$. Then by Claim 5.32(a) we have that $\Delta(y) \leq 4\,\epsilon^* + \epsilon_0 \leq 5\,\epsilon^*$, which implies that $\Delta(x) \leq \Delta(y) + \mathcal{D}(x, y) \leq 6\,\epsilon^* < \epsilon$, contradicting our assumption that every arm in $S_{\lambda_{\max}+1}$ has expected payoff at most $\mu^* - \epsilon$. Claim proved. □

6 THE (SUB)LOGARITHMIC VS. $\sqrt{T}$ REGRET DICHOTOMY

This section concerns the dichotomy between (sub)logarithmic and $\sqrt{t}$ regret for Lipschitz bandits and Lipschitz experts (Theorems 1.6 and 1.9, respectively). We focus on the restriction of these results to compact metric spaces:

Theorem 6.1. *Fix a compact metric space $(X, \mathcal{D})$. The following dichotomies hold:*

(a) The Lipschitz MAB problem on $(X, \mathcal{D})$ is either $f(t)$-tractable for every $f \in \omega(\log t)$, or it is not $g(t)$-tractable for any $g \in o(\sqrt{t})$.

(b) The Lipschitz experts problem on $(X, \mathcal{D})$ 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 and uniformly Lipschitz payoffs.

In both cases, (sub)logarithmic tractability occurs if and only if X is countable.

We also prove two auxiliary results: the $(\log t)$-intractability for Lipschitz bandits on infinite metric spaces (Theorem 1.7), and an algorithmic result via a more intuitive oracle access to the metric space (for metric spaces of finite *Cantor-Bendixson rank*, a classic notion from point-set topology).

The section is organized as follows. We provide a joint analysis for Lipschitz bandits and Lipschitz experts: an overview in Section 6.1, the lower bound is in Section 6.2, and the algorithmic result is in Section 6.3. The two auxiliary results are, respectively, in Sections 6.4 and 6.5.

6.1 Regret Dichotomies: An Overview of the Proof

We identify a simple topological property (existence of a topological well-ordering) that entails the algorithmic result, and another topological property (existence of a perfect subspace) that entails the lower bound.

Definition 6.2. 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, then X is called *well-orderable*. A metric space $(X, \mathcal{D})$ 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 [33], but the particular definition given here is new, to the best of our knowledge.

The proof of Theorem 6.1 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 6.3. *For any compact metric space $(X, \mathcal{D})$, the following are equivalent: (i) X is a countable set, (ii) $(X, \mathcal{D})$ is well-orderable, and (iii) no metric subspace of $(X, \mathcal{D})$ is perfect.*[34]

Lemma 6.3 follows from classical theorems of Cantor [33] and Mazurkiewicz and Sierpinski [74]. We provide a proof in Appendix C for the sake of making our exposition self-contained.

Extension to arbitrary metric spaces. We extend Theorem 6.1 to the corresponding dichotomies for arbitrary metric spaces using the reduction to complete metric spaces in Appendix B, and the $o(t)$-intractability result for non-compact metric spaces in Theorem 1.10 (which is proved independently in Section 7).

For Lipschitz MAB, the argument is very simple. First, we reduce from arbitrary metric spaces to complete metric spaces: We show that the Lipschitz MAB problem is $f(t)$-tractable on a given metric space if and only if it is $f(t)$-tractable on the completion thereof (see Appendix B). Second, we reduce from complete metric spaces to compact metric spaces using Theorem 1.10: By this theorem, the Lipschitz MAB problem is not $o(t)$-tractable if the metric space is complete but not compact. Thus, we obtain the desired dichotomy for Lipschitz MAB on arbitrary metric spaces, as stated in Theorem 1.6.

For Lipschitz experts, the argument is slightly more complicated, because the reduction to complete metric spaces only applies to the lower bound. Let $(X, \mathcal{D})$ be an arbitrary metric space, and let $(X^*, \mathcal{D}^*)$ denote the metric completion thereof. First, if $(X^*, \mathcal{D}^*)$ is not compact, then by Theorem 1.10 the Lipschitz experts problem is not $o(t)$-tractable. Therefore, it remains to consider the case that $(X^*, \mathcal{D}^*)$ is compact. Note that Theorem 6.1 applies to $(X^*, \mathcal{D}^*)$. If X^* is not countable, then by Theorem 6.1 the problem is not $o(\sqrt{t})$-tractable on $(X^*, \mathcal{D}^*)$, and therefore it is not $o(\sqrt{t})$-tractable on $(X, \mathcal{D})$ (see Appendix B). If X^* is countable, then the algorithm and analysis in Section 6.3 apply to X, too, and guarantee $O(1)$-tractability. Thus, we obtain the desired dichotomy for Lipschitz experts on arbitrary metric spaces, as stated in Theorem 1.9.

[34] For arbitrary metric spaces, we have (ii) $\iff$ (iii) and (i)$\Rightarrow$(ii), but not (ii)$\Rightarrow$(i).

6.2 Lower Bounds via a Perfect Subspace

In this section, we prove the following lower bound:

THEOREM 6.4. *Consider the uniformly Lipschitz experts problem on a metric space $(X, \mathcal{D})$, which has a perfect subspace. Then the problem is not g-tractable for any $g \in o(\sqrt{t})$. In particular, for any such g there exists a distribution $\mathcal{P}$ over problem instances μ such that for any experts algorithm $\mathcal{A}$ we have*

$$\Pr_{\mu \in \mathcal{P}} \left[R_{(\mathcal{A},\, \mu)}(t) = O_\mu(g(t)) \right] = 0. \tag{34}$$

Let us construct the desired distribution over problem instances. First, we use the existence of a perfect subspace to construct a ball-tree (cf. Definition 5.3).

LEMMA 6.5. *For any metric space with a perfect subspace there exists a ball-tree in which each node has exactly two children.*

PROOF. Consider a metric space $(X, \mathcal{D})$ with a perfect subspace $(Y, \mathcal{D})$. 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 $B(y, r/3)$ contains another point $y' \in Y$. Let $r' = \mathcal{D}(y, y')/2$ and define the two children of (y, r) as (y, r') and (y', r'). □

Now let us use the ball-tree to construct the distribution on payoff functions. (We will re-use this construction in Section 8.3.1.) In what follows, we consider a metric space $(X, \mathcal{D})$ with a fixed ball-tree T. For each $i \geq 1$, let D_i be the set of all depth-i nodes in the ball-tree. Recall that an *end* in a ball-tree is an infinite path from the root: $\mathbf{w} = (w_0, w_1, w_2, \ldots)$, where $w \in D_i$ for all i. For each tree node $w = (x_0, r_0)$ define the "bump function" $F_w : X \to [0, 1]$ as in Equation (24):

$$F_w(x) = \begin{cases} \min\{r_0 - \mathcal{D}(x, x_0),\ r_0/2\} & \text{if } x \in B(x_0, r_0), \\ 0 & \text{otherwise.} \end{cases} \tag{35}$$

The construction is parameterized by a sequence $\delta_i, \delta_2, \delta_3, \ldots \in (0, 1)$, which we will specify later.

Definition 6.6. 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 λ there is an associated end $\mathbf{w}(\lambda)$ defined by $\mathbf{w} = (w_0, w_1, \ldots)$ where w_0 is the root and for $i > 0$, w_i is the unique child of w_{i-1} that belongs to λ.

CONSTRUCTION 6.7. *For any lineage λ let us define a problem instance $\mathbb{P}_\lambda$ (probability measure on payoff functions) via the following sampling rule. First every tree node w independently samples a random sign $\sigma(w) \in \{+1, -1\}$ so that $\mathbb{E}[\sigma(w)] = \delta_i$ if w is the depth $i \geq 1$ node in $\mathbf{w}(\lambda)$, and choosing the sign uniformly at random otherwise. Define the payoff function π associated with a particular sign pattern $\sigma(\cdot)$ as follows:*

$$\pi = \frac{1}{2} + \frac{1}{3} \sum_{w \in T \setminus D_0} \sigma(w)\, F_w. \tag{36}$$

Let $\mu_\lambda(x) = \mathbb{E}_{\pi \sim \mathbb{P}_\lambda}[\pi(x)]$ denote the expectation of $\pi(x)$ under distribution $\mathbb{P}_\lambda$.

Let $\mathcal{P}_T$ be the distribution over problem instances $\mathbb{P}_\lambda$ in which λ is a complete lineage sampled uniformly at random; that is, each node samples one of its children independently and uniformly at random, and λ is the set of sampled children.

Remark. By Lemma 5.8, the payoff function in Equation (36) is Lipschitz on $(X, \mathcal{D})$ for any sign pattern $\sigma(\cdot)$. Therefore $\mathbb{P}_\lambda$ is an instance of uniformly Lipschitz experts problem, for each lineage λ.

To complete Construction 6.7, it remains to specify the δ_i's. Fix function $g()$ from Theorem 6.4. For each $i \geq 1$, let $r_i^* = \min\{r : (x,r) \in D_i\}$ be the smallest radius among depth-i nodes in the ball-tree. Note that $r_i^* \leq 4^{-i}$. Choose a number n_i large enough that $g(n) < \frac{1}{24\,i}\, r_i^* \sqrt{n}$ for all $n > n_i$; such n_i exists because $g \in o(\sqrt{t})$. Let $\delta_i = n_i^{-1/2}$.

Discussion. For a complete lineage λ, the expected payoffs are given by

$$\mu_\lambda = \frac{1}{2} + \frac{1}{3} \sum_{i=1}^{\infty} \delta_i \, F_{w_i}, \tag{37}$$

where $\mathbf{w}(\lambda) = (w_0, w_1, \ldots)$ is the end associated with λ. For the special case of MAB it would suffice to construct a problem instance with expected payoffs given by Equation (37), without worrying about lineages or random sign patterns. This would be a "weighted" version of the lower-bounding construction from Section 5.1.

However, for the full-feedback problem it is essential that the sum in Equation (36) is over all tree nodes (except the root), rather than the end $\mathbf{w}(\lambda)$. If the sum were over $\mathbf{w}(\lambda)$, then a single sample of the payoff function π would completely inform the learner of the location of $\mathbf{w}(\lambda)$ in the tree. (Just look for the nested rings on which π varies, and they form a target whose bulls-eye is $\mathbf{w}(\lambda)$.) Instead, we fill the whole metric space with "static" in the form of a hierarchically nested set of rings on which π varies, where the only special distinguishing property of the rings that zero in on $\mathbf{w}(\lambda)$ is that there is a slightly higher probability that π increases on those rings. Thus, $\mathbf{w}(\lambda)$ is well-hidden, and in particular is impossible to learn from a single sample of π.

Let us state and prove a salient property of Construction 6.7, which we use to derive the regret lower bound. (This property holds for an arbitrary non-increasing sequence of δ_i's; we will re-use it in Section 8.3.1.)

Lemma 6.8. *Fix a complete lineage λ and tree node $v \in \lambda$. To fix the notation, let us say that v is depth-i node with corresponding ball B of radius r.*

(i) For every event $\mathcal{E}$ in the Borel σ-algebra on $[0,1]^X$,

$$\mathbb{P}_\lambda(\mathcal{E}) / \mathbb{P}_{\lambda \setminus \{v\}}(\mathcal{E}) \in [1 - \delta_i,\ 1 + \delta_i].$$

(ii) If $v \in \mathbf{w}(\lambda)$, then $\sup(\mu_\lambda, B) - \sup(\mu_\lambda, X \setminus B) \geq r\delta_i/6$.

Proof. For part (i), let us treat $\mathcal{E}$ as a set of sign patterns $\sigma : V \to \{\pm 1\}$, where V is the set of all nodes in the ball-tree, and let us treat $\mathbb{P}_\lambda$ as a measure on these sign patterns. For each sign $\beta \in \{\pm 1\}$, let $\mathcal{E}_\beta = \{\sigma \in \mathcal{E} :\ \sigma(v) = \beta\}$ be the set of all sign patterns in $\mathcal{E}$ with a given sign on node u. Note that

$$\mathbb{P}_\lambda(\mathcal{E}) = \sum_{\beta \in \{\pm 1\}} \mathbb{P}_\lambda(\mathcal{E}_\beta) \cdot \mathbb{P}_\lambda\left(\sigma(v) = \beta\right).$$

This equality holds for any lineage, in particular for lineage $\lambda \setminus \{v\}$.

