Batched Stochastic Bandit for Nondegenerate Functions

Yu Liu*,† Yunlu Shu*,‡ Tianyu Wang§

Abstract

This paper studies batched bandit learning problems for nondegenerate functions. Over a compact doubling metric space $(\mathcal{X}, \mathcal{D})$, a function $f : \mathcal{X} \to \mathbb{R}$ is called nondegenerate if there exists $L \geq \lambda > 0$ and $q \geq 1$, such that

$$\lambda \left(\mathcal{D}(\mathbf{x}, \mathbf{x}^*)\right)^q \leq f(\mathbf{x}) - f(\mathbf{x}^*) \leq L \left(\mathcal{D}(\mathbf{x}, \mathbf{x}^*)\right)^q, \quad \forall \mathbf{x} \in \mathcal{X},$$

where $\mathbf{x}^* = \arg\min_{\mathbf{z} \in \mathcal{X}} f(\mathbf{z})$ is the unique minimizer of f over $\mathcal{X}$. In this paper, we introduce an algorithm that solves the batched bandit problem for nondegenerate functions near-optimally. More specifically, we introduce an algorithm, called Geometric Narrowing (GN), whose regret bound is of order $\widetilde{\mathcal{O}}\left(A_+^d \sqrt{T}\right)$, where d is the doubling dimension of $(\mathcal{X}, \mathcal{D})$, and A_+ is a constant independent of d and the time horizon T. In addition, GN only needs $\mathcal{O}(\log \log T)$ batches to achieve this regret. We also provide lower bound analysis for this problem. More specifically, we prove that over some (compact) doubling metric space of doubling dimension d: 1. For any policy π, there exists a problem instance on which π admits a regret of order $\Omega\left(A_-^d \sqrt{T}\right)$, where A_- is a constant independent of d and T; 2. No policy can achieve a regret of order $A_-^d \sqrt{T}$ over all problem instances, using less than $\Omega\left(\log \log T\right)$ rounds of communications. Our lower bound analysis shows that the GN algorithm achieves near optimal regret with minimal number of batches.

1 Introduction

In batched stochastic bandit, an agent collects noisy rewards/losses in batches, and aims to find the best option while exploring the space (Thompson, 1933; Robbins, 1952; Gittins, 1979; Lai and Robbins, 1985; Auer et al., 2002a,b; Perchet et al., 2016; Gao et al., 2019). This setting reflects the key attributes of crucial real-world applications. For example, in experimental design (Robbins, 1952; Berry and Fristedt, 1985), the observations are often noisy and collected in batches (Perchet et al., 2016; Gao et al., 2019). In this paper, we consider batched stochastic bandits for an important class of functions, called "nondegenerate functions".

1.1 Nondegenerate Functions

Over a compact doubling metric space $(\mathcal{X}, \mathcal{D})$, a function $f : \mathcal{X} \to \mathbb{R}$ is called a nondegenerate function if there exists $L \geq \lambda > 0$ and $q \geq 1$, such that

$$\lambda \left(\mathcal{D}(\mathbf{x}, \mathbf{x}^*)\right)^q \leq f(\mathbf{x}) - f(\mathbf{x}^*) \leq L \left(\mathcal{D}(\mathbf{x}, \mathbf{x}^*)\right)^q, \quad \forall \mathbf{x} \in \mathcal{X}, \tag{1}$$

where $\mathbf{x}^* = \arg\min_{\mathbf{z} \in \mathcal{X}} f(\mathbf{z})$ is the unique minimizer of f over $\mathcal{X}$. Nondegenerate functions (Valko et al., 2013; Zhang et al., 2017; Gemp et al., 2024) hold significance as they encompass various important problems. Below we list some important nondegenerate functions.

*Equal contribution
†22110840006@m.fudan.edu.cn
‡22110840008@m.fudan.edu.cn
§wangtianyu@fudan.edu.cn

- **(P0, warm-up example) Revenue curve as a function of price:** As a warm-up example, let the space $(\mathcal{X}, \mathcal{D})$ be $\mathcal{X} = [0,1]$ and $\mathcal{D}(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|$. If $\mathbf{x} \in [0,1]$ models price, then functions satisfying (1) provide a natural model for revenue curve as a function of price, up to a flip of sign (e.g., Chen and Wang, 2023; Perakis and Singhvi, 2023, and references therein).

- **(P) Nonsmooth nonconvex objective over Riemannian manifolds:** Our study introduces a *global* bandit optimization method for a class of nonconvex functions on compact Riemannian manifolds, where nontrivial convex functions do not exist (Yau, 1974). Let $(\mathcal{X}, \mathcal{D})$ be a compact finite-dimensional Riemannian manifold with the metric defined by the geodesic distance (e.g., Petersen, 2006). Then a smooth function with nondegenerate Taylor approximation satisfies (1) near its global minimum $\mathbf{x}^*$. More specifically, we can Taylor approximate the function f near $\mathbf{x}^*$ and get, for $\mathbf{x} = \mathrm{Exp}_{\mathbf{x}^*}(\mathbf{v})$ with some $\mathbf{v} \in T_{\mathbf{x}^*}\mathcal{M}$, $f(\mathbf{x}) \approx \sum_{i=0}^{K} \frac{1}{i!} \varphi_{\mathbf{v}}^{(i)}(\|\mathbf{v}\|)$ where $\varphi_{\mathbf{v}}^{(i)}$ is the i-th derivative of $f \circ \mathrm{Exp}_{\mathbf{x}^*}$ along the direction of $\mathbf{v}$, and $K \geq 2$ is some integer. Since $\mathbf{x}^*$ is a local minimum of f, we have $\varphi_{\mathbf{v}}^{(1)} = 0$ (for all $\mathbf{v}$), and thus $f(\mathbf{x}) - f(\mathbf{x}^*) \approx \frac{1}{q!} \varphi_{\mathbf{v}}^{(q)}(\|\mathbf{v}\|)$ where $q \geq 2$ is the smallest integer such that $\varphi_{\mathbf{v}}^{(q)} \neq 0$ (for some $\mathbf{v}$). If $\varphi_{\mathbf{v}}^{(q)}$ is nontrivial for all $\mathbf{v} \in T_{\mathbf{x}^*}\mathcal{M}$, that is, the leading nontrivial total derivative of f is nondegenerate, then the function f satisfies (1) in a neighborhood of $\mathbf{x}^*$. This justifies the name "nondegenerate". In Figure 2, we provide a specific example of a nondegenerate function over a Riemannian manifold. Over the entire manifold, the objective is nonsmooth and nonconvex.

Also, it is worth emphasizing that nondegenerate functions can possess nonconvexity, nonsmoothness, or discontinuity. As an illustration, consider the following nondegenerate function $f(\mathbf{x})$ defined over the interval $[-2, 2]$, which exhibits discontinuity:

$$f(\mathbf{x}) = \begin{cases} -\mathbf{x}, & \text{if } \mathbf{x} \in [-2, -1) \\ \mathbf{x}^2, & \text{if } \mathbf{x} \in [-1, 1] \\ \mathbf{x} + 1, & \text{if } \mathbf{x} \in (1, 2]. \end{cases} \tag{2}$$

A plot of the function (2) is in Figure 1, $\frac{\mathbf{x}^2}{2}$ (resp. $2\mathbf{x}^2$) is a lower bound (resp. upper bound) for $f(\mathbf{x})$ over $[-2, 2]$. More generally, over a compact Riemannian manifold, with the metric $\mathcal{D}$ defined by the geodesic distance, nondegenerate functions can still possess nonconvexity, nonsmoothness, or discontinuity. A specific example is shown in Figure 2.

Given the aforementioned motivating examples, developing an efficient stochastic bandit/optimization algorithm for nondegenerate functions is of great importance. In addition, we focus our study on the batched feedback setting, which is also important.

1.2 The Batched Bandit Setting

In bandit learning, more specifically stochastic bandit learning, the agent is tasked with sequentially making decisions based on noisy loss/reward samples associated with these decisions. The objective of the agent is to identify the optimal choice while simultaneously learning the expected loss function across the decision space. The effectiveness of the agent's policy is evaluated through regret, which quantifies the difference in loss between the agent's chosen decision and the optimal decision, accumulated over time. More formally, the T-step regret of a policy π is defined as

$$R^\pi(T) := \sum_{t=1}^{T} f(\mathbf{x}_t) - f(\mathbf{x}^*), \tag{3}$$

where $\mathbf{x}_t \in \mathcal{X}$ is the choice of policy π at step t, f is the expected loss function, and $\mathbf{x}^*$ is the optimal choice. Typically, the goal of a bandit algorithm is to achieve a regret rate that grows as slow as possible.

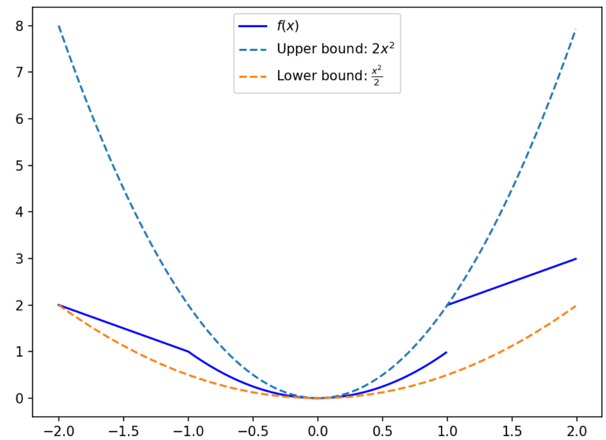


Figure 1: Plot of $f(\mathbf{x})$ defined in (2). $\frac{\mathbf{x}^2}{2}$ (resp. $2\mathbf{x}^2$) is a lower bound (resp. upper bound) for $f(\mathbf{x})$ over $[-2, 2]$. This plot shows that a nondegenerate function can be nonconvex, nonsmooth or discountinuous.

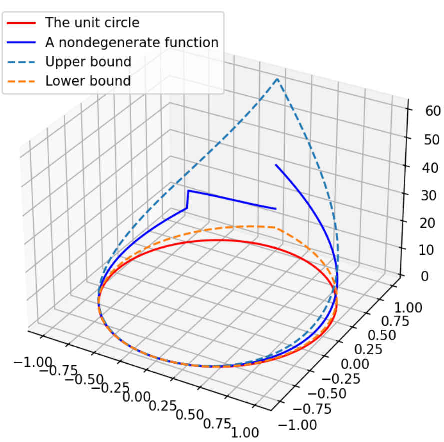


Figure 2: Plot of a nondegenerate function f defined over the unit circle $\mathbb{S}^1$, and the metric is the arc length along the circle. This function is not convex and not continuous, but satisfies the nondegenerate condition.

Remark 1. *For the rest of the paper, we will use a loss minimization formulation for the bandit learning problem. With a flip of sign, we can easily phrase the problem in a reward-maximization language.*

In the context of batched bandit learning, the primary objective remains to be minimizing the growth of regret. However, in this setting, the agent is unable to observe the loss sample immediately after making her decision. Instead, she needs to wait until a *communication point* to collect the loss samples in batches. To elaborate further, in batch bandit problems, the agent in a T-step game dynamically selects a sequence of communication points denoted as $\mathcal{T} = \{t_0, \cdots, t_M\}$, where $0 = t_0 < t_1 < \cdots < t_M = T$ and $M \ll T$. In this setting, loss observations are only communicated to the player at $t_1, \cdots, t_M$. Consequently, for any given time t within the j-th batch ($t_{j-1} < t \leq t_j$), the reward y_t remains unobserved until time t_j. The reward samples are corrupted by mean-zero, *iid*, 1-sub-Gaussian noise. The decision made at time t is solely influenced by the losses received up to time

t_{j-1}. The selection of the communication points $\mathcal{T}$ is adaptive, meaning that the player determines each point $t_j \in \mathcal{T}$ based on the previous operations and observations up to t_{j-1}.

In batched bandit setting, the agent not only aims to minimize regret, but also seeks to minimize the number of communications points required.

For simplicity, we use *nondegenerate bandits* to refer to stochastic bandit problem with nondegenerate loss, and *batched nondegenerate bandits* to refer to batched stochastic bandit problem with nondegenerate loss.

1.3 Our Results

In this paper, we introduce an algorithm, called Geometric Narrowing (GN), that solves batched bandit learning problems for nondegenerate functions in a near-optimal way. The GN algorithm operates by successively narrowing the search space, and satisfies the properties stated in Theorem 1.

Theorem 1. *Let $(\mathcal{X}, \mathcal{D})$ be a compact doubling metric space, and let f be a nondegenerate function defined over $(\mathcal{X}, \mathcal{D})$. Consider a stochastic bandit learning environment where all loss samples are corrupted by iid sub-Gaussian mean-zero noise. For any $T \in \mathbb{N}_+$, with probability exceeding $1 - 2T^{-1}$, the T-step total regret of Geometric Narrowing, written $R^{GN}(T)$, satisfies*

$$R^{GN}(T) \leq K_+ A_+^d \sqrt{T \log T} \log\log \frac{T}{\log T},$$

where d is the doubling dimension of $(\mathcal{X}, \mathcal{D})$, and K_+ and A_+ are constants independent of d and T. In addition, Geometric Narrowing only needs $\mathcal{O}(\log\log T)$ communication points to achieve this regret rate.

As a corollary of Theorem 1, we prove that the simple regret of GN is of order $\mathcal{O}\left(\sqrt{\frac{\log T}{T}} \log\log T\right)$. This result is summarized in Corollary 1.

Corollary 1. *Let $(\mathcal{X}, \mathcal{D})$ be a compact doubling metric space. Let f be a nondegenerate function defined over $(\mathcal{X}, \mathcal{D})$. For any $T \in \mathbb{N}_+$, with probability exceeding $1 - 2T^{-1}$, the GN algorithm finds a point $\mathbf{x}_{out}$ such that $f(\mathbf{x}_{out}) - f(\mathbf{x}^*) \leq \mathcal{O}\left(\sqrt{\frac{\log T}{T}} \log\log T\right)$. In addition, Geometric Narrowing only needs $\mathcal{O}(\log\log T)$ communication points to achieve this rate.*

In addition, we prove that it is hard to outperform GN by establishing lower bound results in Theorems 2, 3 and Corollary 2. Theorem 2 states that no algorithm can uniformly perform better than $\Omega\left(A_-^d \sqrt{T}\right)$ for some A_- independent of d and T.

Theorem 2. *For any $d \geq 1$ and $T \in \mathbb{N}_+$, there exists a compact doubling metric space $(\mathcal{X}_0, \mathcal{D}_0)$ that simultaneously satisfies the following: 1. The doubling dimension of $(\mathcal{X}_0, \mathcal{D}_0)$ is $\lfloor d \rfloor$. 2. For any policy π, there exists a problem instance I defined over $(\mathcal{X}_0, \mathcal{D}_0)$, such that the regret of running π on I satisfies*

$$\mathbb{E}\left[R^\pi(T)\right] \geq K_- A_-^{\lfloor d \rfloor} \sqrt{T}$$

where $\mathbb{E}$ is the expectation whose probability law is induced by running π (for T steps) on the instance I, and K_- and A_- are numbers that do not depend on d or T.