For brevity, denote $\mathbb{P}_0 = \mathbb{P}_{\lambda \setminus \{v\}}$. Observe that $\mathbb{P}_0$ and $\mathbb{P}_\lambda$ differ only in how they set $\sigma(v)$. We can state this property rigorously as follows:

$$\mathbb{P}_\lambda(\mathcal{E}_\beta) = \mathbb{P}_0(\mathcal{E}_\beta) \qquad \text{for each sign } \beta \in \{\pm 1\}.$$

Now, recalling that the event $\{\sigma(v) = 1\}$ is assigned probability $\frac{1}{2}$ under measure $\mathbb{P}_0$, and probability $\frac{1}{2} + \delta_i/2$ under measure $\mathbb{P}_\lambda$, it follows that

$$\mathbb{P}_\lambda(\mathcal{E}) - \mathbb{P}_0(\mathcal{E}) = (\delta_i/2)\ \left(\mathbb{P}_0(\mathcal{E}_+) - \mathbb{P}_0(\mathcal{E}_-)\right),$$
$$|\mathbb{P}_\lambda(\mathcal{E}) - \mathbb{P}_0(\mathcal{E})| \leq (\delta_i/2)\ \left(\mathbb{P}_0(\mathcal{E}_+) + \mathbb{P}_0(\mathcal{E}_-)\right) = \delta_i\, \mathbb{P}_0(\mathcal{E}).$$

For part (ii), write $\mathbf{w}(\lambda) = (w_0, w_1, \dots)$ be the end corresponding to λ. Recall that $v = w_i$. For each w_j, let B_j be the corresponding ball, and let r_j be its radius. Using Equation (37) and the fact that the sequence $(B_j : j \in \mathbb{N})$ is decreasing, it follows that

$$\sup(\mu_\lambda, B) = \frac{1}{2} + \frac{1}{6}\sum_{j=1}^{\infty} \delta_j\, r_j,$$

$$\sup(\mu_\lambda, X \setminus B) = \frac{1}{2} + \frac{1}{6}\sum_{j=1}^{i-1} \delta_j\, r_j,$$

$$\sup(\mu_\lambda, B) - \sup(\mu_\lambda, X \setminus B) = \frac{1}{6}\sum_{j=i}^{\infty} \delta_j\, r_j \geq \delta_i\, r_i/6. \qquad \square$$

Lemma 6.9. *Consider a metric space $(X, \mathcal{D})$ with a ball-tree T. Then (34) holds with $\mathcal{P} = \mathcal{P}_T$.*

To prove this lemma, we define a notion called an (ϵ, δ, k)-ensemble, analogous to the (ϵ, k)-ensembles defined in Section 5.1. As before, it is convenient to articulate this definition in the more general setting of the *feasible experts problem*, in which one is given a set of arms X (not necessarily a metric space) along with a collection $\mathcal{F}$ of Borel probability measures on the set $[0,1]^X$ of functions $\pi : X \to [0,1]$. A problem instance of the feasible experts problem consists of a triple $(X, \mathcal{F}, \mathbb{P})$ where X and $\mathcal{F}$ are known to the algorithm, and $\mathbb{P} \in \mathcal{F}$ is not.

Definition 6.10. 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 π 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 (ϵ, δ, 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 $i > 0$ and every event $\mathcal{E}$ in the Borel σ-algebra of $[0,1]^X$, we have

$$1 - \delta < \mathbb{P}_0(\mathcal{E})/\mathbb{P}_i(\mathcal{E}) < 1 + \delta.$$

(ii) for every $i > 0$, we have $\sup(\mu_i, S_i) - \sup(\mu_i, X \setminus S_i) \geq \epsilon$.

Essentially, the measures $\mathbb{P}_1, \dots, \mathbb{P}_k$ correspond to the children of any given node in the ball-tree. The precise connection to Construction 6.7 is stated below, derived as corollary of Lemma 6.8.

Corollary 6.11. *Fix an arbitrary complete lineage λ in a ball-tree T and a tree node $u \in \mathbf{w}(\lambda)$. Let $u_1, \dots, u_k$ be the children of u. Let u' be the unique child of u contained in λ. Define lineage $\lambda_0 = \lambda \setminus \{u'\}$, and complete lineages $\lambda_i = \lambda_0 \cup \{u_i\}$ for each $i \in [1,k]$. Then the tuple $\vec{\mathbb{P}} = (\mathbb{P}_{\lambda_0}, \mathbb{P}_{\lambda_1}, \dots, \mathbb{P}_{\lambda_k})$ of probability measures from Construction 6.7 constitutes a $(\epsilon, 2\,\delta_j, k)$-ensemble where j is the depth of the tree nodes u_i, r is their radius, and $\epsilon = r\delta_j/6$.*

Proof. Let $S_1, \dots, S_k$ be the balls that correspond to $u_1, \dots, u_k$. Fix u_i, and apply Lemma 6.8 with lineage λ_i and tree node u_i. Then both parts of Definition 6.10 are satisfied for a given i. (For part (i), note that $\lambda_0 = \lambda_i \setminus \{u_i\}$. Observe that Lemma 6.8(i) bounds $\mathbb{P}_\lambda(\mathcal{E})/\mathbb{P}_{\lambda\setminus\{v\}}(\mathcal{E})$, whereas for Definition 6.10, we need to bound the inverse ratio; hence, the bound increases from δ_i to $2 \cdot \delta_i$.) □

Theorem 6.12. *Consider the feasible experts problem on $(X, \mathcal{F})$. Let $\vec{\mathbb{P}}$ be an (ϵ, δ, k)-ensemble with $\{\mathbb{P}_1, \dots, \mathbb{P}_k\} \subseteq \mathcal{F}$ 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. To preserve the flow of the article, the proof of this theorem is deferred until Appendix A, where the relevant KL-divergence techniques are developed. The proof of Theorem 6.4 uses Theorem 6.12 for $k = 2$, and the proof of Theorem 1.13 will use it again for large k.

Proof of Lemma 6.9. Let us fix an experts algorithm $\mathcal{A}$ and a function $g \in o(\sqrt{t})$, and consider the distribution over problem instances in Construction 6.7. For each complete lineage λ and tree node $w \in \mathbf{w}(\lambda)$, let w_1, w_2 denote the children of w in the ball-tree, and let w' denote the unique child that belongs to λ. 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 $\vec{\mathbb{P}} = (\mathbb{P}_{\lambda_0}, \mathbb{P}_{\lambda_1}, \mathbb{P}_{\lambda_2})$. By Corollary 6.11, this triple constitutes an $(\epsilon, 2\,\delta_i, 2)$-ensemble where i is the depth of w_1, w_2 in the ball-tree, r is their radius, and $\epsilon = r\,\delta_i/6$. By Theorem 6.12 there exists a problem instance $\alpha(w) \in \{\mathbb{P}_{\lambda_1}, \mathbb{P}_{\lambda_2}\}$ such that for any $t_i < \frac{1}{8}\ln(34) \cdot \delta_i^{-2}$ one has

$$R_{(\mathcal{A},\, \alpha(w))}(t_i) \geq \epsilon t_i/2.$$

Taking $t_i \in (\frac{1}{4}, \frac{1}{8}\ln(34)) \cdot \delta_i^{-2}$, one has

$$R_{(\mathcal{A},\, \alpha(w))}(t_i) \geq \epsilon t_i/2 = r\delta_i t_i/12 > \frac{1}{24} r_i^* \sqrt{t_i},$$

where r_i^* is the smallest radius among all depth-i nodes in the ball-tree. Recalling that we chose n_i large enough that $g(n_i) < \frac{1}{24\,i}\, r_i^* \sqrt{n}$ for all $n > n_i$, and that $n_i = \delta_i^{-2}$, we see that

$$i \cdot g(t_i) < \frac{1}{24} r_i^* \sqrt{t_i} < R_{(\mathcal{A},\, \alpha(w))}(t_i).$$

For each depth i, let us define $\mathcal{E}_i$ to be the set of input distributions $\mathbb{P}_\lambda$ such that λ is a complete lineage whose associated end $\mathbf{w}(\lambda) = (w_0, w_1, \ldots)$ satisfies $w_i = \alpha(w_{i-1})$. Interpreting these sets as random events under the probability distribution $\mathcal{P}_T$, they are mutually independent events each having probability $\frac{1}{2}$. Furthermore, we have proved that there exists a sequence of times $t_i \to \infty$ such that for each i, we have $R_{(\mathcal{A},\, \mathbb{P})}(t_i) > i \cdot g(t_i)$ for any $\mathbb{P} \in \mathcal{E}_i$.

For each complete lineage λ, define the "smallest possible constant" if we were to characterize the algorithm's regret on problem instance $\mathbb{P}_\lambda$ using function g:

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

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 λ. Indeed, if infinitely many events $\mathcal{E}_i$ happen, then event $\{C_\lambda = \infty\}$ happens as well. But the probability that infinitely many events $\mathcal{E}_i$ happen is 1, because for every positive integer n, $\Pr[\cap_{i=n}^{\infty} \overline{\mathcal{E}_i}] = \prod_{i=n}^{\infty} \Pr[\overline{\mathcal{E}_i} \mid \cap_{j=n}^{i-1} \overline{\mathcal{E}_j}] = 0$. □

6.3 Tractability for Compact Well-orderable Metric Spaces

In this section, we prove the main algorithmic result.

Theorem 6.13. *Consider a compact well-orderable metric space $(X, \mathcal{D})$. Then:*

(a) *the Lipschitz MAB problem on $(X, \mathcal{D})$ is f-tractable for every $f \in \omega(\log t)$;*
(b) *the Lipschitz experts problem on $(X, \mathcal{D})$ 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 $(X, \mathcal{D})$ with a topological well-ordering $\prec$ and a payoff function μ. 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 6.14. *There exists an optimal strategy $x^* \in X$ for which it holds that* $\sup(\mu, X \setminus S(x^*)) < \mu^*$.

Proof. Let X^* be the set of all optimal strategies. Since μ 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 μ 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 6.14. Our algorithm is geared toward 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 6.15. A *δ-covering set* of a metric space $(X, \mathcal{D})$ is a subset $S \subset X$ such that each point in X lies within distance δ from some point in S. An oracle $\mathcal{O} = \mathcal{O}(k)$ is a *covering oracle* for $(X, \mathcal{D})$ if it inputs $k \in \mathbb{N}$ and outputs a pair (δ, S) where $\delta = \delta_{\mathcal{O}}(k)$ is a positive number and S is a δ-covering set of X consisting of at most k points. Here $\delta_{\mathcal{O}}(\cdot)$ is any function such that $\delta_{\mathcal{O}}(k) \to 0$ as $k \to \infty$.

Definition 6.16. Given a metric space $(X, \mathcal{D})$ 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* $\mathsf{EXPL}()$.

ALGORITHM 4: Subroutine $\mathsf{EXPL}(k, n, r)$: 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 δ-covering set 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, $\mathsf{EXPL}(k, n, r)$ 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 6.17. *Fix a problem instance and let x^* be the optimal strategy from Lemma 6.14. Consider increasing functions $k, n, T : \mathbb{N} \to \mathbb{N}$ such that $r(t) := 4\sqrt{(\log T(t))/n(t)} \to 0$. Then for any sufficiently large round t, with probability at least $1 - T^{-2}(t)$, the subroutine EXPL with parameters $k(t)$, $n(t)$, $r(t)$ returns x^*.*

Proof. Let us use the notation from Algorithm 4. Fix t and consider a run of $\mathsf{EXPL}(k(t), n(t), r(t))$. 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 δ-covering set, there exists $y^* \in S$ such that $\mathcal{D}(x^*, y^*) \leq \delta$. Let us fix one such y^*. It suffices to prove that y^* is not a loser. Indeed, if $\mu_{\text{av}}(y) - \mu_{\text{av}}(y^*) > 2\,r(t) + \delta$ 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 $\delta = \delta_O(k(t)) < r_0$, where $\delta_O(\cdot)$ 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 $\mathcal{D}(x, y) \leq \delta$. It suffices to prove that x is a loser. Consider some $y^* \in S$ such that $\mathcal{D}(x^*, y^*) \leq \delta$. 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 6.13. Let us fix a function $f \in \omega(\log t)$. Then $f(t) = \alpha(t) \log(t)$ where $\alpha(t) \to \infty$. 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 $r_t \to 0$. The algorithm proceeds in phases of a doubly exponential length[35]. A given phase $i = 1, 2, 3, \ldots$ lasts for $T = 2^{2^i}$ rounds. In this phase, first, we call the exploration subroutine $\mathsf{EXPL}(k_T,\ n_T,\ r_T)$. Let $x_{\text{or}} \in X$ be the point returned by this subroutine. Then, we play x_{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 6.17 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 $i \in \{i_0, i_0 + 1, \ldots, \lceil \log\log t \rceil\}$ 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 $\mathsf{EXPL}()$ on the *free peeks.* For exploitation, we use the point returned by $\mathsf{EXPL}()$ 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 $i = 1, 2, 3, \ldots$ lasts for $T = 2^i$ rounds. In this phase, we run the exploration subroutine $\mathsf{EXPL}(k_T,\ n_T,\ r_T)$ on the *free peeks.* In each round, we *bet* on the point returned by $\mathsf{EXPL}()$ in the previous phase. This completes the description of the algorithm.

By Lemma 6.17 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 $t_0 + O(1)$. □

6.4 The $(\log t)$-intractability for Infinite Metric Spaces: Proof of Theorem 1.7

Consider an infinite metric space $(X, \mathcal{D})$. In view of Theorem 1.10, we can assume that the completion X^* of X is compact. It follows that there exists $x^* \in X^*$ such that $x_i \to x^*$ for some sequence $x_1, x_2, \ldots \in X$. Let $r_i = \mathcal{D}(x_i, x^*)$. Without loss of generality, assume that $r_{i+1} < \frac{1}{2}\, r_i$ for each i, and that the diameter of X is 1.

Let us define an ensemble of payoff functions $\mu_i : X \to [0, 1]$, $i \in \mathbb{N}$, where μ_0 is the "baseline" function, and for each $i \geq 1$ function μ_i is the "counterexample" in which a neighborhood of x_i

[35] The doubly exponential phase length is necessary to get f-tractability. If we employed the more familiar *doubling trick* of using phase length 2^i (as in References [12, 60, 64], for example), then the algorithm would only be $f(t)\ \log t$-tractable.

has slightly higher payoffs. The "baseline" is defined by $\mu_0(x) = \frac{1}{2} - \frac{\mathcal{D}(x,x^*)}{8}$, and the "counterexamples" are given by

$$\mu_i(x) = \mu_0(x) + \nu_i(x), \quad \text{where} \quad \nu_i(x) = \frac{3}{4} \max\left(0, \frac{r_i}{3} - \mathcal{D}(x,x^*)\right).$$

Note that both μ_0 and ν_i are $\frac{1}{8}$-Lipschitz and $\frac{3}{4}$-Lipschitz w.r.t. $(X,\mathcal{D})$, respectively, so μ_i is $\frac{7}{8}$-Lipschitz w.r.t $(X,\mathcal{D})$. 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 μ_0 and μ_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 μ_0 and under μ_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 σ-algebra. Let $\mathbb{P}_i[\cdot]$ and $\mathbb{E}_i[\cdot]$ be, respectively, the distribution and expectation induced by μ_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 6.18. *For any $i \geq 1$ and any event $S \in \sigma(N_i(t))$ it is the case that*

$$\mathbb{P}_i[S] < \frac{1}{3} \leq \mathbb{P}_0[S] \quad \Rightarrow \quad -\ln(\mathbb{P}_i[S]) - \frac{3}{e} \leq O(r_i^2)\ \mathbb{E}_0[N_i(t)]. \tag{38}$$

Remark. The reason our argument proves the regret lower bound in terms of $\log(t)$, rather than some other function of t, is the $\ln(\cdot)$ term in Equation (38), which in turn comes from the $\exp(\cdot)$ term in Claim A.5 (which captures a crucial property of KL-divergence).

Claim 6.18 is proved using KL-divergence techniques, see Appendix A 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 6.18 either $\mathbb{P}_0(S) < \frac{1}{3}$ or the consequent in (38) 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 establishes Theorem 1.7, since $r_i^{-1} \to \infty$ as $i \to \infty$.

6.5 Tractability via More Intuitive Oracle Access

In Theorem 6.13, 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 that 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 6.19. Fix a metric space $(X,\mathcal{D})$. 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 $\textsc{lim}(S)$ denote the *limit set*: the set of all limit points of S. Let $\textsc{lim}(S,0) = S$, and $\textsc{lim}(S,i) = \textsc{lim}(\textsc{lim}(\cdots \textsc{lim}(S)))$, where $\textsc{lim}(\cdot)$ is applied i times. The *Cantor-Bendixson rank* of $(X,\mathcal{D})$ is defined as $\sup\{n : \textsc{lim}(X,n) \neq \emptyset\}$.

Let us say that a *Cantor-Bendixson metric space* is one with a finite Cantor-Bendixson rank. To apply Theorem 6.13, we show that any such metric space is well-orderable.

Lemma 6.20. *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 $(X, \mathcal{D})$ is well-orderable if so is $(\textsc{lim}(X), \mathcal{D})$.

Let $X_1 = X \setminus \textsc{lim}(X)$ and $X_2 = \textsc{lim}(X)$. Suppose $(X_2, \mathcal{D})$ 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 $(X, \mathcal{D})$. 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 $(X_2, \mathcal{D})$, 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 $X_i = \textsc{lim}(X, i) \setminus \textsc{lim}(X, i+1)$, $0 \le i \le n$. 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 6.21. *Consider the Lipschitz MAB/experts problem on a compact metric space $(X, \mathcal{D})$ such that $\textsc{lim}_N(X) = \emptyset$ for some N. Let $\mathcal{O}_i$ be the covering oracle for $X_i = \textsc{lim}(X, i) \setminus \textsc{lim}(X, i+1)$. 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 $(X, \mathcal{D})$ is f-tractable for every $f \in \omega(\log t)$;
(b) the Lipschitz experts problem on $(X, \mathcal{D})$ is 1-tractable, even with a double feedback.

In the rest of this section, consider the setting in Theorem 6.21. We describe the *exploration subroutine* EXPL′(), which is similar to EXPL() in Section 6.3 but does not use the ordering oracle. Then, we prove a version of Lemma 6.17 for EXPL′(). Once we have this lemma, the proof of Theorem 6.21 is identical to that of Theorem 6.13 (and is omitted).

ALGORITHM 5: Subroutine 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 δ_i-covering set S_i of X consisting of at most k points. Let $S = \cup_{l=1}^{n} S_l$. 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 $\mu_{\text{av}}(x) - \mu_{\text{av}}(y) > 2\,r$. 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, 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 6.22. *Fix a problem instance. Consider increasing functions $k, n, T : \mathbb{N} \to \mathbb{N}$ such that $r(t) := 4\sqrt{(\log T(t))\,/n(t)} \to 0$. Then for any sufficiently large t, with probability at least $1 - T^{-2}(t)$, the subroutine EXPL′$(k(t),\ n(t),\ r(t))$ returns an optimal strategy.*

Proof. Use the notation from Algorithm 5. 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 $|\mu_{\text{av}}(x) - \mu(x)| \le r(t)$. 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 μ 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 $(X, \mathcal{D})$ 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 $Y = \textsc{lim}(X, i^*)$.

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 that the δ_{i^*}-covering set 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: PROOF OF THEOREM 1.10

We prove that Lipschitz bandits/experts are $o(t)$-tractable if and only if the completion of the metric space is compact. More formally, we prove Theorem 1.10 (which subsumes Theorem 1.8). We restate the theorem below for the sake of convenience.

Theorem (Theorem 1.10 restated). *The Lipschitz experts problem on metric space $(X, \mathcal{D})$ 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.*

First, we reduce the theorem to that on complete metric spaces, see Appendix B. In what follows, 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 $(X, \mathcal{D})$ and use an extension of algorithm `UniformMesh` (described in the Introduction). 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 Auer et al. [11] 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 Reference [60] (essentially, bounded covering dimension), it suffices to tune each t_i to the corresponding t_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 $N_1, \ldots, N_{i+1}$.

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 `UCB1`, it is easy to see [60] that

$$R_k(t) \leq O\left(\sqrt{N_k\, t \log t}\right) + \epsilon_k\, t \leq \epsilon_k \, \max(t_k^*, t), \quad \text{where} \quad t_k^* = 2\, \frac{N_k}{\epsilon_k^2} \log \frac{N_k}{\epsilon_k^2}. \tag{39}$$

Let us specify phase durations t_i. They are defined very differently from the ones in Reference [60]. In particular, in Reference [60] 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 $N_1, \ldots, N_{i+1}$. Specifically, we define the t_i's inductively as follows:

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

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

Lemma 7.1. *Consider the Lipschitz MAB problem on a compact and complete metric space $(X, \mathcal{D})$. 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 Equation (39). 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 Equation (39) and the facts that $t_k \geq t_k^*$ and $t_k \geq 2\,s_{k-1}$.

For the general case, let $T = s_{k-1} + t$, where $t \in (0, t_k)$. Then by Equation (39), 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 $(X, \mathcal{D})$ 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:

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

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 $(X, \mathcal{D})$ 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 $s_1, s_2, \ldots$, 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 $t_1, t_2, \ldots \in \mathbb{N}$ 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 $I_k = \{m_k + 1, \ldots, m_{k+1}\}$. The intervals I_k form a partition of $\mathbb{N}$ into sets of sizes $4^{t_1}, 4^{t_2}, \ldots$. 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 $B(s_i, r)$:

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

If $J \subseteq \mathbb{N}$ is any set of natural numbers, then 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 π_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 $r_k t_k \geq k g(t_k)$. 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 $j_1, j_2, \ldots$ 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 σ_i sampled at time t when sampling the sequence of i.i.d. payoff functions π_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 $\sigma(i,1) = \sigma(i,2) = \cdots = \sigma(i,t) = 1$ 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 $n = 2^{t_k}$, so that $|I_k| = n^2$, we have

$$\begin{aligned}\mathbb{P}_J\left[\, |S(k,t)| \leq n \,\right] &\leq \sum_{R \subset I_k,\ |R| = n} \mathbb{P}_J[\, S(k,t) \subseteq R \,] \\ &= \binom{n^2}{n} \left(1 - 2^{-t}\right)^{n^2 - n} < \left(n^2 \cdot \left(1 - 2^{-t}\right)^{n-1}\right)^n \\ &< \exp\left(n(2\ln(n) - (n-1)/2^t)\right). \end{aligned} \tag{40}$$

As long as $t \leq t_{k-1}$, the relation $t_k > 2t$ implies $(n-1)/2^t > \sqrt{n}$ so the expression Equation (40) is bounded above by $\exp(-n\sqrt{n} + 2n\ln(n))$, which equals $\exp\left(-8^{t_k} + 2\ln(4) t_k 4^{t_k}\right)$ and is in turn bounded above by $\exp\left(-8^{t_k}/2\right)$.

Let $B(j_{>k})$ denote the union $B(j_{k+1}) \cup B(j_{k+2}) \cup \ldots$, 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 $1, \ldots, t$. We have already demonstrated that for all $t \leq t_k$,

$$\Pr_{\mathbb{P}_J \in \mathcal{P}}(x_t \in B(j_{>k})) \leq 2^{-t_{k+1}} + \sum_{\ell > k} \exp(-8^{t_\ell}/2) < 2^{1 - t_{k+1}}, \tag{41}$$

where the term $2^{-t_{k+1}}$ accounts for the event that $S(\ell, t)$ has at least $2^{t_{k+1}}$ elements, where ℓ in the index of the set I_ℓ containing the number i such that $x_t \in B(i)$, if such an i exists. Equation (41) implies the bound $\mathbb{E}_{\mathbb{P}_J \in \mathcal{P}}[N(t_k, k)] < t_k \cdot 2^{1-t_{k+1}}$. By Markov's inequality, the probability that $N(t_k, k) > t_k/2$ is less than $2^{2-t_{k+1}}$. 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

This section concerns polynomial regret results for Lipschitz experts in metric spaces of (very) high dimension: Theorems 1.11, 1.12, and 1.13, as outlined in Section 1.5.

8.1 The Uniform Mesh (Proof of Theorem 1.11)

We start with a version of algorithm UniformMesh discussed in the Introduction.[36] This algorithm, called UniformMeshExp(b), 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 δ-hitting set[37] 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 UniformMeshExp is naturally described in terms of the log-covering dimension (see Equation (2)). The proof is based on the argument from Kleinberg [60]. We restate it here for the sake of completeness, and to explain how the new dimensionality notion is used.

Theorem 8.1. *Consider the Lipschitz experts problem on a metric space $(X, \mathcal{D})$. For each $b >$ LCD(X), algorithm* UniformMeshExp(b) *achieves regret $R(t) = O(t^{1-1/(b+2)})$.*

Proof. Let $N_\delta = N_\delta(X)$, and let μ 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 δ-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,

$$\Pr[|\mu_T(x) - \mu(x)| < r_T] > 1 - (TN_\delta)^{-3}, \quad \text{where} \quad r_T = \sqrt{8\,\log(T\,N_\delta)\,/T} < 2\delta. \tag{42}$$

Note that δ is chosen specifically to ensure that $r_T \le O(\delta)$.