Theorem 2 implies that the regret bound for GN is near-optimal. Also, we provide a lower bound analysis for the communication lower bound of batched bandit for nondegenerate functions. This result is stated below in Theorem 3.

Theorem 3. *Let $M \in \mathbb{N}_+$ be the total rounds of communications allowed. For any $d \geq 1$ and $T \in \mathbb{N}_+$ $(T \geq M)$, there exists a compact doubling metric space $(\mathcal{X}_0, \mathcal{D}_0)$ that simultaneously satisfies the following: 1. the doubling dimension of $(\mathcal{X}_0, \mathcal{D}_0)$ is $\lfloor d \rfloor$, and 2. for any policy π, there exists a problem instance I defined over $(\mathcal{X}_0, \mathcal{D}_0)$, such that the regret of running π on I for T steps satisfies*

$$\mathbb{E}\left[R^{\pi}(T)\right] \geq K_{-} A_{-}^{\lfloor d \rfloor} \cdot \frac{1}{M^2} \cdot T^{\frac{1}{2} \cdot \frac{1}{1-2^{-M}}},$$

where K_{-} and A_{-} are numbers that do not depend on d, M or T.

By setting M to the order of $\log \log T$ in Theorem 3, we have the following corollary.

Corollary 2. *For any $d \geq 1$ and $T \in \mathbb{N}_+$, there exists a compact doubling metric space $(\mathcal{X}_0, \mathcal{D}_0)$ that simultaneously satisfies the following: 1. The doubling dimension of $(\mathcal{X}_0, \mathcal{D}_0)$ is $\lfloor d \rfloor$; 2. If less than $\Omega(\log \log T)$ rounds of communications are allowed, no policy can achieve a regret of order $A_{-}^{\lfloor d \rfloor} \sqrt{T}$ over all nondegenerate bandit instances defined over $(\mathcal{X}_0, \mathcal{D}_0)$, where A_{-} is a number independent of d and T.*

Corollary 2 implies that the communication complexity of the GN algorithm is near-optimal, since no algorithm can improve GN's communication complexity without worsening the regret.

Note: In Theorem 2, Theoerm 3 and Corollary 2, the specific values of K_{-} and A_{-} may be different at each occurrence.

Our results also suggest a curse-of-dimensionality phenomenon, discussed below in Remark 2.

Remark 2 (Curse of dimensionality)**.** *Our lower bounds (Theorems 2 and 3) grow exponentially in the doubling dimension d. Therefore, no algorithm can uniformly improve this dependence on d, resulting in a phenomenon commonly referred to as curse-of-dimensionality.*

1.4 Implications of Our Results

Our results have several important implications. Firstly, our research gives a distinct method for the stochastic convex optimization with bandit feedback(Shamir, 2013), especially for the strongly-convex and smooth function which is a special kind of nondegenerate function. For the warm-up problem discussed previously in **(P0)**, our GN algorithm provides a solution to the online/dynamic pricing problem (without inventory constraints) (e.g., Chen and Wang, 2023; Perakis and Singhvi, 2023, and references therein). More importantly, our results yield intriguing implications on Riemannian optimization, and offer a new perspective on stochastic Riemannian optimization problems.

(I) Implications on stochastic zeroth-order optimization over Riemmanian manifolds: Our GN algorithm provides a solution for optimizing nondegenerate functions over compact finite-dimensional Riemannian manifolds (with or without boundary). Our results imply that, the global optimum of a large class of nonconvex and nonsmooth functions can be efficiently approximated. As stated in Corollary 2, we show that GN finds the global optimum of the objective at rate $\widetilde{\mathcal{O}}\left(\frac{1}{\sqrt{T}}\right)$. To our knowledge, for stochastic optimization problems, this is the first result that guarantees an $\widetilde{\mathcal{O}}\left(\frac{1}{\sqrt{T}}\right)$ convergence to the global optimum for nonconvex nonsmooth optimization over compact finite-dimensional Riemannian manifolds. In addition, only $\mathcal{O}(\log \log T)$ rounds of communication are needed to achieve this rate.

1.5 Challenges and Our Approach

As the first work that focuses on batched bandit learning for nondegenerate functions, we face several challenges throughout the analysis, especially in the lower bound proof. Unlike existing lower bound analyses, the geometry of the underlying space imposes challenging constraints on the problem instance construction. To further illustrate this challenge, we briefly review the lower bound instance construction for Lipschitz bandits (Kleinberg, 2005; Kleinberg et al., 2008; Bubeck et al., 2009, 2011a), and explain why techniques for Lipschitz bandits do not carry through. Figures 3 illustrate some instance constructions, showing an overall picture (left figure), three instances for Lipschitz bandits lower bound (right top figure), and three instances of a "naive attempt" (right bottom figure). We start with the instances for Lipschitz bandits lower bounds (solid blue line in left figure). In such cases, as the "height" decreases with T, unfortunately the nondegenerate parameter λ also decreases with T. Also, using the solid red line (left figure) instances as a "naive attempt" disrupts key properties of Lipschitz bandit instances. Specifically: (1) As shown in the right top figure, except for in $\{S_i\}_{i=1}^3$, the values of f_{Lip}^i are identical. In contrast, for the "naive attempt" $\{f_{naive}^i\}_{i=1}^3$ in the right bottom figure, the function values vary across the domain *a.e.* (2) From an information-theoretic perspective, distinguishing between f_{Lip}^1 and f_{Lip}^2 is as difficult as differentiating f_{Lip}^1 from f_{Lip}^3, regardless of the distance between their optima. Conversely, telling apart f_{naive}^1 from f_{naive}^2 can be harder than distinguishing f_{naive}^1 from f_{naive}^3, if the optima of f_{naive}^1 and f_{naive}^2 are closer than those of f_{naive}^1 and f_{naive}^3.

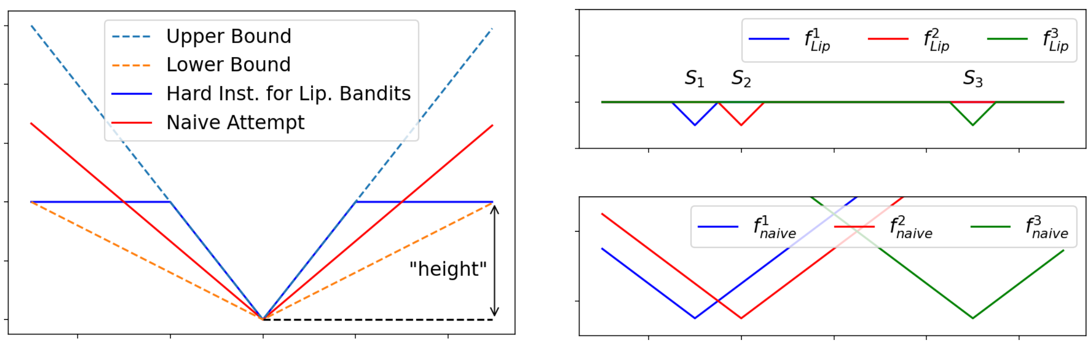


Figure 3: Explanation of the instance for nondegenerate bandits

This implies we cannot change the instances as freely as previously done in the literature, not to mention that all of the lower bound arguments need to take the communication patterns into consideration. To overcome this difficulty, we use a trick called *bitten apple* construction. This trick overcomes the constraints imposed by the nondegenerate property, and is critical in proving a lower bound that scales exponentially with the doubling dimension d. As a result, this trick is critical in justifying the curse-of-dimensionality phenomenon in Remark 2.

For the algorithm design and analysis, we need to carefully utilize the properties of nondegenerate functions to design an algorithm with regret upper bound $\widetilde{\mathcal{O}}(\sqrt{T})$ and communication complexity $\mathcal{O}(\log \log T)$. We need to carefully integrate in the properties of the nondegenerate functions in both the algorithm procedure and the communication pattern. In addition, we design the algorithm in a succinct way, so that the GN algorithm has the following additional advantages.

Proposition 1. *The space complexity of GN does not increase with the time horizon T.*

1.6 Related Works

Compared to many modern machine learning problems, the stochastic Multi-Armed Bandit (MAB) problem has a long history (Thompson, 1933; Robbins, 1952; Gittins, 1979; Lai and Robbins, 1985; Auer et al., 2002a,b). Throughout the years, many solvers for this problem has been invented, including Thompson sampling (Thompson, 1933; Agrawal and Goyal, 2012), the UCB algorithm (Lai and Robbins, 1985; Auer et al., 2002a), exponential weights (Auer et al., 2002b; Arora et al., 2012), and many more; See e.g., (Bubeck and Cesa-Bianchi, 2012; Slivkins, 2019; Lattimore and Szepesvári, 2020) for an exposition.

Throughout the years, multiple variations of the stochastic MAB problems have been intensively investigated, including linear bandits (Auer, 2002; Dani et al., 2007; Chu et al., 2011; Abbasi-Yadkori et al., 2011), Gaussian process bandits (Srinivas et al., 2012; Contal et al., 2014), bandits in metric spaces (Kleinberg, 2005; Kleinberg et al., 2008; Bubeck et al., 2009, 2011a; Podimata and Slivkins, 2021), just to name a few. Among enormous arts on bandit learning, bandits in metric spaces are particularly related to our work. In its early stage, bandits in metric spaces primarily focus on bandit learning over $[0, 1]$ (Agrawal, 1995; Kleinberg, 2005; Auer et al., 2007; Cope, 2009). Afterwards, algorithms for bandits over more general metric spaces were developed (Kleinberg et al., 2008; Bubeck et al., 2009, 2011a,b; Magureanu et al., 2014; Lu et al., 2019; Krishnamurthy et al., 2020; Majzoubi et al., 2020; Feng et al., 2023). In particular, the Zooming bandit algorithm Kleinberg et al. (2008); Slivkins (2014) and the Hierarchical Optimistic Optimization (HOO) algorithm Bubeck et al. (2009, 2011a) were the first algorithms that optimally solve the Lipschitz bandit problem (up to logarithmic factors). Subsequently, Valko et al. (2013) considered an early version of nondegenerate functions, and built its connection to Lipschitz bandits (Kleinberg et al., 2008; Bubeck et al., 2009, 2011a). Valko et al. (2013) proposed StoSOO algorithm for pure exploration of function that is locally smooth with respect to some semi-metric. But, to our knowledge, bandit problems with such functions have not been explored.

In recent years, urged by the rising need for distributed computing and large-scale field experiments (e.g., Berry and Fristedt, 1985; Cesa-Bianchi et al., 2013), the setting of batched feedback has gained attention. Perchet et al. (2016) initiated the study of batched bandit problem, and Gao et al. (2019) settled several important problems in batched multi-armed bandits. Over the last few years, many researchers have contributed to the batched bandit learning problem (Jun et al., 2016; Agarwal et al., 2017; Tao et al., 2019; Han et al., 2020; Karpov et al., 2020; Esfandiari et al., 2021; Ruan et al., 2021; Li and Scarlett, 2022; Agarwal et al., 2022). For example, Han et al. (2020) and Ruan et al. (2021) provide solutions for batched contextual linear bandits. Li and Scarlett (2022) studies batched Gaussian process bandits.

Despite all these works on stochastic bandits and batched stochastic bandits, no existing work focuses on batched bandit learning for nondegenerate functions.

1.6.1 Additional related works from stochastic zeroth-order Riemannian optimization

Since our work has some implications on stochastic zeroth-order Riemannian optimization, we also briefly survey some related works from there; See (Absil et al., 2008; Boumal, 2023) for modern expositions on general Riemannian optimization.

In modern terms, Li et al. (2023a) provided the first oracle complexity analysis for zeroth-order stochastic Riemannian optimization. Afterwards, Li et al. (2023b) introduced a new stochastic zeroth-order algorithm that leverages moving average techniques. In addition to works specific to stochastic zeroth-order Riemannian optimization, numerous researchers have contributed to the field of Riemannian optimization, including (Huang et al., 2015; Gao et al., 2018; Sato et al., 2019; Chen et al., 2020; Gao et al., 2021; Ruszczyński, 2021), just to name a few.

Yet to the best of our knowledge, no prior art from (stochastic zeroth-order) Riemannian optimization literature focuses on approximating the global optimum over a compact Riemannian manifold for functions that can have discontinuities in its domain. Therefore, our results might be of independent interest to the Rimannian optimization community.

Paper Organization. The rest of the paper is organized as follows. In Section 2, we list several basic concepts and conventions for the problem. In Section 3, we introduce the Geometric Narrowing (GN) algorithm. In Section 4, we provide lower bound analysis for batched bandits for nondegenerate functions.

2 Preliminaries

Perhaps we shall begin with the formal definition of doubling metric spaces, since it underpins the entire problem.

Definition 1 (Doubling metric space). *The doubling constant of a metric space $(\mathcal{X}, \mathcal{D})$ is the minimal N such that for all $\mathbf{x} \in \mathcal{X}$, for all $r > 0$, the ball $\mathbb{B}(\mathbf{x}, r) := \{\mathbf{z} \in \mathcal{X} : \mathcal{D}(\mathbf{z}, \mathbf{x}) \leq r\}$ can be covered by N balls of radius $\frac{r}{2}$. A metric space is called doubling if $N < \infty$. The doubling dimension of $\mathcal{X}$ is $d = \log_2(N)$ where N is the doubling constant of $\mathcal{X}$.*

An immediate consequence of the definition of doubling metric spaces is the following proposition.

Proposition 2. *Let $(\mathcal{X}, \mathcal{D})$ be a doubling metric space. For each $\mathbf{x} \in \mathcal{X}$ and $r \in (0, \infty)$, the ball $\mathbb{B}(\mathbf{x}, r)$ can be covered by 2^{kd} balls of radius $r \cdot 2^{-k}$ for any $k \in \mathbb{N}$, where d is the doubling dimension of $(\mathcal{X}, \mathcal{D})$.*

On the basis of doubling metric spaces, we formally define nondegenerate functions.

Definition 2 (Nondegenerate functions). *Let $(\mathcal{X}, \mathcal{D})$ be a doubling metric space. A function $f : \mathcal{X} \to \mathbb{R}$ is called nondegenerate if the followings hold:*

- $\inf_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) > -\infty$ *and f attains its unique minimum at $\mathbf{x}^* \in \mathcal{X}$.*
- *There exist $L \geq \lambda > 0$ and $q \geq 1$, such that $\lambda \left(\mathcal{D}(\mathbf{x}, \mathbf{x}^*)\right)^q \leq f(\mathbf{x}) - f(\mathbf{x}^*) \leq L \left(\mathcal{D}(\mathbf{x}, \mathbf{x}^*)\right)^q$, for all $\mathbf{x} \in \mathcal{X}$.*

The constants L, λ, q are referred to as nondegenerate parameters of function f.