We can neglect the regret incurred when the event in Equation (42) does not hold for some $x \in S$. From now on, let us assume that the event in Equation (42) 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,

$$\mu(x_i^*) \ge \mu_T(x_i^*) - 2\delta \ge \mu_T(x) - 2\delta \ge \mu(x) - 4\delta \ge \mu(x^*) - 5\delta.$$

Thus, the total regret R_{i+1} accumulated in phase $i + 1$ is

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

Thus, the total regret summed over phases is as claimed. □

8.2 Uniformly Lipschitz Experts (Proof of Theorem 1.12)

We now turn our attention to the *uniformly Lipschitz experts problem*, a restricted version of the Lipschitz experts problem in which a problem instance $(X, \mathcal{D}, \mathbb{P})$ satisfies a further property that each function $f \in \texttt{support}(\mathbb{P})$ is itself a Lipschitz function on $(X, \mathcal{D})$. We show that for this version, UniformMeshExp obtains a significantly better regret guarantee, via a more involved analysis. As we will see in the next section, for a wide class of metric spaces including ϵ-uniform tree metrics there is a matching upper bound.

[36] A similar algorithm has been used by Gupta et al. [50] to obtain regret $R(T) = O(\sqrt{T})$ for metric spaces of finite covering dimension.

[37] A subset $S \subset X$ is a *δ-hitting set* for $Y \subset X$ if $Y \subset \cup_{x\in S}\, B(x,\ \delta)$.

Theorem 8.2. *Consider the uniformly Lipschitz experts problem with full feedback. Fix a metric space $(X, \mathcal{D})$. For each $b > \mathtt{LCD}(X)$ such that $b \geq 2$, $\mathtt{UniformMeshExp}(b-2)$ 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 μ 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 δ-hitting set chosen in this phase. (The specific choice of δ 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 δ-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 $y_i^* = \operatorname{argmax}(\mu, S)$ be the optimal strategy in the chosen sample, and let $x_i^* = \operatorname{argmax}(\nu, S)$ be the algorithm's "best guess." The crux is to show that

$$\Pr[\, \mu(y_i^*) - \mu(x_i^*) \leq O(\delta \log T)\,] > 1 - T^{-3}. \tag{43}$$

Once Equation (43) is established, the remaining steps is exactly as the proof of Theorem 8.1.

Proving Equation (43) 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 $(X, \mathcal{D}, \mathbb{P})$ 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

$$\Pr\left[\, |\Delta(x,y)| \leq \mathcal{D}(x,y)\, \sqrt{8\, \log(T\,N)/T} \,\right] > 1 - (TN)^{-3}. \tag{44}$$

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

To take advantage of Equation (44), let us define the following structure that we call the *covering tree* of the metric space $(X, \mathcal{D})$. 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 $C(u)$ 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 $X(v)$, $v \in C(u)$ 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

$$\begin{aligned} \sigma(u) &= \operatorname{argmax}(\mu,\ \mathcal{X}(u) \cap S), \\ \rho(u) &= \operatorname{argmax}(\nu,\ \mathcal{X}(u) \cap S), \end{aligned} \tag{45}$$

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 $v, w \in C(u)$, we have $|\,\Delta(\sigma(v),\ \sigma(w))\,| \leq 4\delta$.
(ii) for any $x, y \in S$ such that $\mathcal{D}(x, y) \leq \delta$, we have $|\Delta(x, y)| \leq 4\delta$.

Claim 8.3. *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 $j \le n$ and apply the Chernoff-style bound Equation (44) with $N = |\mathcal{T}_j|$ and $(x, y) = (\sigma(v), \sigma(w))$. Since $|\mathcal{T}_l| \le 2^{2^{lb}}\,|\mathcal{T}_{l-1}|$ for each sufficiently large l, it follows that $\log|\mathcal{T}_j| \le C + \sum_{l=1}^{j} 2^{lb} \le C + \frac{4}{3}\,2^{jb}$, 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, δ and n), the "slack" in Equation (44) is at most 4δ:

$$\mathcal{D}(x,y)\sqrt{8\,\log(T\,N)/T} \le 3\,\mathcal{D}(x,y)\sqrt{\log(N)/T} \le 4\,2^{-j}\sqrt{2^{bj}/2^{bn}} = 4\delta\,2^{-(n-j)(b-2)/2} \le 4\delta.$$

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

To prove (ii), apply Equation (44) 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 $\Delta^*(u) \ge 0$. The central argument of this proof is the following upper bound on $\Delta^*(u)$.

Claim 8.4. *In a clean phase, $\Delta^*(u) \le O(\delta)(n-j)$ for each $j \le n$ 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 δ. 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 $u, v \in C(u)$ such that $\sigma(u) \in X(v)$ and $\rho(u) \in X(w)$. Since the tie-breaking rules in Equation (45) 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 $\Delta^*(w) \le O(\delta)(n-j-1)$ by induction, and $\Delta(\sigma(v),\, \sigma(w)) \le 4\delta$ by part (i) in the definition of the clean phase. □

To complete the proof of Equation (43), 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 8.4 (applied for $\mathcal{T}_0 = \{u_0\}$), we have

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

8.3 Regret Characterization (Proof of Theorem 1.13)

As it turns out, the log-covering dimension is not the right notion to characterize optimal regret for arbitrary metric spaces. We need a more refined version: the *max-min-log-covering dimension*, defined in Equation (3), similar to the max-min-covering dimension.

Theorem 8.5. *Fix a metric space $(X, \mathcal{D})$ and let $b =$ `MaxMinLCD`(X). The Lipschitz experts problem on $(X, \mathcal{D})$ is (t^γ)-tractable for any $\gamma > \frac{b+1}{b+2}$, and not (t^γ)-tractable for any $\gamma < \frac{b-1}{b}$.*

For the lower bound, we use a suitably "thick" version of the ball-tree from Section 6.2 in conjunction with the (ϵ, δ, k)-ensemble idea from Section 6.2, see Section 8.3.1. For the algorithmic result, we combine the "naive" experts algorithm (`UniformMeshExp`) with (an extension of) the *transfinite fat decomposition* technique from Section 5, see Section 8.3.2.

The lower bound in Theorem 8.5 holds for the uniformly Lipschitz experts problem. It follows that the upper bound in Theorem 8.2 is optimal for metric spaces such that

$\texttt{MaxMinLCD}(X) = \texttt{LCD}(X)$, e.g., for ϵ-uniform tree metrics. In fact, we can plug the improved analysis of `UniformMeshExp` from Theorem 8.2 into the algorithmic technique from Theorem 8.5 and obtain a matching upper bound in terms of the `MaxMinLCD`. Thus (in conjunction with Theorem 1.9), we have a complete characterization for regret:

Theorem 8.6. *Consider the uniformly Lipschitz experts problem with full feedback. Fix a metric space* $(X, \mathcal{D})$ *with uncountably many points, and let* $b = \texttt{MaxMinLCD}(X)$*. The problem on* $(X, \mathcal{D})$ *is* (t^γ)*-tractable for any* $\gamma > \max(\frac{b-1}{b}, \frac{1}{2})$*, and not* (t^γ)*-tractable for any* $\gamma < \max(\frac{b-1}{b}, \frac{1}{2})$*.*

The proof of the upper bound in Theorem 8.6 proceeds exactly that in Theorem 8.5, except that we use a more efficient analysis of `UniformMeshExp`.

8.3.1 The `MaxMinLCD` *Lower Bound: Proof for Theorem 8.6.* If $\texttt{MaxMinLCD}(X) = d$, and $\gamma < \frac{d-1}{d}$, then let us first fix constants b and c such that $b < c < d$ and $\gamma < \frac{b-1}{b}$. Let $Y \subseteq X$ be a subspace such that $c \leq \inf\{\texttt{LCD}(Z) : \text{open, nonempty } Z \subseteq Y\}$. We will repeatedly use the following packing lemma that relies on the fact that $b < \texttt{LCD}(U)$ for all nonempty subsets $U \subseteq Y$.

Lemma 8.7. *For any nonempty open* $U \subseteq Y$ *there exists* $r_0 > 0$ *such that for all* $r \in (0, r_0)$*,* U *contains more than* $2^{r^{-b}}$ *disjoint balls of radius* r*.*

Proof. Let r_0 be a positive number such that for all positive $r < r_0$, every covering of U requires more than $2^{r^{-b}}$ balls of radius $2r$. Such an r_0 exists, because $\texttt{LCD}(U) > b$. Now for any positive $r < r_0$ let $\mathcal{P} = \{B_1, B_2, \ldots, B_M\}$ 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 $\{B_i^+ \mid 1 \leq i \leq M\}$, i.e., they form a covering of Y. Hence, $M \geq 2^{r^{-b}}$ as desired. □

Using the packing lemma, we recursively construct a ball-tree on metric space $(Y, \mathcal{D})$ with very high node degrees. Specifically, let us say that a ball-tree has *log-strength* b if each tree node with children of radius r has at least $2^{r^{-b}}$ children. For convenience, all tree nodes of the same depth will have the same radius r_i. Then each node at depth $i-1$ has at least $n_i = \lceil 2^{r_i^{-b}} \rceil$ children.

Claim 8.8. *There exists a ball-tree* T *on* $(Y, \mathcal{D})$ *with log-strength* b*, in which all tree nodes of the same depth* i *have the same radius* r_i*.*

Proof. The root of the ball tree is centered at any point in Y and has radius $r_0 = \frac{1}{4}$. For each successive $i \geq 1$, let $r_i \in (0, r_{i-1}/4)$ be a positive number small enough that for every depth $i-1$ tree node $w = (x, r_{i-1})$, the sub-ball $B(x, r_{i-1}/2)$ contains $n_i = \lceil 2^{r_i^{-b}} \rceil$ disjoint balls of radius r_i. (Denote by $\mathcal{B}_w$ the collection of the corresponding disjoint extensive-form balls.) Such r_i exists by Lemma 8.7. The set of children of w is defined to be $\mathcal{B}_w$. □

We re-use Construction 6.7 for metric space $(Y, \mathcal{D})$ and ball-tree T, with $\delta_i \equiv \frac{1}{3}$. Thus, we construct a problem instance $\mathbb{P}_\lambda$ for each lineage over λ, and a distribution $\mathcal{P}_T$ over problem instances $\mathbb{P}_\lambda$. Recall that a problem instance is a distribution over (deterministic) payoff functions $\pi : X \to [0, 1]$, which are Lipschitz by Lemma 5.8.

Fix a complete lineage λ, and let $\mathbf{w}(\lambda) = (w_0, w_1, \ldots)$ be the associated end of the ball-tree. For each $i \geq 1$, let B_i be the ball in $(Y, \mathcal{D})$ corresponding to tree node w_i. Let $\mu = \mathbb{E}_{\pi \sim \mathbb{P}_\lambda}[\pi]$ be the expected payoff function corresponding to $\mathbb{P}_Q$. Then then μ achieves its maximum value

$\frac{1}{2} + \frac{1}{18} \sum_{i=0}^{\infty} r_i$ at the unique point $x^* \in \cap_{i=0}^{\infty} B_i$. At any point $x \notin B_j$, we have

$$\mu(x^*) - \mu(x) \geq \left(\frac{1}{18} \sum_{i=j}^{\infty} r_i \right) - \left(\frac{1}{18} \sum_{i=j+1}^{\infty} r_i \right) = \frac{1}{18} r_j.$$

We now finish the lower bound proof as in the proof of Lemma 6.9. Fix depth $i-1$ node w in the ball-tree, and let $w^1, w^2, \ldots, w^{n_i}$ be the children of w in the ball-tree. Let $\lambda(w)$ be the unique child of w contained in the lineage λ. Consider the sets $\lambda_0 = \lambda \setminus \lambda(w)$ and $\lambda_j = \lambda_0 \cup \{w^j\}$ for $j = 1, 2, \ldots, n_i$. By Corollary 6.11, the distributions $(\mathbb{P}_{\lambda_0}, \mathbb{P}_{\lambda_1}, \ldots, \mathbb{P}_{\lambda_{n_i}})$ constitute an (ϵ, δ, k)-ensemble for $\epsilon = r_i/18$, $\delta = \frac{1}{3}$, 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 of

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

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

8.3.2 The `MaxMinLCD` *Upper Bound: Proofs for Theorem 8.5 and Theorem 8.6.* First, let us incorporate the analysis of `UniformMeshExp`(b) via the following lemma.

Lemma 8.9. *Consider an instance $(X, \mathcal{D}, \mathbb{P})$ of the Lipschitz experts problem, and let $x^* \in X$ be an optimal point. Fix subset $U \subset X$, which contains x^*, and let $b >$ `LCD`(U). Then for any sufficiently large T and $\delta = T^{-1/(b+2)}$ the following holds:*

(a) Let S be a δ-hitting set for U of cardinality $|S| \leq N_\delta(U)$. 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 a 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 Section 5 with respect to the log-covering dimension rather than the covering dimension.

Definition 8.10. Fix a metric space $(X, \mathcal{D})$. Let β denote an arbitrary ordinal. A *transfinite LCD decomposition* of depth β 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 $\{$ordinals $\nu \leq \beta$: V 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 `LCD`$(S_\lambda \setminus U) \leq b$.

The existence of suitable decompositions and the connection to `MaxMinLCD` is derived exactly as in Proposition 5.28.