Before proceeding further, we introduce the following notations and conventions for convenience.

- For two set $S, S' \subset \mathcal{X}$, define

$$D(S, S') := \sup_{\mathbf{x} \in S, \mathbf{x}' \in S'} \mathcal{D}(\mathbf{x}, \mathbf{x}'). \tag{4}$$

- For any $z > 0$, define $[z]_2 := 2^{\lceil \log_2 z \rceil}$.
- Throughout the paper, all numbers except for the time horizon T, doubling dimension d, and rounds of communications M, are regarded as constants.

3 The Geometric Narrowing Algorithm

Our algorithm for solving batched nondegenerate bandits is called Geometric Narrowing (GN). As the name suggests, the GN algorithm progressively narrows down the search space, and eventually lands in a small neighborhood of $\mathbf{x}^*$. To achieve this, we need to identify the specific regions of the space that should be eliminated. Additionally, we want to achieve a near-optimal regret rate using only approximately $\log \log T$ batches.

Perhaps the best way to illustrate the idea of the algorithm is through visuals. In Figure 4, we provide an example of how function evaluations and nondegenerate properties jointly narrow down the search space. Yet a naive utilization of the observations in Figure 4 is insufficient to design an efficient algorithm. Indeed, the computational cost grows quickly as the number of function value samples accumulates, even for the toy example shown in Figure 4. To overcome this, we succinctly summarize the observations illustrated in Figure 4 as an algorithmic procedure.

In addition to the narrowing procedure shown in Figure 4, we also need to determine the batching mechanism, in order to achieve the $\mathcal{O}(\log \log T)$ communication bound. This communication scheme

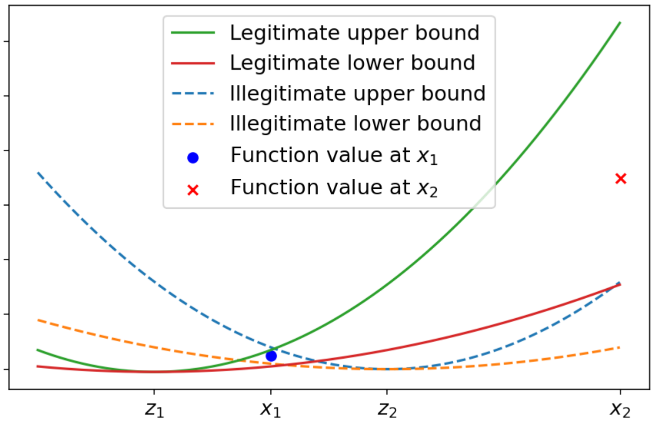


Figure 4: Illustration of the execution procedure of the GN algorithm over an interval. The function values at $\mathbf{x}_1$ and $\mathbf{x}_2$ jointly narrow down the range of $\mathbf{x}^*$. To ensure the function values at $\mathbf{x}_1$ and $\mathbf{x}_2$ fall between the upper and lower bounds for the nondegenerate function, the minimum of the function has to reside in a certain range. In this figure, the solid lines show a pair of legitimate bound, implying that the underlying functions may take its minimum at $\mathbf{z}_1$; the dashed lines show a pair of legitimate bound, implying that the underlying functions cannot take its minimum at $\mathbf{z}_2$, neither in a neighborhood of $\mathbf{z}_2$.

is described through a radius sequence in Definition 3. The procedure of GN is in Algorithm 1. In Figure 5, we demonstrate an example run of the GN algorithm.

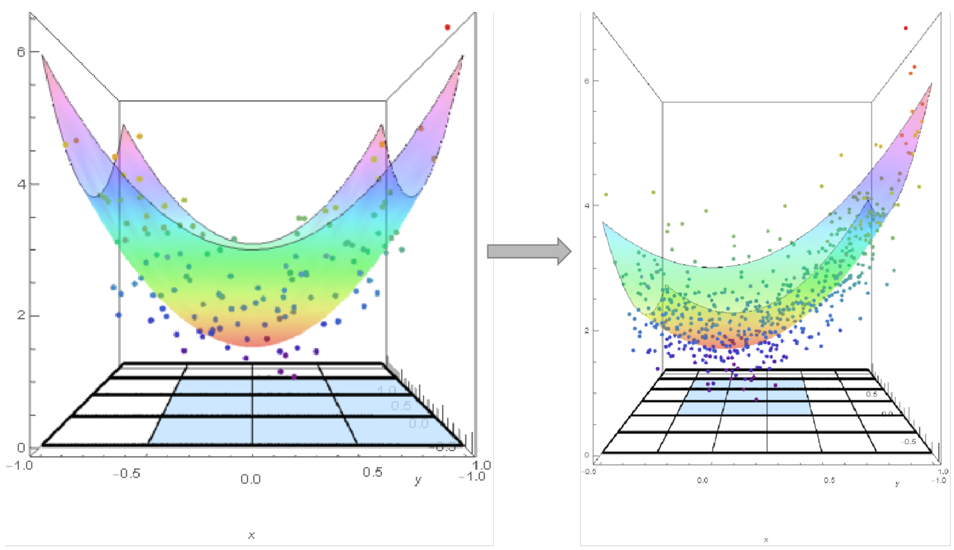


Figure 5: An example run of the GN algorithm. The surface shows the expected loss function, and the scattered points are loss samples over the current domain. These two plots describe the delete and split operations between adjacent batches of a GN run.

Definition 3. *For $d > 0$ and $q \geq 1$, we define $\hat{c}_1 = \frac{1}{2(2q+d)\log 2} \log \frac{T}{\log T}$ and $\hat{c}_{i+1} = \hat{\eta}\hat{c}_i$ for $i = 1, 2 \ldots$, where $\hat{\eta} = \frac{q+d}{2q+d}$. Then we define a sequence $\{\hat{r}_m\}_m$ by $\hat{r}_m = 2^{-\sum_{i=1}^{m} \hat{c}_i}$ for $m = 1, 2 \ldots$. On the basis*

Algorithm 1 Geometric Narrowing (GN) for Nondegenerate Functions

1: **Input.** Space $(\mathcal{X}, \mathcal{D})$; time horizon T; Number of batches $2M$.
/* Without loss of generality, let the diameter of $\mathcal{X}$ be 1: $\mathrm{Dim}(\mathcal{X}) = 1$. */
2: **Initialization.** Rounded Radius sequence $\{\bar{r}_m\}_{m=1}^{2M}$ defined in Definition 3; The first communication point $t_0 = 0$; Cover $\mathcal{X}$ by $\bar{r}_1$-balls, and define $\mathcal{A}_1^{pre}$ as the collection of these balls.
3: Compute $n_m = \frac{16 \log T}{\lambda^2 \bar{r}_m^{2q}}$ for $m = 1, \cdots, 2M$.
4: **for** $m = 1, 2, \cdots, 2M$ **do**
5: If $\bar{r}_m > \bar{r}_{m-1}$, then **continue**. /* Skip the rest of the steps in the current iteration, and enter the next iteration. */
6: For each ball $B \in \mathcal{A}_m^{pre}$, play arms $\mathbf{x}_{B,1}, \cdots, \mathbf{x}_{B,n_m}$, all located at the region of B.
7: Collect the loss samples $y_{B,1}, \cdots, y_{B,n_m}$ associated with $\mathbf{x}_{B,1}, \cdots, \mathbf{x}_{B,n_m}$. Compute the average loss for each B, $\widehat{f}_m(B) := \frac{\sum_{i=1}^{n_m} y_{B,i}}{n_m}$ for each ball $B \in \mathcal{A}_m^{pre}$. Find $\widehat{f}_m^{\min} = \min_{B \in \mathcal{A}_m} \widehat{f}_m(B)$. Let $B_m^{\min}$ be the ball where $\widehat{f}_m^{\min}$ is obtained.
8: Define

$$\mathcal{A}_m := \left\{ B \in \mathcal{A}_m^{pre} : D(B, B_m^{\min}) \leq \left(2 + \left(\frac{\lambda + L}{\lambda} \right)^{\frac{1}{q}} \right) \bar{r}_m \right\}.$$

9: For each ball $B \in \mathcal{A}_m$, use $(\bar{r}_m / \bar{r}_{m+1})^d$ balls of radius $\bar{r}_{m+1}$ to cover B, and define $\mathcal{A}_{m+1}^{pre}$ as the collection of these balls.
/* Due to Definition 1, we can cover $B \in \mathcal{A}_m$ by $(\bar{r}_m / \bar{r}_{m+1})^d$ balls of radius $\bar{r}_{m+1}$. */
10: Compute $t_{m+1} = t_m + (\bar{r}_m / \bar{r}_{m+1})^d \cdot |\mathcal{A}_m| \cdot n_{m+1}$. If $t_{m+1} \geq T$ then **break**.
11: **end for**
12: **Cleanup:** Pick a point in the region that is not eliminated, and play this point. Repeat this operation until all T steps are used.
13: **Output (optional):** Arbitrarily pick $\mathbf{x}_{out} \in \cup_{B \in \mathcal{A}_{2M}} B$ as an approximate for $\mathbf{x}^*$. /* This output step is optional, and only used for best arm identification or stochastic optimization tasks. */

of $\{\hat{c}_m\}_m$, we define $\hat{l}_m = \lfloor \sum_{i=1}^m \hat{c}_i \rfloor$ and $\hat{u}_m = \lceil \sum_{i=1}^m \hat{c}_i \rceil$. Then we define Rounded Radius (RR) Sequence*: $\bar{r}_m$, $m = 1, \ldots, 2M$:*

$$\bar{r}_m = \begin{cases} \bar{r}_{2k-1} = 2^{-\hat{l}_k} = 2^{-\lfloor \sum_{i=1}^k \hat{c}_i \rfloor} & \text{if } m = 2k-1, k = 1, \ldots, M \\ \bar{r}_{2k} = 2^{-\hat{u}_k} = 2^{-\lceil \sum_{i=1}^k \hat{c}_i \rceil} & \text{if } m = 2k, k = 1, \ldots, M. \end{cases}$$

From the above definition, we have $\bar{r}_{2k} \leq \hat{r}_k \leq \bar{r}_{2k-1}$ for $k = 1, \ldots, M$.

3.1 Analysis of the GN Algorithm

We start with the following simple concentration lemma.

Lemma 1. *Under Theorem 1's assumption, define*

$$\mathcal{E} := \left\{ \left| \widehat{f}_m(B) - \mathbb{E}\left[\widehat{f}_m(B) \right] \right| \leq \sqrt{\frac{4 \log T}{n_m}}, \quad \forall 1 \leq m \leq 2M, \ \forall B \in \mathcal{A}_m^{pre} \right\}.$$

It holds that $\mathbb{P}(\mathcal{E}) \geq 1 - 2T^{-1}$.

Proof of Lemma 1. Fix a ball $B \in \mathcal{A}_m^{pre}$. Recall the average loss of $B \in \mathcal{A}_m^{pre}$ is defined as

$$\widehat{f}_m(B) = \frac{\sum_{i=1}^{n_m} y_{B,i}}{n_m}.$$

We also have

$$\mathbb{E}\left[\widehat{f}_m(B)\right] = \frac{\sum_{i=1}^{n_m} f(\mathbf{x}_{B,i})}{n_m}.$$

Since $\widehat{f}_m(B) - \mathbb{E}\left[\widehat{f}_m(B)\right]$ is centered at zero, and is $\frac{1}{n_m}$-sub-Gaussian (e.g., Section 2.3 in Boucheron et al., 2013), applying the Chernoff bound gives

$$\mathbb{P}\left(\left|\widehat{f}_m(B) - \mathbb{E}\left[\widehat{f}_m(B)\right]\right| \geq \sqrt{\frac{4\log T}{n_m}}\right) \leq \frac{2}{T^2}.$$

Apparently, there are no more than T balls that contain observations. Thus a union bound over these balls finishes the proof. $\square$

Next in Lemma 2, we show that under event $\mathcal{E}$, the GN algorithm has nice properties.

Lemma 2. *Under event $\mathcal{E}$ (defined in Lemma 1), the following properties hold:*

- *The optimal point $\mathbf{x}^*$ is not removed;*
- *For any $B \in \mathcal{A}_m$, $\mathcal{D}(\mathbf{x}, \mathbf{x}^*) \leq \left(2 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)\bar{r}_m$ for all $\mathbf{x} \in \cup_{B \in \mathcal{A}_m} B$.*

Proof. For each m, let B_m^* denote the ball in $\mathcal{A}_m$ such that $B_m^* \ni \mathbf{x}^*$. For each m and $B \in \mathcal{A}_m$, we use $\mathbf{x}_m(B)$ to denote the center of the ball B. Let $\mathcal{E}$ be true. We know

$$\begin{aligned}
0 &\geq \widehat{f}_m(B_m^{\min}) - \widehat{f}_m(B_m^*) \\
&= \underbrace{\widehat{f}_m(B_m^{\min}) - f(\mathbf{x}_m(B_m^{\min}))}_{①} + \underbrace{f(\mathbf{x}_m(B_m^{\min})) - f(\mathbf{x}^*)}_{②} + \underbrace{f(\mathbf{x}^*) - f(\mathbf{x}_m(B_m^*))}_{③} + \underbrace{f(\mathbf{x}_m(B_m^*)) - \widehat{f}_m(B_m^*)}_{④} \\
&\geq -4\sqrt{\frac{\log T}{n_m}} + \lambda\left(\mathcal{D}\left(\mathbf{x}_m(B_m^{\min}), \mathbf{x}^*\right)\right)^q - L\bar{r}_m^q,
\end{aligned}$$

where for ① and ④ we use Lemma 1, for ③ we use property of the nondegenerate function, and ② is evidently nonnegative.

Since $4\sqrt{\frac{\log T}{n_m}} = \lambda \bar{r}_m^q$, we know that with high probability

$$\mathcal{D}\left(\mathbf{x}_m(B_m^{\min}), \mathbf{x}^*\right) \leq \left(\frac{\lambda+L}{\lambda}\right)^{1/q} \cdot \bar{r}_m.$$

Let B_m^* be the cube in $\mathcal{A}_m^{pre}$ that contains $\mathbf{x}^*$. This implies that $D(B_m^*, B_m^{\min}) \leq \left(2 + \left(\frac{\lambda+L}{\lambda}\right)^{1/q}\right)\bar{r}_m$, and thus the optimal arm is not eliminated. Since (1) the optimal arm is not eliminated, and (2) the diameter of $\cup_{B \in \mathcal{A}_m} B$ is no larger than $\left(2 + 2\left(\frac{\lambda+L}{\lambda}\right)^{1/q}\right)\bar{r}_m$, we have also proved the second item. $\square$

With Lemmas 1 and 2 in place, we are ready to prove Theorem 1.

Proof of Theorem 1. For each m, we introduce $S_m := \cup_{B \in \mathcal{A}_m} B$ to simplify notation. By the algorithm procedure, the diameter of S_m is bounded by $\left(3 + 2\left(\frac{\lambda+L}{\lambda}\right)^{1/q}\right)\bar{r}_m$. By Proposition 2, we have

$$|\mathcal{A}_m| \leq \left\lceil 3 + 2\left(\frac{\lambda+L}{\lambda}\right)^{1/q}\right\rceil_2^d,$$

which gives

$$|\mathcal{A}_m^{pre}| \leq \left(\frac{\bar{r}_{m-1}}{\bar{r}_m}\right)^d \left[3 + 2\left(\frac{\lambda+L}{\lambda}\right)^{1/q}\right]_2^d. \tag{5}$$

Note that the m-th batch incurs no regret if it is skipped. Thus it suffices to consider the case where the m-th batch is not skipped. For $m = 2k-1$, we can bound the regret in the $(2k-1)$-th batch (denoted by R_{2k-1}) by

$$\begin{aligned} R_{2k-1} &\leq |\mathcal{A}_{2k-1}^{pre}| \cdot n_{2k-1} \cdot L \cdot \left(2 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^q \bar{r}_{2k-2}^q \\ &\leq \left(\frac{\bar{r}_{m-1}}{\bar{r}_m}\right)^d \left[3 + 2\left(\frac{\lambda+L}{\lambda}\right)^{1/q}\right]_2^d \cdot n_{2k-1} \cdot L \cdot \left(2 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^q \bar{r}_{2k-2}^q, \end{aligned}$$

where the first line uses Lemma 2. Plugging $n_{2k-1} = \frac{16 \log T}{\lambda^2 \bar{r}_{2k-1}^{2q}}$ into the above inequality gives

$$\begin{aligned} R_{2k-1} &\leq \left(\frac{\bar{r}_{m-1}}{\bar{r}_m}\right)^d \left[3 + 2\left(\frac{\lambda+L}{\lambda}\right)^{1/q}\right]_2^d \cdot \frac{16 \log T}{\lambda^2 \bar{r}_{2k-1}^{2q}} \cdot L \cdot \left(2 + \left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^q \bar{r}_{2k-2}^q \\ &\leq L \left(2 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^q \left[3 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right]_2^d \frac{16 \log T}{\lambda^2} \bar{r}_{2k-2}^{q+d} \bar{r}_{2k-1}^{-2q-d} \\ &\leq L \left(2 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^q \left[3 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right]_2^d \frac{16 \log T}{\lambda^2} \hat{r}_{k-1}^{q+d} \hat{r}_k^{-2q-d}, \end{aligned}$$

where the last inequality follows from the definitions of $\hat{r}_m$ and $\bar{r}_m$.

By definition of the sequence $\{\hat{r}_m\}$, we have, for any m, $\hat{r}_{m-1}^{q+d} \hat{r}_m^{-2q-d} = 2^{-(q+d)\sum_{i=1}^{m-1} \hat{c}_i + (2q+d)\sum_{i=1}^{m} \hat{c}_i} = 2^{(2q+d)\hat{c}_m + q\sum_{i=1}^{m-1} \hat{c}_i} = 2^{(2q+d)\hat{c}_1}$. Thus we can upper bound R_{2k-1} by

$$R_{2k-1} \leq L \left(2 + \left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^q \left[3 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right]_2^d \cdot \frac{16}{\lambda^2} \cdot \sqrt{T \log T}. \tag{6}$$

For $m = 2k$, the regret in batch $2k$ (written R_{2k}) is bounded by

$$R_{2k} \leq |\mathcal{A}_{2k}^{pre}| \cdot n_{2k} \cdot L \cdot \left(2 + \left(\frac{\lambda+L}{\lambda}\right)^{1/q}\right)^q \bar{r}_{2k}^q.$$

Bringing (5) and definition of n_{2k} into the above inequality, and noticing $\frac{\bar{r}_{m-1}}{\bar{r}_m} \leq 2$ (for even m) gives

$$\begin{aligned} R_{2k} &\leq L 2^d \left(2 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^q \left[3 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right]_2^d \frac{16 \log T}{\lambda^2} \bar{r}_{2k}^{-q} \\ &\leq L 2^{d+q} \left(2 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^q \left[3 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right]_2^d \frac{16 \log T}{\lambda^2} \hat{r}_k^{-q}, \end{aligned}$$

where for the last inequality, we use definitions of $\bar{r}_m$ and $\hat{r}_m$ to get $\bar{r}_{2k}^{-1} \leq 2 \cdot \hat{r}_k^{-1}$. Again by definition of $\widehat{r}_m$, we have $\hat{r}_m^{-1} \leq 2^{\hat{c}_1 \frac{1}{1-\hat{\eta}}}$, and thus the regret in batch $2k$ is at most

$$R_{2k} \leq L 2^{d+q} \left(2 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^q \left[3 + 2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right]_2^d \frac{16}{\lambda^2} \sqrt{T \log T}. \tag{7}$$

For the cleanup phase, the regret (written R_{2B+1}) is bounded by

$$\begin{aligned}
R_{2M+1} &\le L\left(2+2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^{q} \bar{r}_{2M}^{q} T \\
&\le L\left(2+2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^{q} \sqrt{T\log T}\left(\frac{T}{\log T}\right)^{\frac{1}{2}\hat{\eta}^{M}}. \qquad (8)
\end{aligned}$$

Let there be in total $2M+1$ batches. Collecting terms from (6), (7) and (8) gives

$$\begin{aligned}
R^{GN}(T) \le\ & L\left(2+2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^{q}\left[3+2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right]_2^{d}\frac{16}{\lambda^2}\sqrt{T\log T}\cdot M \\
&+ L2^{d+q}\left(2+2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^{q}\left[3+2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right]_2^{d}\frac{16}{\lambda^2}\sqrt{T\log T}\cdot M \\
&+ L\left(2+2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^{q}\sqrt{T\log T}\left(\frac{T}{\log T}\right)^{\frac{1}{2}\hat{\eta}^{M}}
\end{aligned}$$

Now choose $M=\hat{M}^{*}=\frac{\log\log\frac{T}{\log T}}{\log\frac{1}{\hat{\eta}}}$, we have $\hat{\eta}^{\hat{M}^{*}}=\left(\log\frac{T}{\log T}\right)^{-1}$, then

$$\begin{aligned}
R^{GN}(T) \le\ & L(2^{d+q}+1)\left(2+2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right)^{q} \\
&\cdot\left(\frac{16}{\lambda^2}\left[3+2\left(\frac{\lambda+L}{\lambda}\right)^{\frac{1}{q}}\right]_2^{d}\frac{\log\log\frac{T}{\log T}}{\log(2q+d)-\log(q+d)}+e^{\frac{1}{2}}\right)\sqrt{T\log T}.
\end{aligned}$$

With this choice of M, only $\mathcal{O}\left(\log\log T\right)$ batches are needed. Q.E.D.

$\square$

Following the proof of Theorem 1, we can readily prove Corollary 1.

Proof of Corollary 1. Let the event $\mathcal{E}$ be true. From Definition 3, we know

$$\begin{aligned}
\bar{r}_{2M}^{q} &\le 2^{-q\hat{c}_1\frac{1-\hat{\eta}^{M}}{1-\hat{\eta}}} \\
&= 2^{-\frac{q}{2(2q+d)\log 2}\log\frac{T}{\log T}\cdot\frac{1-\hat{\eta}^{M}}{1-\hat{\eta}}} \\
&= \left(\frac{T}{\log T}\right)^{-\frac{q}{2(2q+d)}\cdot\frac{1-\left(\frac{q+d}{2q+d}\right)^{M}}{\frac{q}{2q+d}}} \\
&= \left(\frac{T}{\log T}\right)^{-\frac{1}{2}\left(1-\left(\frac{q+d}{2q+d}\right)^{M}\right)} \\
&= \sqrt{\frac{\log T}{T}}\cdot\left(\frac{T}{\log T}\right)^{\frac{1}{2}\hat{\eta}^{M}}.
\end{aligned}$$

Let $M=\frac{\log\log\frac{T}{\log T}}{\log\frac{1}{\hat{\eta}}}$, we have $\hat{\eta}^{M}=\left(\log\frac{T}{\log T}\right)^{-1}$, and thus

$$\bar{r}_{2M}^{q} \le e^{\frac{1}{2}}\sqrt{\frac{\log T}{T}}.$$

By Lemma 2, we know, under event $\mathcal{E}$,

$$f(\mathbf{x}_{out}) - f(\mathbf{x}^*) \le L\mathcal{D}\left(\mathbf{x}_{out}, \mathbf{x}^*\right)^q \le L\left(2 + 2\left(\frac{\lambda + L}{\lambda}\right)^{\frac{1}{q}}\right)^q \bar{r}_{2M}^q \le e^{\frac{1}{2}} L\left(2 + 2\left(\frac{\lambda + L}{\lambda}\right)^{\frac{1}{q}}\right)^q \sqrt{\frac{\log T}{T}}$$

We conclude the proof by noticing that the $\mathcal{E}$ holds true with probability exceeding $1 - \frac{2}{T}$. □

4 Lower Bound Analysis

First of all, we need to identify a particular doubling metric space to work with. Hinted by the celebrated Assouad's embedding theorem, we turn to the Euclidean space with a specific metric. For any d, the doubling metric space we choose is $(\mathbb{R}^{\lfloor d \rfloor}, \|\cdot\|_\infty)$. One important reason for this choice is that the doubling dimension of this space equal its dimension as a vector space. Throughout the rest of this paper, without loss of generality, we let d be an integer, and consider the metric space $(\mathbb{R}^d, \|\cdot\|_\infty)$.

Remark 3. *By Assouad's embedding theorem, one can embed a separable metric space* $(\mathcal{X}, \mathcal{D})$ *with doubling number* N *into a Euclidean space with some distortion, hence our research works in general doubling metric space.*

After settling the metric space to work with, we still need to overcome previously unencountered challenges. To further illustrate these challenges, let us review the lower bound strategy for Lipschitz bandits. In proving the lower bound for Lipschitz bandits (Kleinberg, 2005; Kleinberg et al., 2008; Bubeck et al., 2011a), one essentially use the packing/covering number for the underlying space, and this packing number essentially serves as number of arms in the lower bound proof. For our problem, however, the lower bound argument for Lipschitz does not carry through. The reasons are:

- First and foremost, a nondegenerate function may be discontinuous. Restriction to Lipschitz bandit instances rules out a large class of problem instances.
- More importantly, in the lower bound argument for Lipschitz bandits, one construct instances with small "peaks" in the domain. We then let the height of the peak to decrease with the total time horizon T, so that no algorithm can quickly find the peaks for all instances. However, for nondegenerate functions, the nondegenerate parameters do not depend on T. Therefore, we are not allowed to tweak the landscape of the instances as freely as previously done for Lipschitz bandits.

On top of the above challenges, we need to incorporate the communication pattern into the entire analysis. To tackle all these difficulties, we use a *bitten-apple* trick in the instance construction. Specific examples of bitten-apple instances are shown in Figures 6 and 7.

4.1 The instances

To formally define the instances, we first-of-all partition the space $\mathbb{R}^d$ into 2^d orthants $O_1, O_2, \cdots, O_{2^d}$. We represent the natural numbers $1, 2, \cdots, 2^d$ by a sequence of $+/-$ signs. That is, for any $k = 1, 2, \cdots, 2^d$, we use $\left(s_1^k, \cdots, s_d^k\right) \in \{-1, +1\}^d$ to represent k. This representation is equivalent to writing k as a base-two number. For $k = 1, 2, \cdots, 2^d$ and a number $\epsilon \in (0, 1)$, we define $\mathbf{x}_{k,\epsilon}^* = \left(s_1^k \epsilon, s_2^k \epsilon, \cdots, s_d^k \epsilon\right)$. Clearly, $\|\mathbf{x}_{k,\epsilon}^*\|_\infty = \epsilon$ for $k = 1, 2, \cdots, 2^d$. As a convention, we let O_1 be the orthant associated with $(+, +, \cdots, +)$.

Firstly, we introduce a sequence of reference communication points $\mathcal{T}_r = \{T_1, \cdots, T_M\}$ and the corresponding gaps $\{\epsilon_1^q, \cdots, \epsilon_M^q\}$, defined as

$$T_j = \lfloor T^{\frac{1-2^{-j}}{1-2^{-M}}} \rfloor, \qquad \epsilon_j^q = \frac{1}{4} \cdot \frac{\sqrt{2}}{2} \cdot \frac{\sqrt{2^d - 1}}{2^q + 2} \cdot \frac{1}{M} \cdot T^{-\frac{1}{2} \cdot \frac{1-2^{1-j}}{1-2^{-M}}}, \qquad j \in [M]. \tag{9}$$

Then we construct collections of instances $\mathcal{I}_1, \cdots, \mathcal{I}_M$. Each instance is defined by a mean loss function f and a noise distribution. For our purpose, we let the noise be standard Gaussian. That is, the observed loss samples at $\mathbf{x}$ are *iid* from the Gaussian distribution $\mathcal{N}(f(\mathbf{x}), 1)$. For $1 \le j \le M-1$, we let $\mathcal{I}_j = \{I_{j,k}\}_{k=1}^{2^d-1}$ and the expected loss function of $I_{j,k}$ is defined as

$$f_{j,k}^{\epsilon_j}(\mathbf{x}) = \begin{cases} \|\mathbf{x} - \mathbf{x}_{k,\epsilon_j}^*\|_\infty^q - \|\mathbf{x}_{k,\epsilon_j}^*\|_\infty^q, & \text{if } \mathbf{x} \in \mathbb{B}(\mathbf{x}_{k,\epsilon_j}^*, \epsilon_j) \backslash \mathbb{B}(0, \frac{\epsilon_j}{2}), \\ \|\mathbf{x} - \mathbf{x}_{2^d, \frac{\epsilon_M}{3}}^*\|_\infty^q - \|\mathbf{x}_{2^d, \frac{\epsilon_M}{3}}^*\|_\infty^q, & \text{if } \mathbf{x} \in \mathbb{B}(\mathbf{x}_{2^d, \frac{\epsilon_M}{3}}^*, \frac{\epsilon_M}{3}) \backslash \mathbb{B}(0, \frac{\epsilon_M}{6}), \\ \|\mathbf{x}\|_\infty^q, & \text{otherwise.} \end{cases} \tag{10}$$

For $j = M$, we let $\mathcal{I}_M = \{I_M\}$ and the expected loss function of I_M is defined as

$$f_{M,k}^{\epsilon_M}(\mathbf{x}) = \begin{cases} \|\mathbf{x} - \mathbf{x}_{2^d, \frac{\epsilon_M}{3}}^*\|_\infty^q - \|\mathbf{x}_{2^d, \frac{\epsilon_M}{3}}^*\|_\infty^q, & \text{if } \mathbf{x} \in \mathbb{B}(\mathbf{x}_{2^d, \frac{\epsilon_M}{3}}^*, \frac{\epsilon_M}{3}) \backslash \mathbb{B}(0, \frac{\epsilon_M}{6}), \\ \|\mathbf{x}\|_\infty^q, & \text{otherwise.} \end{cases} \tag{11}$$

Note that $f_{M,k}^{\epsilon_M}(\mathbf{x})$ is independent of k. Here we keep the subscript k for notational consistency. Figures 6 and 7 plot examples of $f_{j,k}^{\epsilon_j}$ and $f_{M,k}^{\epsilon_M}$.