Lemma 8.11. *For every compact metric space $(X, \mathcal{D})$,* `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 $(X, \mathcal{D})$ and $b >$ `MaxMinLCD`(X), and let $\{S_\lambda\}_{0 \leq \lambda \leq \beta}$ be a transfinite LCD decomposition of depth β and dimension b. For each $x \in X$, let the *depth* of x be the maximal ordinal λ such that $x \in S_\lambda$. (Such an ordinal exists by Definition 8.10(b).)

Access to the metric space. The algorithm requires two oracles: the *depth oracle* `Length`$(\cdot)$ and the *covering oracle* $\mathcal{D} -$ `Cov`$(\cdot)$. Both oracles input a finite collection $\mathcal{F}$ of open balls $B_0, B_1, \ldots, B_n$, given via the centers and the radii, and return a point in X. Let B be the union of these balls, and let $\overline{B}$

be the closure of B. A call to oracle $\texttt{Length}(\mathcal{F})$ returns an arbitrary point $x \in \overline{B} \cap S_\lambda$, where λ is the maximum ordinal such that S_λ intersects $\overline{B}$. (Such an ordinal exists by Definition 8.10(b).) Given a point $y^* \in X$ of depth λ, a call to oracle $\mathcal{D} - \texttt{Cov}(y^*, \mathcal{F})$ either reports that B covers S_λ, or it returns an arbitrary point $x \in S_\lambda \setminus B$. A call to $\mathcal{D} - \texttt{Cov}(\emptyset, \mathcal{F})$ is equivalent to the call $\mathcal{D} - \texttt{Cov}(y^*, \mathcal{F})$ for some $y^* \in S_0$.

The covering oracle will be used to construct δ-nets as follows. First, using successive calls to $\mathcal{D} - \texttt{Cov}(\emptyset, \mathcal{F})$ one can construct a δ-net for X. Second, given a point $y^* \in X$ of depth λ and a collection of open balls whose union is B, using successive calls to $\mathcal{D} - \texttt{Cov}(y^*, \cdot)$ one can construct a δ-net for $S_\lambda \setminus B$. The second usage is geared toward 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 8.10(c), we have $\texttt{LCD}(S_\lambda \setminus B) < b$, and one can apply Lemma 8.9.

The algorithm. Our algorithm proceeds in phases $i = 1, 2, 3, \ldots$ 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 $j = 0, 1, 2, \ldots$, 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 $\texttt{Length}(\mathcal{F})$, where

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

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

$$\{B(x, r) : \ x \in \mathcal{N} \quad \text{and} \quad \Delta_T(x) > 2(r_T + r)\}, \tag{46}$$

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

$$\Pr[|\mu_T(x) - \mu(x)| < r_T] > 1 - (T\,|\mathcal{N}|)^{-3} \quad \text{for each } x \in \mathcal{N}. \tag{47}$$

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) λ, algorithm uses the covering oracle to construct a δ-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.[38]

Sketch of the analysis. The proof roughly follows that of Theorem 5.2. Call a phase *clean* if the event in Equation (47) 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

[38] The interesting case here is $|\mathcal{N}'| \le Q_T$. If $\mathcal{N}'$ contains too many points, then the choice of A_i is not essential for the analysis.

argument is similar to the one in Lemma 6.17. 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 $O(\delta logT)$ of the optimum. The reason is that, letting λ be the depth of y^*_{i-1}, one can show that for a sufficiently large T the set B (defined in Equation (46)) contains $S_{\lambda+1}$ and does not contain some optimal strategy. By definition of the transfinite LCD decomposition, we have $\mathtt{LCD}(S_\lambda \setminus U) < b$, so in our construction the quota Q_T on the number of active points permits A_i to be a δ-cover of $S_\lambda \setminus U$. Now, we can use Lemma 8.9 to guarantee the "quality" of x^*_i. The final regret computation is similar to the one in the proof of Theorem 8.1.

9 CONCLUSIONS

Kleinberg et al. [64] (i.e., Sections 4 and 5 of this article) introduced the Lipschitz MAB problem and motivated a host of open questions. Many of these questions have been addressed in the follow-up work, including Kleinberg and Slivkins [63] (i.e., the rest of this article), and the work described in Section 2. Below, we describe the current state of the open questions.

First, the adaptive refinement technique from Section 4 can potentially be used in other settings in explore-exploit learning where one has side information on similarity between arms. Specific potential applications include adversarial MAB, Gaussian Process Bandits, and dynamic pricing. Also, stronger analysis of this technique appears possible in the context of ranked bandits (see Slivkins et al. [92] for details).

Second, it is desirable to consider MAB with more general structure on payoff functions. A particularly attractive target would be structures that subsume Lipschitz MAB and Linear MAB.

Third, a recurring theme in algorithm design is structural results that assert that a problem instance either has simple structure, or it contains a specific type of complex substructure that empowers the lower bound analysis. Our work contributes another example of this theme, in the form of dichotomy results in point-set topology (e.g., existence of a transfinite fat decomposition versus existence of a ball tree). It would potentially be interesting to find other applications of this technique.

APPENDIX

A KL-DIVERGENCE TECHNIQUES

All lower bounds in this article heavily use the notion of Kullback-Leibler divergence (*KL-divergence*). Our usage of the KL-divergence techniques is encapsulated in several statements in the body of the article (Theorem 5.7, Theorem 6.12, and Claim 6.18), whose proofs are fleshed out in this Appendix and may be of independent interest. To make this appendix more self-contained, we restate the relevant definitions and theorem statements from the body of the article, and provide sufficient background.

A.1 Background

Definition A.1. Let Ω be a finite set with two probability measures p, q. Their *KL-divergence* is the sum

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

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 Ω and taking values in some set Γ, then the *conditional KL-divergence* of p and q given Y is the sum

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

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 Ω provided that q is absolutely continuous with respect to p. For details, see Reference [61], Chapter 2.7. The following lemma summarizes some standard facts about KL-divergence; for proofs, see References [39, 61].

Lemma A.2. *Let p, q be two probability measures on a measure space $(\Omega, \mathcal{F})$ and let Y be a random variable defined on Ω and taking values in some finite set Γ. Define a pair of probability measures p_Y, q_Y on Γ by specifying that $p_Y(y) = p(Y = y), q_Y(y) = q(Y = y)$ for each $y \in \Gamma$. Then,*

$$\mathsf{KL}(p; q) = \mathsf{KL}(p; q \mid Y) + \mathsf{KL}(p_Y; q_Y),$$

and $\mathsf{KL}(p; q \mid Y)$ is non-negative.

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 A.3. *Let Ω be a sample space, and suppose p, q are two probability measures on Ω^n, the set of n-tuples of elements of Ω. For a sample point $\vec{\omega} \in \Omega^n$, let ω^i denote its first i components. If p^i, q^i denote the probability measures induced on Ω^i by p (respectively, q), then*

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

Proof. For $m = 1, 2, \ldots, n$, the formula $\mathsf{KL}(p^m; q^m) = \sum_{i=1}^{m} \mathsf{KL}(p^i; q^i \mid \omega^{i-1})$ follows by induction on m, using Lemma A.2. □

The following three lemmas will also be useful in our lower bound argument. They may have appeared in the literature, but we cannot provide specific citations. We provide proofs for the sake of completeness. Here and henceforth we will use the following notational convention: for real numbers $a, b \in [0, 1]$, $\mathsf{KL}(a; b)$ denotes the KL-divergence $\mathsf{KL}(p; q)$ where p, q are probability measures on $\{0, 1\}$ such that $p(\{1\}) = a$, $q(\{1\}) = b$. In other words,

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

Lemma A.4. *For any $0 < \epsilon < y \leq 1$, $\mathsf{KL}(y - \epsilon; y) < \epsilon^2 / y(1 - y)$.*

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

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

□

Lemma A.5. *Let Ω be a sample space with two probability measures p, q whose KL-divergence is κ. 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 $\kappa = \mathsf{KL}(p;q)$ 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 A.2 with Y as the indicator random variable of $\mathcal{E}$, we obtain

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

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 A.6. *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, $\mathsf{KL}(p;q) < \delta^2$.

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 Ω, 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 $x \geq -\frac{1}{2}$.

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

□

A.2 Bandit Lower Bound via (ϵ, k)-ensembles

We consider an MAB problem with i.i.d. payoffs where the algorithm is given a set of arms X and a collection $\mathcal{F}$ of feasible payoff functions $X \to [0,1]$. We call it the *feasible MAB problem* on $(X, \mathcal{F})$. We will consider 0-1 payoffs; then for a problem instance with payoff function $f \in \mathcal{F}$, the reward from each action $x \in X$ is 1 with probability $f(x)$, and 0 otherwise.

Definition (Definition 5.6, restated). Consider the feasible MAB problem on $(X, \mathcal{F})$. An (ϵ, k)-*ensemble* is a collection of subsets $\mathcal{F}_1, \ldots, \mathcal{F}_k \subset \mathcal{F}$ such that there exist mutually disjoint subsets $S_1, \ldots, S_k \subset X$ and a function $\mu_0 : X \to [\frac{1}{3}, \frac{2}{3}]$ such that for each $i = 1 \ldots k$ and each function $\mu_i \in \mathcal{F}_i$ the following holds: (i) $\mu_i \equiv \mu_0$ on each S_ℓ, $\ell \neq i$, and (ii) $\sup(\mu_i, S_i) - \sup(\mu_0, X) \geq \epsilon$, and (iii) $0 \leq \mu_i - \mu_0 \leq 2\epsilon$ on S_i.

Theorem (Theorem 5.7, restated). *Consider the feasible MAB problem with 0-1 payoffs. Let $\mathcal{F}_1, \ldots, \mathcal{F}_k$ be an (ϵ, k)-ensemble, where $k \geq 2$ and $\epsilon \in (0, \frac{1}{24})$. Then for any $t \leq \frac{1}{128}\, k\, \epsilon^{-2}$ and any bandit algorithm there exist at least $k/2$ distinct i's such that the regret of this algorithm on any payoff function from $\mathcal{F}_i$ is at least $\frac{1}{60}\,\epsilon t$.*

Proof. Let us specify the notation. Let $\Omega = X \times \{0, 1\}$. Since we assume 0-1 payoffs, the t-step history of play of a bandit algorithm $\mathcal{A}$ can be expressed by an element of Ω^t, indicating the sequence of arms selected and payoffs received. Thus, an algorithm $\mathcal{A}$ and a payoff function μ together determine a probability distribution on Ω^t for every natural number t. Fix any (possibly randomized) algorithm $\mathcal{A}$ and consider the distribution p determined by $\mathcal{A}$ when the payoff function is μ_0. Recall the mutually disjoint sets $S_1, S_2, \ldots, S_k$ in the definition of an (ϵ, k)-ensemble. For $1 \leq i \leq k$ and $1 \leq u \leq t$, let $Y_{i,u}$ be the indicator random variable of the event $x_u \in S_i$, where x_u denotes the arm selected by $\mathcal{A}$ at time u. Let $Z_i = \sum_{u=1}^{t} Y_{i,u}$.

Since $\sum_{i=1}^{k}\ \mathbb{E}_p\,[Z_i] \leq t$, there must be at least $k/2$ indices i such that $\mathbb{E}_p[Z_i] \leq t/k \leq 1/128\,\epsilon^2$. Fix one such i, and an arbitrary $\mu_i \in \mathcal{F}_i$. In what follows, we will show that $R_{(\mathcal{A},\,\mu_i)}(t) \geq \epsilon t/60$.

Let $(x_u, y_u) \in X \times \{0, 1\} = \Omega$ denote the arm selected and the payoff received at time u, and let q denote the distribution on Ω^t determined by $\mathcal{A}$ and μ_i. We have

$$\begin{aligned}
\mathsf{KL}(p^u; q^u \,|\, \omega^{u-1}) &= \sum_{\omega^u \in \Omega^u} p^u(\omega^u)\ \ln\left(\frac{p^u(\omega^u \,|\, \omega^{u-1})}{q^u(\omega^u \,|\, \omega^{u-1})}\right)\\
&= \sum_{\omega^u \in \Omega^u} p^u(\omega^u)\ \ln\left(\frac{p^u(x_u \,|\, \omega^{u-1})}{q^u(x_u \,|\, \omega^{u-1})} \cdot \frac{p^u(y_u \,|\, x_u, \omega^{u-1})}{q^u(y_u \,|\, x_u, \omega^{u-1})}\right)\\
&= \sum_{\omega^u \in \Omega^u} p^u(\omega^u)\ \ln\left(\frac{p^u(y_u \,|\, x_u, \omega^{u-1})}{q^u(y_u \,|\, x_u, \omega^{u-1})}\right)
\end{aligned}$$

[the distribution of x_u given ω^{u-1} depends only on $\mathcal{A}$, not on distribution p versus q]

$$\begin{aligned}
&= \sum_{\omega^{u-1} \in \Omega^{u-1}} \int_{x_u \in X} \sum_{y_u \in \{0,1\}} p^u(y_u \,|\, x_u, \omega^{u-1})\ \ln\left(\frac{p^u(y_u \,|\, x_u, \omega^{u-1})}{q^u(y_u \,|\, x_u, \omega^{u-1})}\right)\ \mathrm{d}\,p^u(\,\cdot\,,\ \omega^{u-1})\\
&= \sum_{\omega^{u-1} \in \Omega^{u-1}} \int_{x_u \in X} \mathsf{KL}(\mu_0(x_u);\ \mu_i(x_u) \,|\, x_u, \omega^{u-1})\ \mathrm{d}\,p^u(\,\cdot\,,\ \omega^{u-1})\\
&= \sum_{\omega^{u-1} \in \Omega^{u-1}} \int_{x_u \in S_i} \mathsf{KL}(\mu_0(x_u)); \mu_i(x_u) \,|\, x_u, \omega^{u-1})\ \mathrm{d}\,p^u(\,\cdot\,,\ \omega^{u-1})
\end{aligned}$$

[because $\mu_0 = \mu_i(x_u)$ when $x_u \notin S_i$.]

$$\leq \sum_{\omega^{u-1} \in \Omega^{u-1}} \int_{x_u \in S_i} \frac{4\,\epsilon^2}{\mu_i(x_u)(1 - \mu_i(x_u))}\ \mathrm{d}\,p^u(\,\cdot\,,\ \omega^{u-1})$$

[by Lemma A.4 and property (iii) in the definition of "ensemble"]

$$\leq p^u(x_u \in S_i) \cdot \frac{4\epsilon^2}{3/16}.$$

The last inequality holds, because $\mu_i(x_u) \in [\frac{1}{3}, \frac{3}{4}]$. The latter holds by property (iii) in the definition of the "ensemble" and the assumptions that $\mu_0 \in [\frac{1}{3}, \frac{2}{3}]$ and $\epsilon \leq \frac{1}{24}$.

Now, we can write

$$\begin{aligned}
\mathsf{KL}(p; q) = \sum_{u=1}^{t} \mathsf{KL}(p^u; q^u \,|\, \omega^{u-1}) &\leq \left(\sum_{u=1}^{t} p^u(x_u \in S_i)\right) \cdot \frac{64\ \epsilon^2}{3}\\
&= \mathbb{E}\,[Z_i] \cdot \frac{64\ \epsilon^2}{3}\ \ \leq \frac{1}{128\,\epsilon^2} \cdot \frac{64\epsilon^2}{3} = \frac{1}{6}.
\end{aligned}$$

Let $\mathcal{E}$ be the event that $Z_i \leq \frac{5t}{3k}$. By Markov's inequality, $p(\mathcal{E}) \geq 0.4$. Now, using Lemma A.5 along with the bound $\mathsf{KL}(p;q) \leq 1/6$, a short calculation leads to the bound $q(\mathcal{E}) \geq 0.1$, and consequently,

$$\begin{aligned}\mathbb{E}_q[t - Z_i] &\geq q(\mathcal{E})\mathbb{E}_q[t - Z_i \mid \mathcal{E}] \\ &\geq 0.1 \cdot \left(t - \frac{5t}{3k}\right) \geq 0.1 \cdot \left(t - \frac{5t}{6}\right) = \frac{t}{60}.\end{aligned}$$

Assuming the payoff function is μ_i, the regret of algorithm $\mathcal{A}$ increases by ϵ each time it chooses a arm $x_u \notin S_i$. Hence,

$$R_{(\mathcal{A},\, \mu_i)}(t) \geq \epsilon \mathbb{E}_q[t - Z_i] \geq \epsilon t/60. \qquad \square$$

A.3 Experts Lower Bound via (ϵ, δ, k)-ensembles

We consider the *feasible experts problem*, in which one is given an action set X along with a collection $\mathcal{F}$ of Borel probability measures on the set $[0,1]^X$ of functions $\pi : X \to [0,1]$. A problem instance of the feasible experts problem consists of a triple $(X, \mathcal{F}, \mathbb{P})$ where X and $\mathcal{F}$ are known to the algorithm, and $\mathbb{P} \in \mathcal{F}$ is not. In each round the payoff function π is sampled independently from $\mathbb{P}$, so that for each action $x \in X$ the (realized) payoff is $\pi(x)$.

Definition A.7 (Definition 6.10, Restated). Consider a set X and a $(k+1)$-tuple $\vec{\mathbb{P}} = (\mathbb{P}_0, \mathbb{P}_1, \ldots, \mathbb{P}_k)$ of Borel probability measures on $[0,1]^X$, the set of $[0,1]$-valued payoff functions π 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 (ϵ, δ, k)-*ensemble* if there exist pairwise disjoint subsets $S_1, S_2, \ldots, S_k \subseteq X$ for which the following properties hold:

(i) for every $i > 0$ and every event $\mathcal{E}$ in the Borel σ-algebra of $[0,1]^X$, we have

$$1 - \delta < \mathbb{P}_0(\mathcal{E})/\mathbb{P}_i(\mathcal{E}) < 1 + \delta.$$

(ii) for every $i > 0$, we have $\sup(\mu_i, S_i) - \sup(\mu_i, X \setminus S_i) \geq \epsilon$.

Theorem A.8 (Theorem 6.12, Restated). *Consider the feasible experts problem on $(X, \mathcal{F})$. Let $\vec{\mathbb{P}}$ be an (ϵ, δ, k)-ensemble with $\{\mathbb{P}_1, \ldots, \mathbb{P}_k\} \subseteq \mathcal{F}$ 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 $R_{(\mathcal{A},\, \mathbb{P}_i)}(t) \geq \epsilon t/2$.*

Proof. Let $\Omega = [0,1]^X$. Using Property (i) of an (ϵ, δ, k)-ensemble combined with Lemma A.6, we find that $\mathsf{KL}(\mathbb{P}_i; \mathbb{P}_0) < \delta^2$.

Let $\mathcal{A}$ be an experts algorithm whose random bits are drawn from a sample space Γ with probability measure ν. 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 A.3), $\mathsf{KL}(p_i^s; p_0^s) < s\delta^2 < \ln(17k)/2$. 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 A.5 implies

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

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 $i \in \{1, \ldots, k\}$, the expected number of satisfactory $s \in \{1, \ldots, t\}$ is less than $t/4$, so by Markov's inequality, for at least half of the $i \in \{1, \ldots, k\}$, the number of satisfactory $s \in \{1, \ldots, t\}$ is less than $t/2$. Property (ii) of an (ϵ, δ, k)-ensemble guarantees that every unsatisfactory s contributes at least

ϵ 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 $R_{(\mathcal{A},\,\mathbb{P}_i)}(t) \geq \epsilon t/2$. □

A.4 Proof of Claim 6.18

Recall that in Section 6.4, we defined a pair of payoff functions μ_0, μ_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 A.4, $\mathsf{KL}(\mu_0(x); \mu_i(x)) < r_i^2/3$ for all $x \in X$, and $\mathsf{KL}(\mu_0(x); \mu_i(x)) = 0$ 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 μ_0, μ_i, this defines two different probability measures p_0^t, p_i^t on the set Ω^t of possible t-step histories. Let ω^t denote a sample point in Ω^t. The bounds derived in the previous paragraph imply that for any non-negative integer s,

$$\mathsf{KL}(p_0^{s+1}; p_i^{s+1} \mid \omega^s) < \frac{1}{3} r_i^2 \mathbb{P}_0(x_{s+1} \in B_i). \tag{48}$$

Summing Equation (48) for $s = 0, 1, \ldots, t-1$ and applying Lemma A.3, we obtain

$$\mathsf{KL}(p_0^t; p_i^t) < \frac{1}{3} r_i^2 \sum_{s=1}^{t} \mathbb{P}_0(x_s \in B_i) = \frac{1}{3} r_i^2 \mathbb{E}_0(N_i(t)), \tag{49}$$

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

The bound stated in Claim 6.18 now follows by applying Lemma A.5 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.

B REDUCTION TO COMPLETE METRIC SPACES

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

Lemma B.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 μ 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, then algorithm $\mathcal{A}_X$ instead plays $x = x(y,t)$. Let π be the observed payoff. Algorithm $\mathcal{A}_X$ draws an independent 0-1 random sample with expectation π, 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 μ 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" 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 $p(t) \to 1$. 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 $\mathcal{I}_\in$ (call these distributions q_1 and q_2). Let us consider and look at their KL-divergence. By the chain rule (See Lemma A.2), 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 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 C makes a mistake (on $\mathcal{I}_2$) with constant probability. □

Lemma B.2. *Consider the Lipschitz experts 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 B.1. □

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

C TOPOLOGICAL EQUIVALENCES: PROOF OF LEMMA 6.3

Let us restate the lemma, for the sake of convenience. Recall that it includes an equivalence result for compact metric spaces, and two implications for arbitrary metric spaces:

Lemma C.1. *For any compact metric space (X, d), the following are equivalent: (i) X is a countable set, (ii) (X, d) is well-orderable, (iii) no metric subspace of (X, d) is perfect. For an arbitrary metric space, we have (ii) $\iff$ (iii) and (i)$\Rightarrow$(ii), but not (ii)$\Rightarrow$(i).*

(compact metric spaces). Let us prove the assertions in the circular order.

(i) implies (iii). Let us prove the contrapositive: If (X, d) has a perfect subspace Y, then X is uncountable. We have seen that if (X, d) has a perfect subspace Y, then it has a ball-tree. Every end ℓ of the ball-tree (i.e., infinite path starting from the root) corresponds to a nested sequence of balls. The closures of these balls have the finite intersection property, hence their their intersection is non-empty. Pick an arbitrary point of the intersection and call if $x(\ell)$. Distinct ends ℓ, ℓ' correspond to distinct points $x(\ell), x(\ell')$, because if (y, r_y), (z, r_z) are siblings in the ball-tree, which are ancestors of ℓ and ℓ', respectively, then the closures of $B(y, r_y)$ and $B(z, r_z)$ are disjoint and they contain $x(\ell), x(\ell')$, respectively. Thus, we have constructed a set of distinct points of X, one for each end of the ball-tree. There are uncountably many ends, so X is uncountable.

(iii) implies (ii). Let β be some ordinal of strictly larger cardinality than X. Let us define a transfinite sequence $\{x_\lambda\}_{\lambda\leq\beta}$ of points in X using transfinite recursion,[39] by specifying that x_0 is any isolated point of X, and that for any ordinal $\lambda > 0$, x_λ is any isolated point of the subspace (Y_λ, d), where $Y_\lambda = X \setminus \{x_\nu : \nu < \lambda\}$, as long as Y_λ is nonempty. (Such isolated point exists, since by our assumption subspace (Y_λ, d) is not perfect.) If Y_λ is empty, then define, e.g., $x_\lambda = x_0$. Now, Y_λ is empty for some ordinal λ, because, otherwise, we obtain a mapping from X onto an ordinal β whose cardinality exceeds the cardinality of X. Let $\beta_0 = \min\{\lambda : Y_\lambda = \emptyset\}$. Then every point in X has been indexed by an ordinal number $\lambda < \beta_0$, and so we obtain a well-ordering of X. By construction, for every $x = x_\lambda$, we can define a radius $r(x) > 0$ such that $B(x, r(x))$ is disjoint from the set of points $\{x_\nu : \nu > \lambda\}$. Any initial segment S of the well-ordering is equal to the union of the balls $\{B(x, r(x)) : x \in S\}$, hence is an open set in the metric topology. Thus, we have constructed a topological well-ordering of X.

(ii) implies (i). Suppose we have a binary relation $\prec$ that is a topological well-ordering of (X, d). Let $S(n)$ denote the set of all $x \in X$ such that $B(x, \frac{1}{n})$ is contained in the set $P(x) = \{y : y \preceq x\}$. By the definition of a topological well-ordering, we know that for every x, $P(x)$ is an open set, hence $x \in S(n)$ for sufficiently large n. Therefore $X = \cup_{n\in\mathbb{N}} S(n)$. Now, the definition of $S(n)$ implies that every two points of $S(n)$ are separated by a distance of at least $1/n$. (If x and z are distinct points of $S(n)$ and $x \prec z$, then $B(x, \frac{1}{n})$ is contained in the set $P(x)$, which does not contain z, hence $d(x, z) \geq \frac{1}{n}$.) Thus, by compactness of (X, d) set $S(n)$ is finite. □

(ARBITRARY METRIC SPACES). For implications *(i)⇒(ii)* and *(iii)⇒(ii)*, the proof above does not in fact use compactness. An example of an uncountable but well-orderable metric space is $(\mathbb{R}, d)$, where d is a uniform metric. It remains to prove that *(ii)⇒(iii)*.