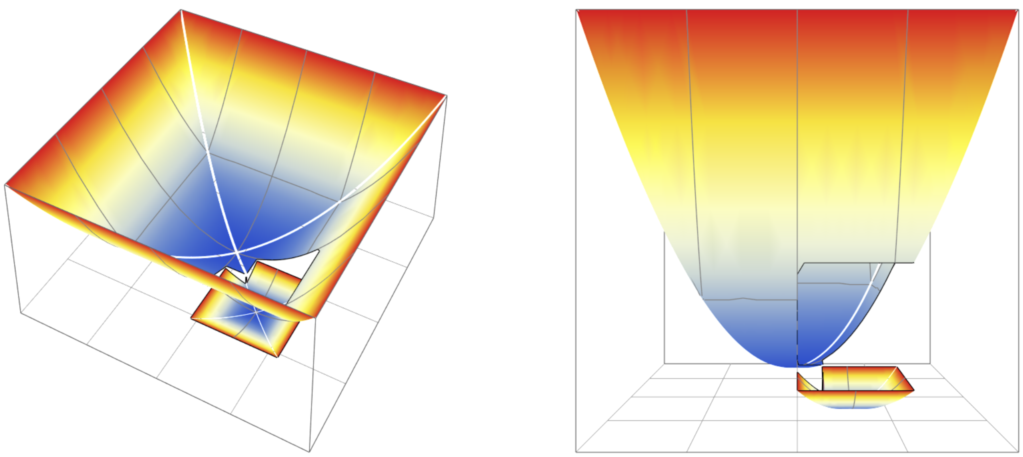


Figure 6: Example plot of $f_{M,k}^{\epsilon_M}(\mathbf{x})$ with $d = q = 2$. The two graphs come from different views of the same function.

On the basis of $\{f_{j,k}\}_{j \in [M], k \in [2^d-1]}$, we construct another series of problem instances $\{I_{j,k,l}\}_{j \in [M], k \in [2^d-1], l \in [2^d]}$:

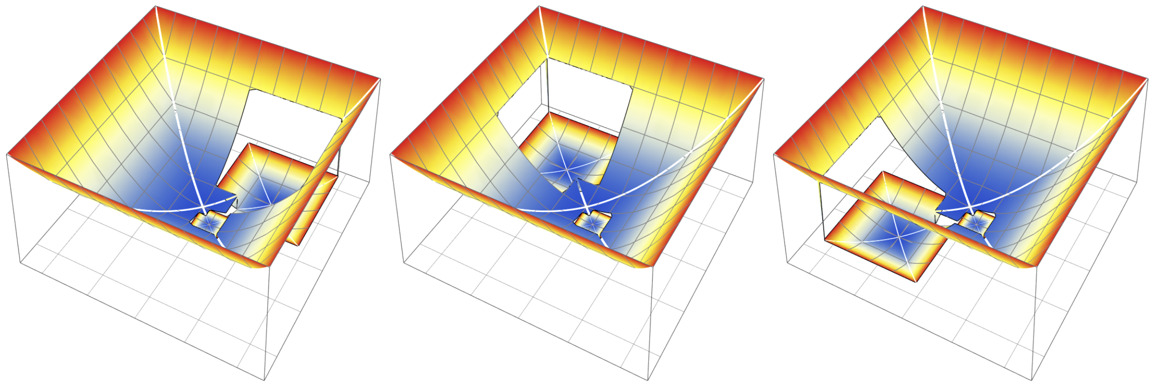
- For $j < M$, $l \neq k$ and $l < 2^d$, the loss function of problems instance $I_{j,k,l}$ is defined as

$$f_{j,k,l}^{\epsilon_j}(\mathbf{x}) = \begin{cases} \|\mathbf{x} - \mathbf{x}_{k,\epsilon_j}^*\|_\infty^q - \|\mathbf{x}_{k,\epsilon_j}^*\|_\infty^q, & \text{if } \mathbf{x} \in \mathbb{B}(\mathbf{x}_{k,\epsilon_j}^*, \epsilon_j) \backslash \mathbb{B}(0, \frac{\epsilon_j}{2}), \\ \|\mathbf{x} - 2^{\frac{1}{q}} \cdot \mathbf{x}_{l,\epsilon_j}^*\|_\infty^q - \|2^{\frac{1}{q}} \cdot \mathbf{x}_{l,\epsilon_j}^*\|_\infty^q, & \text{if } \mathbf{x} \in \mathbb{B}(2^{\frac{1}{q}} \cdot \mathbf{x}_{l,\epsilon_j}^*, 2^{\frac{1}{q}} \cdot \epsilon_j) \backslash \mathbb{B}(0, \frac{2^{\frac{1}{q}} \cdot \epsilon_j}{2}), \\ \|\mathbf{x} - \mathbf{x}_{2^d, \frac{\epsilon_M}{3}}^*\|_\infty^q - \|\mathbf{x}_{2^d, \frac{\epsilon_M}{3}}^*\|_\infty^q, & \text{if } \mathbf{x} \in \mathbb{B}(\mathbf{x}_{2^d, \frac{\epsilon_M}{3}}^*, \frac{\epsilon_M}{3}) \backslash \mathbb{B}(0, \frac{\epsilon_M}{6}), \\ \|\mathbf{x}\|_\infty^q, & \text{otherwise.} \end{cases}$$

- For $j < M$, $l = k < 2^d$, we let $f_{j,k,l}^{\epsilon_j}(\mathbf{x}) := f_{j,k}^{\epsilon_j}(\mathbf{x})$, which is defined in (10).

- For $j < M$, $k < 2^d$, and $l = 2^d$, we define

$$f_{j,k,2^d}^{\epsilon_j}(\mathbf{x}) = \begin{cases} \|\mathbf{x} - \mathbf{x}_{k,\epsilon_j}^*\|_\infty^q - \|\mathbf{x}_{k,\epsilon_j}^*\|_\infty^q, & \text{if } \mathbf{x} \in \mathbb{B}(\mathbf{x}_{k,\epsilon_j}^*, \epsilon_j) \backslash \mathbb{B}(0, \frac{\epsilon_j}{2}), \\ \|\mathbf{x} - 2^{\frac{1}{q}} \cdot \mathbf{x}_{2^d,\epsilon_j}^*\|_\infty^q - \|2^{\frac{1}{q}} \cdot \mathbf{x}_{2^d,\epsilon_j}^*\|_\infty^q, & \text{if } \mathbf{x} \in \mathbb{B}(2^{\frac{1}{q}} \cdot \mathbf{x}_{2^d,\epsilon_j}^*, 2^{\frac{1}{q}} \cdot \epsilon_j) \backslash \mathbb{B}(0, \frac{2^{\frac{1}{q}} \cdot \epsilon_j}{2}), \\ \|\mathbf{x}\|_\infty^q, & \text{otherwise.} \end{cases}$$



(a) $f^{\epsilon_j}_{j,1}(\mathbf{x})$ (b) $f^{\epsilon_j}_{j,2}(\mathbf{x})$ (c) $f^{\epsilon_j}_{j,3}(\mathbf{x})$

Figure 7: Example instance $f^{\epsilon_j}_{j,k}(\mathbf{x})$ with $d = q = 2$ for $1 \le j \le M-1$. The above three graphs from left to right show $f^{\epsilon_j}_{j,k}$ for $k = 1, 2, 3$.

- For $j = M$, $k < 2^d$, and $l < 2^d$, the corresponding loss function is defined as

$$f^{\epsilon_M}_{M,k,l}(\mathbf{x}) = \begin{cases} \|\mathbf{x} - 2^{\frac{1}{q}} \cdot \mathbf{x}^*_{l,\frac{\epsilon_M}{3}}\|^q_\infty - \|2^{\frac{1}{q}} \cdot \mathbf{x}^*_{l,\frac{\epsilon_M}{3}}\|^q_\infty, & \text{if } \mathbf{x} \in \mathbb{B}(2^{\frac{1}{q}} \cdot \mathbf{x}^*_{l,\frac{\epsilon_M}{3}}, 2^{\frac{1}{q}} \cdot \frac{\epsilon_M}{3}) \backslash \mathbb{B}(0, \frac{2^{\frac{1}{q}} \cdot \epsilon_M}{6}), \\ \|\mathbf{x} - \mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|^q_\infty - \|\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|^q_\infty, & \text{if } \mathbf{x} \in \mathbb{B}(\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}, \frac{\epsilon_M}{3}) \backslash \mathbb{B}(0, \frac{\epsilon_M}{6}), \\ \|\mathbf{x}\|^q_\infty, & \text{otherwise.} \end{cases}$$

- For $j = M$, $k < 2^d$ and $l = 2^d$, we define $f^{\epsilon_M}_{M,k,2^d}(\mathbf{x}) := f^{\epsilon_M}_{M,k}(\mathbf{x})$. For the case where $j = M$, we keep the subscript k for the same reason as in (11).

Figure 8 depicts the partitioning of space for the function $f^{\epsilon_j}_{j,k,l}$ $(j < M, l < 2^d, l \neq k)$. In orthant O_k, O_l and O_{2^d}, the function $f^{\epsilon_j}_{j,k,l}$ differs from $\|\mathbf{x}\|^q_\infty$ in a region of a bitten-apple shape.

First of all, we verify that these functions are nondegenerate functions.

Proposition 3. *The functions* $\{f^{\epsilon_j}_{j,k}\}_{j\in[M],k\in[2^d-1]}$ *are nondegenerate with parameters independent of time horizon T, the doubling dimension d, and rounds of communications M.*

Proof of Proposition 3. We first consider $f^{\epsilon_j}_{j,k}$. Note that the minimum of $f^{\epsilon_j}_{j,k}$ is obtained at $\mathbf{x}^*_{k,\epsilon_j}$.

The lower bound:

For $\mathbf{x} \in \mathbb{B}(\mathbf{x}^*_{k,\epsilon_j}, \epsilon_j) \backslash \mathbb{B}(0, \frac{\epsilon_j}{2})$, we have $f^{\epsilon_j}_{j,k}(\mathbf{x}) - f^{\epsilon_j}_{j,k}(\mathbf{x}^*_{k,\epsilon_j}) = \|\mathbf{x} - \mathbf{x}^*_{k,\epsilon_j}\|^q_\infty$, which clearly satisfies the nondegenerate condition.

For $\mathbf{x} \in \mathbb{B}(\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}, \frac{\epsilon_M}{3}) \backslash \mathbb{B}(0, \frac{\epsilon_M}{6})$, since $\epsilon_M \le \epsilon_j$ for all $j = 1, 2, \cdots, M$, we have,

$$\begin{aligned} &\|\mathbf{x} - \mathbf{x}^*_{k,\epsilon_j}\|^q_\infty \le \left(2\|\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|_\infty + \|\mathbf{x}^*_{k,\epsilon_j}\|_\infty\right)^q \le 3^q \|\mathbf{x}^*_{k,\epsilon_j}\|^q_\infty \\ &\le 3^q \cdot 3^q \left(\|\mathbf{x}^*_{k,\epsilon_j}\|_\infty - \|\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|_\infty\right)^q \le 9^q \left(\|\mathbf{x}^*_{k,\epsilon_j}\|^q_\infty - \|\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|^q_\infty\right) \\ &\le 9^q \left(\|\mathbf{x}^*_{k,\epsilon_j}\|^q_\infty - \|\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|^q_\infty + \|\mathbf{x} - \mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|^q_\infty\right) = 9^q \left(f^{\epsilon_j}_{j,k}(\mathbf{x}) - f^{\epsilon_j}_{j,k}(\mathbf{x}^*_{k,\epsilon_j})\right). \end{aligned}$$

For $\mathbf{x}$ in other parts of the domain, we have

$$f^{\epsilon_j}_{j,k}(\mathbf{x}) - f^{\epsilon_j}_{j,k}(\mathbf{x}^*_{k,\epsilon_j}) = \|\mathbf{x}\|^q_\infty + \|\mathbf{x}^*_{k,\epsilon_j}\|^q_\infty \ge \frac{1}{2^{q-1}} \|\mathbf{x} - \mathbf{x}^*_{k,\epsilon_j}\|^q_\infty,$$

where the last inequality uses convexity of $\|\cdot\|^q_\infty$ and Jensen's inequality.

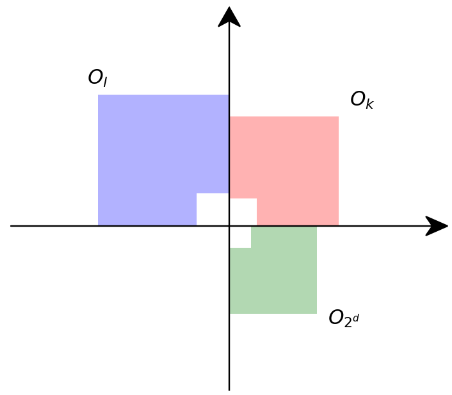


Figure 8: An illustration of how the space is partitioned for function $f_{j,k,l}^{\epsilon_j}$. In some particular orthant, the function $f_{j,k,l}^{\epsilon_j}$ differs from $\|\mathbf{x}\|_\infty^q$ in regions that resemble a bitten-apple shape. Such regions are illustrated as shaded areas in the figure.

The upper bound:
For $\mathbf{x} \in \mathbb{B}(\mathbf{x}^*_{k,\epsilon_j}, \epsilon_j) \backslash \mathbb{B}(0, \frac{\epsilon_j}{2})$, the nondegenerate condition holds true.
For $\mathbf{x} \in \mathbb{B}(\mathbf{x}^*_{2^d, \frac{\epsilon_M}{3}}, \frac{\epsilon_M}{3}) \backslash \mathbb{B}(0, \frac{\epsilon_M}{6})$,

$$\begin{aligned} & f_{j,k}^{\epsilon_j}(\mathbf{x}) - f_{j,k}^{\epsilon_j}(\mathbf{x}^*_{k,\epsilon_j}) = \|\mathbf{x} - \mathbf{x}^*_{2^d,\epsilon_M}\|_\infty^q + \epsilon_j^q - \left(\frac{\epsilon_M}{3}\right)^q \\ \leq\ & 2^{q-1}\left(\|\mathbf{x} - \mathbf{x}^*_{k,\epsilon_j}\|_\infty^q + \|\mathbf{x}^*_{k,\epsilon_j} - \mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|_\infty^q\right) + \epsilon_j^q \\ =\ & 2^{q-1}\|\mathbf{x} - \mathbf{x}^*_{k,\epsilon_j}\|_\infty^q + 2^{q-1}\left(\epsilon_j + \frac{\epsilon_M}{3}\right)^q + \epsilon_j^q \leq \left(2^q+1\right)^2 \|\mathbf{x} - \mathbf{x}^*_{k,\epsilon_j}\|_\infty^q, \end{aligned}$$

where the inequality on the first line uses convexity of $\|\cdot\|_\infty^q$ and Jensen's inequality.

For $\mathbf{x}$ in other parts of the domain, we have

$$f_{j,k}^{\epsilon_j}(\mathbf{x}) - f_{j,k}^{\epsilon_j}(\mathbf{x}^*_{k,\epsilon_j}) = \|\mathbf{x}\|_\infty^q + \|\mathbf{x}^*_{k,\epsilon_j}\|_\infty^q \leq 2^{q-1}\left(\|\mathbf{x} - \mathbf{x}^*_{k,\epsilon_j}\|_\infty^q + \|\mathbf{x}^*_{k,\epsilon_j}\|_\infty^q\right) + \|\mathbf{x}^*_{k,\epsilon_j}\|_\infty^q.$$

Since $\|\mathbf{x}^*_{k,\epsilon_j}\|_\infty \leq 2\|\mathbf{x} - \mathbf{x}^*_{k,\epsilon_j}\|_\infty$ for $\mathbf{x} \notin \left(\mathbb{B}(\mathbf{x}^*_{k,\epsilon_j}, \epsilon_j) \backslash \mathbb{B}(0, \frac{\epsilon_j}{2})\right)$, we continue from the above inequality and get

$$f_{j,k}^{\epsilon_j}(\mathbf{x}) - f_{j,k}^{\epsilon_j}(\mathbf{x}^*_{k,\epsilon_j}) \leq \left(2^q+1\right)^2 \|\mathbf{x} - \mathbf{x}^*_{k,\epsilon_j}\|_\infty^q.$$

Following the same procedure, we can check that the nondegenerate condition holds true for the function $f_{M,k}^{\epsilon_M}$.

□

Proposition 4. *The functions* $\{f_{j,k,l}^{\epsilon_j}\}_{j\in[M],k\in[2^d-1],l\in[2^d]}$ *are nondegenerate with parameters independent of time horizon* T*, the doubling dimension* d*, and rounds of communications* M*.*

Proof of Proposition 4. For $j \leq M-1$, first, consider $l < 2^d$. For $\mathbf{x} \in \mathbb{B}(\mathbf{x}^*_{k,\epsilon_j}, \epsilon_j) \backslash \mathbb{B}(0, \frac{\epsilon_j}{2})$, we have

$$\begin{aligned} & f_{j,k,l}^{\epsilon_j}(\mathbf{x}) - f_{j,k,l}^{\epsilon_j}(2^{1/q}\cdot\mathbf{x}^*_{l,\epsilon_j}) = \|\mathbf{x} - \mathbf{x}^*_{k,\epsilon_j}\|_\infty^q - \|\mathbf{x}^*_{k,\epsilon_j}\|_\infty^q + \|2^{1/q}\cdot\mathbf{x}^*_{l,\epsilon_j}\|_\infty^q \\ \geq\ & \|\mathbf{x}^*_{l,\epsilon_j}\|_\infty^q \geq \left(\frac{1}{6}\|\mathbf{x} - 2^{1/q}\cdot\mathbf{x}^*_{l,\epsilon_j}\|_\infty\right)^q, \end{aligned}$$

and

$$
\begin{aligned}
&f^{\epsilon_j}_{j,k,l}(\mathbf{x}) - f^{\epsilon_j}_{j,k,l}(2^{1/q}\cdot \mathbf{x}^*_{l,\epsilon_j}) = \|\mathbf{x}-\mathbf{x}^*_{k,\epsilon_j}\|^q_\infty - \|\mathbf{x}^*_{k,\epsilon_j}\|^q_\infty + \|2^{1/q}\cdot \mathbf{x}^*_{l,\epsilon_j}\|^q_\infty \\
&\le 2^{q-1}\|\mathbf{x}-2^{1/q}\cdot \mathbf{x}^*_{l,\epsilon_j}\|^q_\infty + 2^{q-1}\|\mathbf{x}^*_{k,\epsilon_j} - 2^{1/q}\cdot \mathbf{x}^*_{l,\epsilon_j}\|^q_\infty - \|\mathbf{x}^*_{k,\epsilon_j}\|^q_\infty + \|2^{1/q}\cdot \mathbf{x}^*_{l,\epsilon_j}\|^q_\infty \\
&\le 2^{q-1}\|\mathbf{x}-2^{1/q}\mathbf{x}^*_{l,\epsilon_j}\|^q_\infty + 2^{q-1}\cdot 3^q\epsilon_j^q - \epsilon_j^q + 2\epsilon_j^q \le (3^q+1)^2\|\mathbf{x}-2^{1/q}\mathbf{x}^*_{l,\epsilon_j}\|^q_\infty,
\end{aligned}
$$

where the last inequality uses that $\epsilon_j \le \|\mathbf{x}-2^{1/q}\cdot \mathbf{x}^*_{l,\epsilon_j}\|_\infty$.

For $\mathbf{x}$ in other parts of the domain, we use Proposition 3. For the case where $l=2^d$, we also apply Proposition 3.

For $j=M$, the proof follows analogously. □

In addition, we prove that the loss functions we construct satisfy the following properties.

Proposition 5. *For any $j=1,2,\cdots,M-1$ and $k=1,2,\cdots,2^d-1$, it holds that*

$$
\left|f^{\epsilon_j}_{j,k}(\mathbf{x}) - f^{\epsilon_M}_{M,k}(\mathbf{x})\right| \le \begin{cases} (2^q+2)\epsilon_j^q, & \text{if } \mathbf{x}\in \mathbb{B}(\mathbf{x}^*_{k,\epsilon_j},\epsilon_j)\backslash \mathbb{B}(0,\frac{\epsilon_j}{2}), \\ 0, & \text{otherwise.}\end{cases}
$$

Proof. For $\mathbf{x}\in \mathbb{B}(\mathbf{x}^*_{k,\epsilon_j},\epsilon_j)\backslash \mathbb{B}(0,\frac{\epsilon_j}{2})$, it holds that

$$
\left|f^{\epsilon_j}_{j,k}(\mathbf{x}) - f^{\epsilon_M}_{M,k}(\mathbf{x})\right| = \left|\|\mathbf{x}-\mathbf{x}^*_{k,\epsilon_j}\|^q_\infty - \|\mathbf{x}^*_{k,\epsilon_j}\|^q_\infty - \|\mathbf{x}\|^q_\infty\right| \le \epsilon_j^q + \epsilon_j^q + 2^q\epsilon_j^q = (2^q+2)\epsilon_j^q.
$$

For $\mathbf{x}\notin \mathbb{B}(\mathbf{x}^*_{k,\epsilon_j},\epsilon_j)\backslash \mathbb{B}(0,\frac{\epsilon_j}{2})$, $f^{\epsilon_j}_{j,k}(\mathbf{x})$ is identical to $f^{\epsilon_M}_{M,k}(\mathbf{x})$. This concludes the proof. □

Now for simplicity, we introduce the following notation: For $k=1,2,\cdots,2^d$, define

$$
S^\epsilon_k := \mathbb{B}(\mathbf{x}^*_{k,\epsilon},\epsilon).
$$

Proposition 6. *It holds that*

- *If $j<M$, $k<2^d$ and $l\neq k$*

$$
|f^{\epsilon_j}_{j,k,l}(\mathbf{x}) - f^{\epsilon_j}_{j,k,k}(\mathbf{x})| \le \begin{cases} 2(2^q+2)\epsilon_j^q, & \text{if } \mathbf{x}\in S_l^{2^{1/q}\epsilon_j} \\ 0, & \text{otherwise.}\end{cases}
$$

- *Also, if $k<2^d$ and $l<2^d$,*

$$
|f^{\epsilon_M}_{M,k,l}(\mathbf{x}) - f^{\epsilon_M}_{M,k,2^d}(\mathbf{x})| \le \begin{cases} 2(2^q+2)\epsilon_M^q, & \text{if } \mathbf{x}\in S_l^{2^{1/q}\epsilon_M} \\ 0, & \text{otherwise.}\end{cases}
$$

- *On instance $I_{j,k,l}$ ($j\in[M], k\in[2^d-1], l\in[2^d]$), pulling an arm that is not in $S_l^{2^{1/q}\epsilon_j}$ incurs a regret no smaller than $\frac{\epsilon_j^q}{3^q}$.*

Proof. **The first item.**

Case I: $j<M$ **and** $l<2^d$**.** For $\mathbf{x}\in \mathbb{B}(2^{\frac{1}{q}}\cdot \mathbf{x}^*_{l,\epsilon_j}, 2^{\frac{1}{q}}\cdot\epsilon_j)\backslash \mathbb{B}(0,\frac{2^{\frac{1}{q}}\cdot\epsilon_j}{2}) \subseteq S_l^{2^{1/q}\epsilon_j}$, it holds that

$$
\left|f^{\epsilon_j}_{j,k,l}(\mathbf{x}) - f^{\epsilon_j}_{j,k,k}(\mathbf{x})\right| = \left|\|\mathbf{x}-2^{\frac{1}{q}}\cdot \mathbf{x}^*_{l,\epsilon_j}\|^q_\infty - \|2^{\frac{1}{q}}\cdot \mathbf{x}^*_{l,\epsilon_j}\|^q_\infty - \|\mathbf{x}\|^q_\infty\right| \le 2\left(2^q+2\right)\epsilon_j^q.
$$

For $\mathbf{x}\notin S_l^{2^{1/q}\epsilon_j} = \mathbb{B}\left(2^{1/q}\cdot \mathbf{x}^*_{l,\epsilon_j}, 2^{1/q}\cdot\epsilon_j\right)$, $f^{\epsilon_j}_{j,k,l}(\mathbf{x})$ is identical to $f^{\epsilon_j}_{j,k,k}(\mathbf{x})$.

Case II: $j < M$ **and** $l = 2^d$**.** For $\mathbf{x} \in \mathbb{B}(2^{\frac{1}{q}} \cdot \mathbf{x}^*_{2^d,\epsilon_j}, 2^{\frac{1}{q}} \cdot \epsilon_j)$, it holds that

$$\begin{aligned}
&\left| f^{\epsilon_j}_{j,k,l}(\mathbf{x}) - f^{\epsilon_j}_{j,k,k}(\mathbf{x}) \right| \\
&\le \max \begin{cases} \left| \|\mathbf{x} - 2^{\frac{1}{q}} \cdot \mathbf{x}^*_{2^d,\epsilon_j}\|^q_\infty - \|2^{\frac{1}{q}} \cdot \mathbf{x}^*_{2^d,\epsilon_j}\|^q_\infty - \|\mathbf{x}\|^q_\infty \right|, & \text{if } ①; \\ \left| \|\mathbf{x} - 2^{\frac{1}{q}} \cdot \mathbf{x}^*_{2^d,\epsilon_j}\|^q_\infty - \|2^{\frac{1}{q}} \cdot \mathbf{x}^*_{2^d,\epsilon_j}\|^q_\infty - \|\mathbf{x} - \mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|^q_\infty + \|\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|^q_\infty \right|, & \text{if } ②; \\ \left| \|\mathbf{x}\|^q_\infty - \|\mathbf{x} - \mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|^q_\infty + \|\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|^q_\infty \right| & \text{if } ③; \\ 0, & \text{if } ④ \end{cases} \\
&\le 2\left(2^q + 2\right)\epsilon_j^q,
\end{aligned}$$

where ① stands for $\mathbf{x} \in \mathbb{B}\left(2^{1/q} \cdot \mathbf{x}^*_{2^d,\epsilon_j}, 2^{1/q} \cdot \epsilon_j\right) \backslash \mathbb{B}\left(0, \frac{2\epsilon_M}{3}\right)$, ② stands for

$$\mathbf{x} \in \mathbb{B}\left(\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}, \frac{\epsilon_M}{3}\right) \backslash \mathbb{B}\left(0, \frac{2^{1/q}\epsilon_j}{2}\right),$$

③ stands for

$$\mathbf{x} \in \mathbb{B}\left(\mathbf{x}^*_{2^d,\frac{2^{1/q}\cdot\epsilon_j}{4}}, \frac{2^{1/q}\epsilon_j}{4}\right) \backslash \mathbb{B}\left(0, \frac{\epsilon_M}{6}\right),$$

and ④ stands for $\mathbf{x}$ in other parts of $\mathbb{B}\left(2^{1/q} \cdot \mathbf{x}^*_{2^d,\epsilon_j}, 2^{1/q} \cdot \epsilon_j\right)$, and the last inequality uses that $\epsilon_M \le \epsilon_j$ for $j \le M$. The above derivation is valid even if some of ①–④ are empty.

Outside of $\mathbb{B}(2^{1/q} \cdot \mathbf{x}^*_{2^d,\epsilon_j}, 2^{1/q} \cdot \epsilon_j)$, $f^{\epsilon_j}_{j,k,2^d}$ is identical to $f^{\epsilon_j}_{j,k,k}$.

The second item. For $\mathbf{x} \in \mathbb{B}\left(2^{\frac{1}{q}} \cdot \mathbf{x}^*_{l,\frac{\epsilon_M}{3}}, 2^{\frac{1}{q}} \cdot \frac{\epsilon_M}{3}\right)$,

$$\begin{aligned}
|f^{\epsilon_M}_{M,k,l}(\mathbf{x}) - f^{\epsilon_M}_{M,k,2^d}(\mathbf{x})| &= \left| \|\mathbf{x} - 2^{\frac{1}{q}} \cdot \mathbf{x}^*_{l,\frac{\epsilon_M}{3}}\|^q_\infty - \|2^{\frac{1}{q}} \cdot \mathbf{x}^*_{l,\frac{\epsilon_M}{3}}\|^q_\infty - \|\mathbf{x}\|^q_\infty \right| \\
&\le 2\epsilon_M^q + \frac{2 \cdot 2^q}{3^q}\epsilon_M^q \le 2(2^q + 2)\epsilon_M^q.
\end{aligned}$$

Outside of $\mathbb{B}\left(2^{\frac{1}{q}} \cdot \mathbf{x}^*_{l,\frac{\epsilon_M}{3}}, 2^{\frac{1}{q}} \cdot \frac{\epsilon_M}{3}\right)$, $f^{\epsilon_M}_{M,k,l}(\mathbf{x})$ is identical to $f^{\epsilon_M}_{M,k,2^d}(\mathbf{x})$.