Suppose there exists a topological well-ordering $\prec$. For each subset $Y \subseteq X$ and an element $\lambda \in Y$ let $Y_\prec(\lambda) = \{y \in Y : y \preceq \lambda\}$ be the corresponding initial segment.

We claim that $\prec$ induces a topological well-ordering on any subset $Y \subseteq X$. We need to show that for any $\lambda \in Y$ the initial segment $Y_\prec(\lambda)$ is open in the metric topology of (Y, d). Indeed, fix $y \in Y_\prec(\lambda)$. The initial segment $X_\prec(\lambda)$ is open by the topological well-ordering property of X, so $B_X(y, \epsilon) \subset X_\prec(\lambda)$ for some $\epsilon > 0$. Since $Y_\prec(\lambda) = X_\prec(\lambda) \cap Y$ and $B_Y(y, \epsilon) = B_X(y, \epsilon) \cap Y$, it follows that $B_Y(y, \epsilon) \subset Y_\prec(\lambda)$. Claim proved.

Suppose the metric space (X, d) has a perfect subspace $Y \subset X$. Let λ be the $\prec$-minimum element of Y. Then, $Y_\prec(\lambda) = \{\lambda\}$. However, by the previous claim $\prec$ is a topological well-ordering of (Y, d), so the initial segment $Y_\prec(\lambda)$ is open in the metric topology of (Y, d). Since (Y, d) is perfect, $Y_\prec(\lambda)$ must be infinite, contradiction. This completes the *(ii)⇒(iii)* direction. □

D LOG-COVERING DIMENSION: THE EARTHMOVER DISTANCE EXAMPLE

We flesh out the example from Section 1.5. Fix a metric space $(X, \mathcal{D})$ of finite diameter and covering dimension $\kappa < \infty$. Let $\mathcal{P}_X$ denote the set of all probability measures over X. Let $(\mathcal{P}_X, W_1)$ be the space of all probability measures over $(X, \mathcal{D})$ under the Wasserstein W_1 metric, a.k.a., the Earthmover distance:

$$W_1(\nu, \nu') = \inf \mathbb{E}\left[\|Y - Y'\|_2\right],$$

where the infimum is taken over all joint distributions (Y, Y') on $X \times X$ with marginals ν and ν', respectively (for any two $\nu, \nu' \in \mathcal{P}_X$).

THEOREM D.1. *The log-covering dimension of* $(\mathcal{P}_X, W_1)$ *is* κ.

[39] "Transfinite recursion" is a theorem in set theory that asserts that to define a function F on ordinals, it suffices to specify, for each ordinal λ, how to determine $F(\lambda)$ from $F(\nu)$, $\nu < \lambda$.

For the sake of completeness: for any $\mu, \mu' \in \mathcal{P}_X$, the Wasserstein W_1 metric, a.k.a., the Earth-mover distance, is defined as $W_1(\nu, \nu') = \inf \mathbb{E}[\, \mathcal{D}(Y, Y')\,]$, where the infimum is taken over all joint distributions (Y, Y') on $X \times X$ with marginals ν and ν', respectively.

In the remainder of this subsection, we prove Theorem D.1.

(Theorem D.1: Upper Bound). Let us cover $(\mathcal{P}_X, W_1)$ with balls of radius $\frac{2}{k}$ for some $k \in \mathbb{N}$. Let S be a $\frac{1}{k}$-net in (X, d); note that $|S| = O(k^{\kappa})$ for a sufficiently large k. Let P be the set of all probability distributions p on (X, d) such that $\mathtt{support}(p) \subset S$ and for every point $x \in S$, $p(x)$ is a rational number with denominator k^{d+1}. The cardinality of P is bounded above by $(k^{\kappa+1})^{k^{\kappa}}$. It remains to show that balls of radius $\frac{2}{k}$ centered at the points of P cover the entire space $(\mathcal{P}_X, W_1)$. This is true because:

- every distribution q is $\frac{1}{k}$-close to a distribution p with support contained in S (let p be the distribution defined by randomly sampling a point of (X, d) from q and then outputting the closest point of S);
- every distribution with support contained in S is $\frac{1}{k}$-close to a distribution in P (round all probabilities down to the nearest multiple of $k^{-(\kappa+1)}$; this requires moving only $\frac{1}{k}$ units of stuff). □

To prove the lower bound, we make a connection to the Hamming metric.

Lemma D.2. *Let (X, d) be any metric space, and let H denote the Hamming metric on the Boolean cube $\{0, 1\}^n$. If $S \subseteq X$ is a subset of even cardinality $2n$ and ϵ is a lower bound on the distance between any two points of S, then there is a mapping $f : \{0, 1\}^n \to \mathcal{P}_X$ such that for all $a, b \in \{0, 1\}^n$,*

$$W_1(f(a), f(b)) \geq \frac{\epsilon}{n}\, H(a, b). \tag{50}$$

Proof. Group the points of S arbitrarily into pairs $S_i = \{x_i, y_i\}$, where $i = 1, \ldots, n$. For $a \in \{0, 1\}^n$ and $1 \leq i \leq n$, define $t_i(a) = x_i$ if $a_i = 0$, and $t_i(a) = y_i$ otherwise. Let $f(a)$ be the uniform distribution on the set $\{t_1(a), \ldots, t_n(a)\}$. To prove Equation (50), note that if i is any index such that $a_i \neq b_i$, then $f(a)$ assigns probability $\frac{1}{n}$ to $t_i(a)$ while $f(b)$ assigns zero probability to the entire ball of radius ϵ centered at $t_i(a)$. Consequently, the $\frac{1}{n}$ units of probability at $t_i(a)$ have to move a distance of at least ϵ when shifting from distribution $f(a)$ to $f(b)$. Summing over all indices i such that $a_i \neq b_i$, we obtain Equation (50). □

The following lemma, asserting the existence of asymptotically good binary error-correcting codes, is well known, e.g., see References [48, 99].

Lemma D.3. *Suppose δ, ρ are constants satisfying $0 < \delta < \frac{1}{2}$ and $0 \leq \rho < 1 + \delta \log_2(\delta) + (1 - \delta)\log_2(1 - \delta)$. For every sufficiently large n, the Hamming cube $\{0, 1\}^n$ contains more than $2^{\rho n}$ points, no two of which are nearer than distance δn in the Hamming metric.*

Combining these two lemmas, we obtain an easy proof for the lower bound in Theorem D.1.

(Theorem D.1: Lower Bound). Consider any $\gamma < \kappa$. The hypothesis on the covering dimension of (X, d) implies that for all sufficiently small ϵ, there exists a set S of cardinality $2n$—for some $n > \epsilon^{-\gamma}$—such that the minimum distance between two points of S is at least 5ϵ. Now let C be a subset of $\{0, 1\}^n$ having at least $2^{n/5}$ elements, such that the Hamming distance between any two points of C is at least $n/5$. Lemma D.3 implies that such a set C exists, and we can then apply Lemma D.2 to embed C in $\mathcal{P}_X$, obtaining a subset of $\mathcal{P}_X$ whose cardinality is at least $2^{\epsilon^{-\gamma}/5}$, with distance at least ϵ between every pair of points in the set. Thus, any ϵ-covering of $\mathcal{P}_X$ must contain at least $2^{\epsilon^{-\gamma}/5}$ sets, implying that $\mathtt{LCD}(\mathcal{P}_X, W_1) \geq \gamma$. As γ was an arbitrary number less than κ, the proposition is proved. □

REFERENCES

[1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. 2011. Improved algorithms for linear stochastic bandits. In *Proceedings of the 25th Advances in Neural Information Processing Systems (NIPS'11).* 2312–2320.

[2] Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. 2008. Competing in the Dark: An Efficient Algorithm for Bandit Linear Optimization. In *Proceedings of the 21st Conference on Learning Theory (COLT'08).* 263–274.

[3] Ittai Abraham and Dahlia Malkhi. 2005. Name independent routing for growth bounded networks. In *Proceedings of the 17th ACM Symposium on Parallel Algorithms and Architectures (SPAA).* 49–55.

[4] Rajeev Agrawal. 1995. The continuum-armed bandit problem. *SIAM J. Control Optimiz.* 33, 6 (1995), 1926–1951.

[5] Shipra Agrawal, Vashist Avadhanula, Vineet Goyal, and Assaf Zeevi. 2016. A near-optimal exploration-exploitation approach for assortment selection. In *Proceedings of the 17th ACM Conference on Economics and Computation (ACM EC'16).* 599–600.

[6] Shipra Agrawal and Nikhil R. Devanur. 2014. Bandits with concave rewards and convex knapsacks. In *Proceedings of the 15th ACM Conference on Economics and Computation (ACM EC'14).*

[7] Kareem Amin, Michael Kearns, and Umar Syed. 2011. Bandits, query learning, and the haystack dimension. In *Proceedings of the 24th Conference on Learning Theory (COLT'11).*

[8] J.-Y. Audibert and S. Bubeck. 2010. Regret bounds and minimax policies under partial monitoring. *J. Mach. Learn. Res.* 11 (2010), 2785–2836.

[9] J.-Y. Audibert, R. Munos, and Cs. Szepesvári. 2009. Exploration-exploitation trade-off using variance estimates in multi-armed bandits. *Theoret. Comput. Sci.* 410 (2009), 1876–1902.

[10] Peter Auer. 2002. Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.* 3 (2002), 397–422.

[11] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. 2002. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* 47, 2–3 (2002), 235–256.

[12] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. 2002. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.* 32, 1 (2002), 48–77.

[13] Peter Auer and Ronald Ortner. 2010. UCB revisited: Improved regret bounds for the stochastic multi-armed bandit problem. *Periodica Mathematica Hungarica* 61 (2010), 55–65.

[14] Peter Auer, Ronald Ortner, and Csaba Szepesvári. 2007. Improved rates for the stochastic continuum-armed bandit problem. In *Proceedings of the 20th Conference on Learning Theory (COLT'07).* 454–468.

[15] Baruch Awerbuch and Robert Kleinberg. 2008. Online linear optimization and adaptive routing. *J. Comput. Syst. Sci.* 74, 1 (Feb. 2008), 97–114.

[16] Mohammad Gheshlaghi Azar, Alessandro Lazaric, and Emma Brunskill. 2014. Online stochastic optimization under correlated bandit feedback. In *Proceedings of the 31st International Conference on Machine Learning (ICML'14).* 1557–1565.

[17] Moshe Babaioff, Shaddin Dughmi, Robert D. Kleinberg, and Aleksandrs Slivkins. 2015. Dynamic pricing with limited supply. *ACM Trans. Econ. Comput.* 3, 1 (2015), 4.

[18] Moshe Babaioff, Robert Kleinberg, and Aleksandrs Slivkins. 2015. Truthful mechanisms with implicit payment computation. *J. ACM* 62, 2 (2015), 10.

[19] Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. 2014. Characterizing truthful multi-armed bandit mechanisms. *SIAM J. Comput.* 43, 1 (2014), 194–230.

[20] Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. 2018. Bandits with knapsacks. *J. ACM* 65, 3 (2018).

[21] Dirk Bergemann and Juuso Välimäki. 2006. Bandit problems. In *The New Palgrave Dictionary of Economics*, 2nd ed., Steven Durlauf and Larry Blume (Eds.). Macmillan Press.

[22] Donald Berry and Bert Fristedt. 1985. *Bandit Problems: Sequential Allocation of Experiments.* Chapman & Hall.

[23] Donald A. Berry, Robert W. Chen, Alan Zame, David C. Heath, and Larry A. Shepp. 1997. Bandit problems with infinitely many arms. *Ann. Stat.* 25, 5 (1997), 2103–2116.

[24] Omar Besbes and Assaf Zeevi. 2009. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operat. Res.* 57, 6 (2009), 1407–1420.

[25] Avrim Blum. 1997. Empirical support for winnow and weighted-majority-based algorithms: Results on a calendar scheduling domain. *Mach. Learn.* 26 (1997), 5–23.

[26] Avrim Blum, Vijay Kumar, Atri Rudra, and Felix Wu. 2003. Online learning in online auctions. In *Proceedings of the 14th ACM-SIAM Symposium on Discrete Algorithms (SODA'03).* 202–204.

[27] Sébastien Bubeck and Nicolo Cesa-Bianchi. 2012. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Found. Trends Mach. Learn.* 5, 1 (2012).

[28] Sébastien Bubeck and Rémi Munos. 2010. Open loop optimistic planning. In *Proceedings of the 23rd Conference on Learning Theory (COLT'10).* 477–489.

[29] Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. 2008. Online optimization in X-armed bandits. In *Proceedings of the 21st Advances in Neural Information Processing Systems (NIPS'08)*. 201–208.