The third item. For this part, we detail a proof for the case where $j < M$, $l < 2^d$ and $l \ne k$. The other cases are proved using similar arguments.

Case I: $j < M$**,** $l < 2^d$**, and** $l \ne k$**.** When $\mathbf{x} \notin S_l^{2^{\frac{1}{q}}\epsilon_j}$, it holds that

$$\begin{aligned}
&f^{\epsilon_j}_{j,k,l}(\mathbf{x}) - f^{\epsilon_j}_{j,k,l}(2^{\frac{1}{q}} \cdot \mathbf{x}^*_{l,\epsilon_j}) = f^{\epsilon_j}_{j,k,l}(\mathbf{x}) + \|2^{\frac{1}{q}} \cdot \mathbf{x}^*_{l,\epsilon_j}\|^q_\infty \\
&\ge \min \begin{cases} 2\|\mathbf{x}^*_{l,\epsilon_j}\|^q_\infty - \|\mathbf{x}^*_{k,\epsilon_j}\|^q_\infty, & \text{if } \mathbf{x} \in \mathbb{B}\left(\mathbf{x}^*_{k,\epsilon_j}, \epsilon_j\right) \backslash \mathbb{B}\left(0, \frac{\epsilon_j}{2}\right) \\ 2\|\mathbf{x}^*_{l,\epsilon_j}\|^q_\infty - \|\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|^q_\infty, & \text{if } \mathbf{x} \in \mathbb{B}\left(\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}, \frac{\epsilon_M}{3}\right) \backslash \mathbb{B}\left(0, \frac{\epsilon_M}{6}\right) \\ 2\|\mathbf{x}^*_{l,\epsilon_j}\|^q_\infty, & \text{if } \mathbf{x} \text{ is in other parts of } \mathbb{R}^d \backslash \mathbb{B}\left(2^{\frac{1}{q}} \cdot \mathbf{x}^*_{l,\epsilon_j}, 2^{\frac{1}{q}}\epsilon_j\right). \end{cases} \\
&\ge \epsilon_j^q \ge \frac{\epsilon_j^q}{3^q}.
\end{aligned}$$

Case II: $j = M$ **and** $l < 2^d$**.** Recall that the instance does not depend on k in this case. When $\mathbf{x} \notin S_l^{2^{\frac{1}{q}}\epsilon_j}$, it holds that

$$\begin{aligned}
&f^{\epsilon_M}_{M,k,l}(\mathbf{x}) - f^{\epsilon_M}_{M,k,l}(2^{\frac{1}{q}} \cdot \mathbf{x}^*_{l,\frac{\epsilon_M}{3}}) = f^{\epsilon_M}_{M,k,l}(\mathbf{x}) + \|2^{\frac{1}{q}} \cdot \mathbf{x}^*_{l,\frac{\epsilon_M}{3}}\|^q_\infty \\
&\ge \min\left\{2\|\mathbf{x}^*_{l,\frac{\epsilon_M}{3}}\|^q_\infty - \|\mathbf{x}^*_{2^d,\frac{\epsilon_M}{3}}\|^q_\infty, 2\|\mathbf{x}^*_{l,\frac{\epsilon_M}{3}}\|^q_\infty\right\} \ge \frac{\epsilon_j^q}{3^q}.
\end{aligned}$$

Case III: $j < M$**,** $l = 2^d$**, (and** $k < 2^d$**).** For this case, when $\mathbf{x} \notin S_{2^d}^{2^{\frac{1}{q}}\epsilon_j}$, it holds that

$$f_{j,k,2^d}^{\epsilon_j}(\mathbf{x}) - f_{j,k,2^d}^{\epsilon_j}(2^{\frac{1}{q}} \cdot \mathbf{x}_{2^d,\epsilon_j}^*) = f_{j,k,2^d}^{\epsilon_j}(\mathbf{x}) + \|2^{\frac{1}{q}} \cdot \mathbf{x}_{2^d,\epsilon_j}^*\|_\infty^q$$
$$\geq \min\left\{2\|\mathbf{x}_{2^d,\epsilon_j}^*\|_\infty^q - \|\mathbf{x}_{k,\epsilon_j}^*\|_\infty^q, 2\|\mathbf{x}_{2^d,\epsilon_j}^*\|_\infty^q\right\} \geq \epsilon_j^q \geq \frac{\epsilon_j^q}{3^q}.$$

There are some other cases. They are **Case IV:** $j < M$**,** $l < 2^d$**, and** $k = l$**;** and **Case V:** $j = M$**,** $l = 2^d$**, (and** $k < 2^d$**).** The proof for Cases IV-V uses the same argument as that for the previous cases. Now we combine all cases to conclude the proof.

□

4.2 The information-theoretical argument

First of all, we state below a classic result of Bretagnolle and Huber (Bretagnolle and Huber, 1978); See (e.g., Lattimore and Szepesvári, 2020) for a modern reference.

Lemma 3 (Bretagnolle–Huber)**.** *For two distributions P, Q over the same probability space, it holds that*

$$D_{TV}(P, Q) \leq \sqrt{1 - e^{-D_{kl}(P\|Q)}} \leq 1 - \frac{1}{2}\exp\left(-D_{kl}(P \parallel Q)\right).$$

The proof consists of two major steps. In the first step, we prove that for any policy π, there exists a long batch with high chance. In the second step, on the basis of existence of a long batch, we prove that there exists a bitten-apple instance (defined in Section 4.1) on which no policy performs better the lower bound in Theorem 3. Next we focus on proving the first step.

For a policy π that communicates at $t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_M$, we consider a set of events

$$A_j := \{t_{j-1} < T_{j-1} \text{ and } t_j \geq T_j\}, \tag{12}$$

where T_j is the reference communication point defined in (9). Whenever the event A_j is true, the j-th batch is large. Next we prove that some of A_j occurs under some instances, thus proving the existence of a long batch. Before proceeding, we introduce the following notation for simplicity.

For any policy π, we define

$$p_j := \frac{1}{2^d - 1}\sum_{k=1}^{2^d-1} \mathbb{P}_{j,k}(A_j), \qquad j = 1, 2, \cdots, M. \tag{13}$$

where $\mathbb{P}_{j,k}(A_j)$ denotes the probability of the event A_j under the instance $I_{j,k}$ and policy π. Next in Lemma 4, we show that with constant chance, there is a long batch.

Lemma 4. *For any policy π that adaptively determines the communications points, it holds that $\sum_{j=1}^M p_j \geq \frac{7}{8}$, where p_j is defined in (13).*

Proof of Lemma 4. Fix an arbitrary policy π. For each t, let $\mathbb{P}_{j,k}^t$ (resp. $\mathbb{P}_{M,k}^t$) be the probability of $(\mathbf{x}_t, y_t)$ governed by running π in environment $f_{j,k}^{\epsilon_j}$ (resp. $f_{M,k}^{\epsilon_M}$), i.e. $\mathbb{P}_{j,k}^t = \mathbb{P}_{j,k}^t\left(\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \cdots, \mathbf{x}_{t_{j-1}}, y_{t_{j-1}}\right)$. The event A_j is determined by the observations up to time T_{j-1}, since communication point t_j is determined given the previous time grid $\{t_1, t_2, \cdots, t_{j-1}\}$ under a fixed policy π. To further illustrate this fact, we first notice that the event $A_j' := \{t_{j-1} < T_{j-1}\}$ is fully determined by observations up to T_{j-1}. If $t_{j-1} \geq T_{j-1}$, then the failure of A_j', thus the failure of A_j, is known by time T_{j-1}. If $t_{j-1} < T_{j-1}$, then based on observations up to time $t_{j-1} < T_{j-1}$, the policy π determines t_j, thus A_j. In both cases, the success of A_j is fully determined by observations up to time T_{j-1}. It is also worth emphasizing that the policy π does not communicate at $\{T_j\}_{j\in[M]}$. We use $\{T_j\}_{j\in[M]}$ only as a reference. With the above argument, we get

$$|\mathbb{P}_{M,k}(A_j) - \mathbb{P}_{j,k}(A_j)| = |\mathbb{P}_{M,k}^{T_{j-1}}(A_j) - \mathbb{P}_{j,k}^{T_{j-1}}(A_j)| \leq D_{TV}\left(\mathbb{P}_{M,k}^{T_{j-1}}, \mathbb{P}_{j,k}^{T_{j-1}}\right). \tag{14}$$

By Lemma 3,

$$\frac{1}{2^d-1}\sum_{k=1}^{2^d-1} D_{TV}\left(\mathbb{P}_{M,k}^{T_{j-1}}, \mathbb{P}_{j,k}^{T_{j-1}}\right) \leq \frac{1}{2^d-1}\sum_{k=1}^{2^d-1}\sqrt{1-\exp\left(-D_{kl}\left(\mathbb{P}_{M,k}^{T_{j-1}}\|\mathbb{P}_{j,k}^{T_{j-1}}\right)\right)}. \tag{15}$$

Note that $f_{j,k}^{\epsilon_j}$ differs from $f_{M,k}^{\epsilon_M}$ only in $\mathbb{B}(\mathbf{x}^*_{k,\epsilon_j}, \epsilon_j)\backslash\mathbb{B}(0, \frac{\epsilon_j}{2})$. Hence the chain rule for KL-divergence gives, for any $t \in [T_{j-1}, T_j)$,

$$\begin{aligned}
& D_{kl}\left(\mathbb{P}_{M,k}^t\|\mathbb{P}_{j,k}^t\right) \\
={}& D_{kl}\left(\mathbb{P}_{M,k}^t\left(\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \cdots, \mathbf{x}_{T_{j-1}}, y_{T_{j-1}}\right)\|\mathbb{P}_{j,k}^t\left(\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \cdots, \mathbf{x}_{T_{j-1}}, y_{T_{j-1}}\right)\right) \\
={}& D_{kl}\left(\mathbb{P}_{M,k}^t\left(\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \cdots, \mathbf{x}_{T_{j-1}-1}, y_{T_{j-1}-1}\right)\|\mathbb{P}_{j,k}^t\left(\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \cdots, \mathbf{x}_{T_{j-1}-1}, y_{T_{j-1}-1}\right)\right) \\
& + \mathbb{E}_{\mathbb{P}_{M,k}^t}\left[D_{kl}\left(\mathcal{N}\left(f_{M,k}^{\epsilon_M}(\mathbf{x}_{T_{j-1}}), 1\right)\|\mathcal{N}\left(f_{j,k}^{\epsilon_j}(\mathbf{x}_{T_{j-1}}), 1\right)\right)\right] \\
& + D_{kl}\left(\mathbb{P}_{M,k}^t\left(\mathbf{x}_{T_{j-1}}|\mathbf{x}_1, y_1, \cdots, \mathbf{x}_{T_{j-1}-1}, y_{T_{j-1}-1}\right)\|\mathbb{P}_{j,k}^t\left(\mathbf{x}_{T_{j-1}}|\mathbf{x}_1, y_1, \cdots, \mathbf{x}_{T_{j-1}-1}, y_{T_{j-1}-1}\right)\right)
\end{aligned} \tag{16}$$

where $\mathcal{N}(\mu, 1)$ is the Gaussian random variable of mean μ and variance 1. Under the fixed policy π, $\mathbf{x}_{T_{j-1}}$ is fully determined by choices and observations before it. Thus

$$D_{kl}\left(\mathbb{P}_{M,k}^t\left(\mathbf{x}_{T_{j-1}}|\mathbf{x}_1, y_1, \cdots, \mathbf{x}_{T_{j-1}-1}, y_{T_{j-1}-1}\right)\|\mathbb{P}_{j,k}^t\left(\mathbf{x}_{T_{j-1}}|\mathbf{x}_1, y_1, \cdots, \mathbf{x}_{T_{j-1}-1}, y_{T_{j-1}-1}\right)\right) = 0.$$

By Proposition 5,

$$\begin{aligned}
D_{kl}\left(\mathcal{N}\left(f_{M,k}^{\epsilon_M}(\mathbf{x}_{T_{j-1}}), 1\right)\|\mathcal{N}\left(f_{j,k}^{\epsilon_j}(\mathbf{x}_{T_{j-1}}), 1\right)\right) &= \frac{1}{2}\left(f_{M,k}^{\epsilon_M}(\mathbf{x}_{T_{j-1}}) - f_{j,k}^{\epsilon_j}(\mathbf{x}_{T_{j-1}})\right)^2 \\
&\leq \frac{(2^q+2)^2}{2}\epsilon_j^{2q}\mathbb{I}_{\{\mathbf{x}_{T_{j-1}}\in S_k^{\epsilon_j}\}}.
\end{aligned}$$

We plug the above results into (16) and get, for any $k \geq 2$,

$$\begin{aligned}
& D_{kl}\left(\mathbb{P}_{M,k}^t\|\mathbb{P}_{j,k}^t\right) \\
={}& D_{kl}\left(\mathbb{P}_{M,k}^t\left(\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \cdots, \mathbf{x}_{T_{j-1}-1}, y_{T_{j-1}-1}\right)\|\mathbb{P}_{j,k}^t\left(\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \cdots, \mathbf{x}_{T_{j-1}-1}, y_{T_{j-1}-1}\right)\right) \\
& + \mathbb{E}_{\mathbb{P}_{M,k}^t}\left[\frac{1}{2}\left(f_{M,k}^{\epsilon_M}(\mathbf{x}_{T_{j-1}}) - f_{j,k}^{\epsilon_j}(\mathbf{x}_{T_{j-1}})\right)^2\right] \\
\leq{}& D_{kl}\left(\mathbb{P}_{M,k}^t\left(\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \cdots, \mathbf{x}_{T_{j-1}-1}, y_{T_{j-1}-1}\right)\|\mathbb{P}_{j,k}^t\left(\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \cdots, \mathbf{x}_{T_{j-1}-1}, y_{T_{j-1}-1}\right)\right) \\
& + \frac{(2^q+2)^2}{2}\mathbb{E}_{\mathbb{P}_{M,k}^t}\left[\epsilon_j^{2q}\mathbb{I}_{\{\mathbf{x}_{T_{j-1}}\in S_k^{\epsilon_j}\}}\right] \\
={}& D_{kl}\left(\mathbb{P}_{M,k}^t\left(\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \cdots, \mathbf{x}_{T_{j-1}-1}, y_{T_{j-1}-1}\right)\|\mathbb{P}_{j,k}^t\left(\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \cdots, \mathbf{x}_{T_{j-1}-1}, y_{T_{j-1}-1}\right)\right) \\
& + \frac{(2^q+2)^2\epsilon_j^{2q}}{2}\mathbb{P}_{M,k}^t\left(\mathbf{x}_{T_{j-1}} \in S_k^{\epsilon_j}\right).
\end{aligned}$$