[30] Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvari. 2011. Online optimization in X-armed bandits. *J. Mach. Learn. Res.* 12 (2011), 1587–1627.

[31] Sébastien Bubeck, Gilles Stoltz, and Jia Yuan Yu. 2011. Lipschitz bandits without the Lipschitz constant. In *Proceedings of the 22nd International Conference on Algorithmic Learning Theory (ALT'11)*. 144–158.

[32] Adam Bull. 2015. Adaptive-treed bandits. *Bernoulli J. Stat.* 21, 4 (2015), 2289–2307.

[33] G. Cantor. 1883. Über unendliche, lineare Punktmannichfaltigkeiten, 4. *Math. Ann.* 21 (1883), 51–58.

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

[35] Nicolò Cesa-Bianchi and Gábor Lugosi. 2006. *Prediction, Learning, and Games.* Cambridge University Press.

[36] Hubert T.-H. Chan, Anupam Gupta, Bruce M. Maggs, and Shuheng Zhou. 2005. On hierarchical routing in bounded-growth metrics. In *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms (SODA'05)*. 762–771.

[37] Richard Cole and Lee-Ad Gottlieb. 2006. Searching dynamic point sets in spaces with bounded doubling dimension. In *Proceedings of the 38th ACM Symposium on Theory of Computing (STOC'06)*. 574–583.

[38] Eric Cope. 2009. Regret and convergence bounds for immediate-reward reinforcement learning with continuous action spaces. *IEEE Trans. Auto. Control* 54, 6 (2009), 1243–1253.

[39] Thomas M. Cover and Joy A. Thomas. 1991. *Elements of Information Theory.* John Wiley & Sons, New York.

[40] Varsha Dani, Thomas P. Hayes, and Sham Kakade. 2007. The price of bandit information for online optimization. In *Proceedings of the 20th Advances in Neural Information Processing Systems (NIPS'07).*

[41] Varsha Dani, Thomas P. Hayes, and Sham Kakade. 2008. Stochastic linear optimization under bandit feedback. In *Proceedings of the 21st Conference on Learning Theory (COLT'08)*. 355–366.

[42] Thomas Desautels, Andreas Krause, and Joel Burdick. 2012. Parallelizing exploration-exploitation tradeoffs with Gaussian process bandit optimization. In *Proceedings of the 29th International Conference on Machine Learning (ICML'12).*

[43] Nikhil Devanur and Sham M. Kakade. 2009. The price of truthfulness for pay-per-click auctions. In *Proceedings of the 10th ACM Conference on Electronic Commerce (EC'09)*. 99–106.

[44] Abraham Flaxman, Adam Kalai, and H. Brendan McMahan. 2005. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms (SODA'05)*. 385–394.

[45] Christodoulos A. Floudas. 1999. *Deterministic Global Optimization: Theory, Algorithms and Applications.* Kluwer Academic Publishers.

[46] Yoav Freund, Robert E. Schapire, Yoram Singer, and Manfred K. Warmuth. 1997. Using and combining predictors that specialize. In *Proceedings of the 29th ACM Symposium on Theory of Computing (STOC'97)*. 334–343.

[47] Aurélien Garivier and Olivier Cappé. 2011. The KL-UCB algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th Conference on Learning Theory (COLT'11).*

[48] E. N. Gilbert. 1952. A comparison of signalling alphabets. *Bell Syst. Tech. J.* 31 (May 1952), 504–522.

[49] John Gittins, Kevin Glazebrook, and Richard Weber. 2011. *Multi-Armed Bandit Allocation Indices.* John Wiley & Sons.

[50] Anupam Gupta, Mike Dinitz, and Kanat Tangwongsan. 2007. *Private communication.*

[51] Anupam Gupta, Robert Krauthgamer, and James R. Lee. 2003. Bounded geometries, fractals, and low–distortion embeddings. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science (FOCS'03)*. 534–543.

[52] Elad Hazan and Satyen Kale. 2011. Better algorithms for benign bandits. *J. Mach. Learn. Res.* 12 (2011), 1287–1311.

[53] Elad Hazan and Nimrod Megiddo. 2007. Online learning with prior information. In *Proceedings of the 20th Conference on Learning Theory (COLT'07)*. 499–513.

[54] J. Heinonen. 2001. *Lectures on Analysis on Metric Spaces.* Springer-Verlag, New York.

[55] Kirsten Hildrum, John Kubiatowicz, and Satish Rao. 2004. Object location in realistic networks. In *Proceedings of the 16th ACM Symposium on Parallel Algorithms and Architectures (SPAA'04)*. 25–35.

[56] Chien-Ju Ho, Aleksandrs Slivkins, and Jennifer Wortman Vaughan. 2016. Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. *J. Artific. Intell. Res.* 55 (2016), 317–359.

[57] Junya Honda and Akimichi Takemura. 2010. An asymptotically optimal bandit algorithm for bounded support models. In *Proceedings of the 23rd Conference on Learning Theory (COLT'10).*

[58] D. R. Karger and M. Ruhl. 2002. Finding nearest neighbors in growth-restricted metrics. In *Proceedings of the 34th ACM Symposium on Theory of Computing (STOC'02)*. 63–66.

[59] Jon Kleinberg, Aleksandrs Slivkins, and Tom Wexler. 2009. Triangulation and embedding using small sets of beacons. *J. ACM* 56, 6 (Sept. 2009).

[60] Robert Kleinberg. 2004. Nearly tight bounds for the continuum-armed bandit problem. In *Proceedings of the 18th Advances in Neural Information Processing Systems (NIPS'04).*

[61] Robert Kleinberg. 2005. *Online Decision Problems with Large Strategy Sets*. Ph.D. Dissertation. MIT.

[62] Robert Kleinberg, Alexandru Niculescu-Mizil, and Yogeshwer Sharma. 2008. Regret bounds for sleeping experts and bandits. In *Proceedings of the 21st Conference on Learning Theory (COLT'08)*. 425–436.

[63] Robert Kleinberg and Aleksandrs Slivkins. 2010. Sharp dichotomies for regret minimization in metric spaces. In *Proceedings of the 21st ACM-SIAM Symposium on Discrete Algorithms (SODA'10)*.

[64] Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. 2008. Multi-armed bandits in metric spaces. In *Proceedings of the 40th ACM Symposium on Theory of Computing (STOC'08)*. 681–690.

[65] Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. 2008. Multi-Armed Bandits in Metric Spaces. *Technical report*. Retrieved from http://arxiv.org/abs/0809.4882.

[66] Robert D. Kleinberg and Frank T. Leighton. 2003. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS'03)*.

[67] Levente Kocsis and Csaba Szepesvari. 2006. Bandit-based Monte-Carlo planning. In *Proceedings of the 17th European Conference on Machine Learning (ECML'06)*. 282–293.

[68] Andreas Krause and Cheng Soon Ong. 2011. Contextual Gaussian process bandit optimization. In *Proceedings of the 25th Advances in Neural Information Processing Systems (NIPS'11)*. 2447–2455.

[69] Tze Leung Lai and Herbert Robbins. 1985. Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math.* 6 (1985), 4–22.

[70] Tyler Lu, Dávid Pál, and Martin Pál. 2010. Showing relevant ads via Lipschitz context multi-armed bandits. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS'10)*.

[71] Stefan Magureanu, Richard Combes, and Alexandre Proutiere. 2014. Lipschitz bandits: Regret lower bound and optimal algorithms. In *Proceedings of the 27th Conference on Learning Theory (COLT'14)*. 975–999.

[72] Odalric-Ambrym Maillard and Rémi Munos. 2010. Online learning in adversarial Lipschitz environments. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD'10)*. 305–320.

[73] Odalric-Ambrym Maillard and Rémi Munos. 2011. Adaptive bandits: Towards the best history-dependent strategy. In *Proceedings of the 24th Conference on Learning Theory (COLT'11)*.

[74] S. Mazurkiewicz and W. Sierpinski. 1920. Contribution à la topologie des ensembles dénombrables. *Fund. Math.* 1 (1920), 17–27.

[75] Manor Mendel and Sariel Har-Peled. 2005. Fast construction of nets in low dimensional metrics, and their applications. In *Proceedings of the 21st ACM Symposium on Computational Geometry (SoCG'05)*. 150–158.

[76] Stanislav Minsker. 2013. Estimation of extreme values and associated level sets of a regression function via selective sampling. In *Proceedings of the 26th Conference on Learning Theory (COLT'13)*. 105–121.

[77] Michael Mitzenmacher and Eli Upfal. 2005. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press.

[78] Rémi Munos. 2011. Optimistic optimization of a deterministic function without the knowledge of its smoothness. In *Proceedings of the 25th Conference on Advances in Neural Information Processing Systems (NIPS'11)*. 783–791.

[79] Rémi Munos. 2014. From bandits to Monte-Carlo tree search: The optimistic principle applied to optimization and planning. *Found. Trends Mach. Learn.* 7, 1 (2014), 1–129.

[80] Rémi Munos and Pierre-Arnaud Coquelin. 2007. Bandit algorithms for tree search. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI'07)*.

[81] Sandeep Pandey, Deepak Agarwal, Deepayan Chakrabarti, and Vanja Josifovski. 2007. Bandits for taxonomies: A model-based approach. In *Proceedings of the SIAM International Conference on Data Mining (SDM'07)*.

[82] Sandeep Pandey, Deepayan Chakrabarti, and Deepak Agarwal. 2007. Multi-armed bandit problems with dependent arms. In *Proceedings of the 24th International Conference on Machine Learning (ICML'07)*.

[83] Filip Radlinski, Robert Kleinberg, and Thorsten Joachims. 2008. Learning diverse rankings with multi-armed bandits. In *Proceedings of the 25th International Conference on Machine Learning (ICML'08)*. 784–791.

[84] Herbert Robbins. 1952. Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* 58 (1952), 527–535.

[85] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. 2000. A metric for distributions with applications to image databases. *Int. J. Comput. Vision* 40, 2 (2000), 99–121.

[86] Manfred Schroeder. 1991. *Fractal, Chaos and Power Laws: Minutes from an Infinite Paradise*. W. H. Freeman and Co.

[87] Shai Shalev-Shwartz and Shai Ben-David. 2014. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.

[88] Aleksandrs Slivkins. 2007. Distance estimation and object location via rings of neighbors. *Distributed Computing* 19, 4 (Mar. 2007), 313–333.

[89] Aleksandrs Slivkins. 2007. Towards fast decentralized construction of locality-aware overlay networks. In *Proceedings of the 26th Annual ACM Symposium on Principles of Distributed Computing (PODC'07)*. 89–98.

[90] Aleksandrs Slivkins. 2011. Multi-armed bandits on implicit metric spaces. In *Proceedings of the 25th Advances in Neural Information Processing Systems (NIPS'11)*.

[91] Aleksandrs Slivkins. 2014. Contextual bandits with similarity information. *J. Mach. Learn. Res.* 15, 1 (2014), 2533–2568.

[92] Aleksandrs Slivkins, Filip Radlinski, and Sreenivas Gollapudi. 2013. Ranked bandits in metric spaces: Learning optimally diverse rankings over large document collections. *J. Mach. Learn. Res.* 14 (Feb. 2013), 399–436.

[93] Aleksandrs Slivkins and Eli Upfal. 2008. Adapting to a changing environment: the Brownian restless bandits. In *Proceedings of the 21st Conference on Learning Theory (COLT'08)*. 343–354.

[94] Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. 2010. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning (ICML'10)*. 1015–1022.

[95] Michel Talagrand. 2005. *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*. Springer.

[96] Kunal Talwar. 2004. Bypassing the embedding: Algorithms for low-dimensional metrics. In *Proceedings of the 36th ACM Symposium on Theory of Computing (STOC'04)*. 281–290.

[97] William R. Thompson. 1933. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25, 3–4 (1933), 285–294.

[98] Michal Valko, Alexandra Carpentier, and Rémi Munos. 2013. Stochastic simultaneous optimistic optimization. In *Proceedings of the 30th International Conference on Machine Learning (ICML'13)*. 19–27.

[99] R. R. Varshamov. 1957. Estimate of the number of signals in error correcting codes. *Doklady Akadamii Nauk* 177 (1957), 739–741.

[100] V. Vovk. 1998. A game of prediction with expert advice. *J. Comput. Syst. Sci.* 56, 2 (1998), 153–173.

[101] Yizao Wang, Jean-Yves Audibert, and Rémi Munos. 2008. Algorithms for infinitely many-armed bandits. In *Advances in Neural Information Processing Systems*. MIT Press, 1729–1736.

[102] Zizhuo Wang, Shiming Deng, and Yinyu Ye. 2014. Close the gaps: A learning-while-doing algorithm for single-product revenue management problems. *Operat. Res.* 62, 2 (2014), 318–331.

Received November 2015; revised May 2018; accepted November 2018