We can then recursively apply chain rule and the above calculation, and obtain

$$D_{kl}\left(\mathbb{P}_{M,k}^t\|\mathbb{P}_{j,k}^t\right) \leq \frac{(2^q+2)^2\epsilon_j^{2q}}{2}\sum_{s\leq T_{j-1}}\mathbb{P}_{M,k}^t\left(\mathbf{x}_s \in S_k^{\epsilon_j}\right)$$

for each $t : T_{j-1} \le t < T_j$. Therefore, we have

$$D_{kl}\left(\mathbb{P}_{M,k}^{T_{j-1}} \| \mathbb{P}_{j,k}^{T_{j-1}}\right) \le \frac{(2^q+2)^2\epsilon_j^{2q}}{2} \sum_{s \le T_{j-1}} \mathbb{P}_{M,k}^{T_{j-1}}\left(\mathbf{x}_s \in S_k^{\epsilon_j}\right), \tag{17}$$

Combining the above inequalities (15) and (17) yields that

$$\begin{aligned}
\frac{1}{2^d-1}\sum_{k=1}^{2^d-1} D_{TV}\left(\mathbb{P}_{M,k}^{T_{j-1}}, \mathbb{P}_{j,k}^{T_{j-1}}\right) &\le \frac{1}{2^d-1}\sum_{k=1}^{2^d-1}\sqrt{1-\exp\left(-D_{kl}\left(\mathbb{P}_{M,k}^{T_{j-1}} \| \mathbb{P}_{j,k}^{T_{j-1}}\right)\right)} \\
&\le \frac{1}{2^d-1}\sum_{k=1}^{2^d-1}\sqrt{1-\exp\left(-\frac{(2^q+2)^2\epsilon_j^{2q}}{2}\sum_{s\le T_{j-1}} \mathbb{P}_{M,k}^{T_{j-1}}\left(\mathbf{x}_s \in S_k^{\epsilon_j}\right)\right)} \\
&\le \sqrt{1-\exp\left(-\frac{(2^q+2)^2\epsilon_j^{2q}}{2(2^d-1)}\sum_{k=1}^{2^d-1}\sum_{s\le T_{j-1}} \mathbb{P}_{M,k}^{T_{j-1}}\left(\mathbf{x}_s \in S_k^{\epsilon_j}\right)\right)}, \qquad (18)
\end{aligned}$$

where the last inequality follows from Jensen. Since $\sum_{k=1}^{2^d-1}\mathbb{P}_{M,k}^{T_j-1}\left(\mathbf{x}_s \in S_k^{\epsilon_j}\right) \le 1$ ($S_k^{\epsilon_j}$ are disjoint), we continue from (18) and get

$$\begin{aligned}
&\sqrt{1-\exp\left(-\frac{(2^q+2)^2\epsilon_j^{2q}}{2(2^d-1)}\sum_{k=1}^{2^d-1}\sum_{s\le T_{j-1}} \mathbb{P}_{M,k}^{T_{j-1}}\left(\mathbf{x}_s \in S_k^{\epsilon_j}\right)\right)} \\
\le &\sqrt{1-\exp\left(-\frac{(2^q+2)^2\epsilon_j^{2q}T_{j-1}}{2(2^d-1)}\right)} \overset{(i)}{\le} \sqrt{1-\exp\left(-\frac{1}{64}\cdot\frac{1}{M^2}\right)} \overset{(ii)}{\le} \frac{1}{8}\cdot\frac{1}{M}, \qquad (19)
\end{aligned}$$

where *(i)* uses definitions of ϵ_j and T_j (9), *(ii)* uses a basic property of the exponential function: $\exp(-x) \ge 1-x$ for each $x \in \mathbb{R}$. Combining (14) and (19) gives that, for each $j = 1, 2, \cdots, M$,

$$|\mathbb{P}_{M,k}(A_j) - p_j| \le \frac{1}{2^d-1}\sum_{k=1}^{2^d-1}|\mathbb{P}_{M,k}(A_j) - \mathbb{P}_{j,k}(A_j)| \le \frac{1}{8M},$$

and thus

$$\sum_{j=1}^{M} p_j \ge \sum_{j=1}^{M}\mathbb{P}_{M,k}(A_j) - \frac{1}{8} \ge \mathbb{P}_{M,k}(\cup_{j=1}^M A_j) - \frac{1}{8} \ge \frac{7}{8},$$

where the last inequality holds since at least one of $\{A_1, A_2, \cdots, A_M\}$ must be true.

□

Now that Lemma 4 is in place, we can prove the existence of a bad bitten-apple instance, which concludes the proof of Theorem 3.

Proof of Theorem 3. Fix any policy π. Let $\mathbb{P}_{j,k,l}$ be the probability of running π on $f_{j,k,l}^{\epsilon_j}$. Let $\mathbb{P}_{j,k,l}^t$ be the probability of $(\mathbf{x}_1, y_1, \mathbf{x}_2, y_2, \cdots, \mathbf{x}_t, y_t)$ governed by running π in environment $f_{j,k,l}^{\epsilon_j}$, Proposition

6 gives that

$$\sup_{I\in\{I_{j,k,l}\}_{j\in[M],k<2^d,l\in[2^d]}} \mathbb{E}\left[R^\pi(T)\right] \geq \frac{1}{M}\sum_{j=1}^{M}\frac{\epsilon_j^q}{3^q}\sum_{t=1}^{T}\frac{1}{2^d-1}\cdot\frac{1}{2^d}\sum_{k=1}^{2^d-1}\sum_{l=1}^{2^d}\mathbb{P}_{j,k,l}^t\left(\mathbf{x}_t\notin S_l^{2^{\frac{1}{q}}\cdot\epsilon_j}\right)$$

$$= \frac{1}{3^q}\cdot\frac{1}{M}\sum_{j=1}^{M}\epsilon_j^q\sum_{t=1}^{T}\frac{1}{2^d-1}\cdot\frac{1}{2^d}\sum_{k=1}^{2^d-1}\sum_{l=1}^{2^d}\left(1-\mathbb{P}_{j,k,l}^t\left(\mathbf{x}_t\in S_l^{2^{\frac{1}{q}}\cdot\epsilon_j}\right)\right)$$

$$\geq \frac{1}{3^q}\cdot\frac{1}{M}\left[\sum_{j=1}^{M-1}\epsilon_j^q\sum_{t=1}^{T}\frac{1}{2^d-1}\cdot\frac{1}{2^d}\sum_{k=1}^{2^d-1}\sum_{l=1}^{2^d}\left(1-\mathbb{P}_{j,k,k}^t\left(\mathbf{x}_t\in S_l^{2^{\frac{1}{q}}\cdot\epsilon_j}\right)-D_{TV}\left(\mathbb{P}_{j,k,l}^t,\mathbb{P}_{j,k,k}^t\right)\right)\right.$$

$$\left.+\epsilon_M^q\sum_{t=1}^{T}\frac{1}{2^d-1}\cdot\frac{1}{2^d}\sum_{k=1}^{2^d-1}\sum_{l=1}^{2^d}\left(1-\mathbb{P}_{M,k,2^d}^t\left(\mathbf{x}_t\in S_l^{2^{\frac{1}{q}}\cdot\epsilon_M}\right)-D_{TV}\left(\mathbb{P}_{M,k,l}^t,\mathbb{P}_{M,k,2^d}^t\right)\right)\right], \tag{20}$$

where the last inequality follows from definition of total-variation distance

$$D_{TV}\left(\mathbb{P}_{j,k,l}^t,\mathbb{P}_{j,k,k}^t\right)\geq\mathbb{P}_{j,k,l}^t\left(\mathbf{x}_t\in S_l^{2^{\frac{1}{q}}\cdot\epsilon_j}\right)-\mathbb{P}_{j,k,k}^t\left(\mathbf{x}_t\in S_l^{2^{\frac{1}{q}}\cdot\epsilon_j}\right)$$

and

$$D_{TV}\left(\mathbb{P}_{M,k,l}^t,\mathbb{P}_{M,k,2^d}^t\right)\geq\mathbb{P}_{M,k,l}^t\left(\mathbf{x}_t\in S_l^{2^{\frac{1}{q}}\cdot\epsilon_M}\right)-\mathbb{P}_{M,k,2^d}^t\left(\mathbf{x}_t\in S_l^{2^{\frac{1}{q}}\cdot\epsilon_M}\right).$$

For the first term on the right side of 20, delete negative number $-\mathbb{P}_{j,k,k}^t(\cdot)$ and bring into the equation $D_{TV}(\mathbb{P},\mathbb{Q})=\frac{1}{2}\int|d\mathbb{P}-d\mathbb{Q}|$, we get

$$\epsilon_j^q\sum_{t=1}^{T}\frac{1}{2^d}\sum_{l=1}^{2^d}\left(1-\mathbb{P}_{j,k,k}^t\left(\mathbf{x}_t\in S_l^{2^{\frac{1}{q}}\cdot\epsilon_j}\right)-D_{TV}\left(\mathbb{P}_{j,k,l}^t,\mathbb{P}_{j,k,k}^t\right)\right)$$

$$\geq\epsilon_j^q\sum_{t=1}^{T}\frac{1}{2^d}\sum_{l\neq k}\left(1-\frac{1}{2}\int\left|d\mathbb{P}_{j,k,k}^t-d\mathbb{P}_{j,k,l}^t\right|\right)$$

$$\geq\epsilon_j^q\sum_{t=1}^{T_j}\frac{1}{2^d}\sum_{l\neq k}\left(1-\frac{1}{2}\int\left|d\mathbb{P}_{j,k,k}^t-d\mathbb{P}_{j,k,l}^t\right|\right)$$

$$\geq\epsilon_j^qT_j\frac{1}{2^d}\sum_{l\neq k}\left(1-\frac{1}{2}\int\left|d\mathbb{P}_{j,k,k}^{T_j}-d\mathbb{P}_{j,k,l}^{T_j}\right|\right) \tag{21}$$

$$=\epsilon_j^qT_j\frac{1}{2^d}\sum_{l\neq k}\frac{1}{2}\left(\int d\mathbb{P}_{j,k,k}^{T_j}+d\mathbb{P}_{j,k,l}^{T_j}-\left|d\mathbb{P}_{j,k,k}^{T_j}-d\mathbb{P}_{j,k,l}^{T_j}\right|\right)$$

$$\geq\epsilon_j^qT_j\frac{1}{2^d}\sum_{l\neq k}\frac{1}{2}\left(\int_{A_j} d\mathbb{P}_{j,k,k}^{T_j}+d\mathbb{P}_{j,k,l}^{T_j}-\left|d\mathbb{P}_{j,k,k}^{T_j}-d\mathbb{P}_{j,k,l}^{T_j}\right|\right)$$

$$=\epsilon_j^qT_j\frac{1}{2^d}\sum_{l\neq k}\frac{1}{2}\left(\int_{A_j} d\mathbb{P}_{j,k,k}^{T_{j-1}}+d\mathbb{P}_{j,k,l}^{T_{j-1}}-\left|d\mathbb{P}_{j,k,k}^{T_{j-1}}-d\mathbb{P}_{j,k,l}^{T_{j-1}}\right|\right), \tag{22}$$

where (21) follows for data processing inequality of total variation distance, and the last equation (22) holds because the observations at time T_j are the same as those at time T_{j-1} under event A_j and the fixed policy π. Further more, we have

$$
\begin{aligned}
&\frac{1}{2}\left(\int_{A_j} d\mathbb{P}_{j,k,k}^{T_{j-1}} + d\mathbb{P}_{j,k,l}^{T_{j-1}} - \left| d\mathbb{P}_{j,k,k}^{T_{j-1}} - d\mathbb{P}_{j,k,l}^{T_{j-1}} \right|\right) \\
&= \frac{\mathbb{P}_{j,k,k}^{T_{j-1}}(A_j) + \mathbb{P}_{j,k,l}^{T_{j-1}}(A_j)}{2} - \frac{1}{2}\int_{A_j} \left| d\mathbb{P}_{j,k,k}^{T_{j-1}} - d\mathbb{P}_{j,k,l}^{T_{j-1}} \right| \\
&\geq \left(\mathbb{P}_{j,k,k}^{T_{j-1}}(A_j) - \frac{1}{2} D_{TV}\left(\mathbb{P}_{j,k,k}^{T_{j-1}}, \mathbb{P}_{j,k,l}^{T_{j-1}}\right)\right) - D_{TV}\left(\mathbb{P}_{j,k,k}^{T_{j-1}}, \mathbb{P}_{j,k,l}^{T_{j-1}}\right) \qquad (23) \\
&= \mathbb{P}_{j,k}(A_j) - \frac{3}{2} D_{TV}\left(\mathbb{P}_{j,k,k}^{T_{j-1}}, \mathbb{P}_{j,k,l}^{T_{j-1}}\right), \qquad (24)
\end{aligned}
$$

where (23) follows from $|\mathbb{P}(A) - \mathbb{Q}(A)| \leq D_{TV}(\mathbb{P}, \mathbb{Q})$, and (24) is attributed to the fact that A_j is determined by the observations up to time T_{j-1}.
Similar to the argument for (17)-(19), we have, for each fixed k

$$
\begin{aligned}
\frac{1}{2^d}\sum_{l\neq k} D_{TV}\left(\mathbb{P}_{j,k,k}^{T_{j-1}}, \mathbb{P}_{j,k,l}^{T_{j-1}}\right) &\leq \frac{1}{2^d}\sum_{l\neq k}\sqrt{1 - \exp\left(-D_{kl}\left(\mathbb{P}_{j,k,k}^{T_{j-1}} \| \mathbb{P}_{j,k,l}^{T_{j-1}}\right)\right)} \\
&\leq \frac{1}{2^d}\sum_{l\neq k}\sqrt{1 - \exp\left(-\frac{(2^q+2)^2\left(2^{\frac{1}{q}}\cdot\epsilon_j\right)^{2q}}{2}\sum_{s\leq T_{j-1}} \mathbb{P}_{j,k,k}^{T_{j-1}}\left(\mathbf{x}_s \in S_l^{2^{\frac{1}{q}}\cdot\epsilon_j}\right)\right)} \\
&\leq \frac{2^d-1}{2^d}\sqrt{1 - \exp\left(-\frac{(2^q+2)^2\left(2^{\frac{1}{q}}\cdot\epsilon_j\right)^{2q}}{2(2^d-1)}\sum_{l\neq k}\sum_{s\leq T_{j-1}} \mathbb{P}_{j,k,k}^{T_{j-1}}\left(\mathbf{x}_s \in S_l^{2^{\frac{1}{q}}\cdot\epsilon_j}\right)\right)} \\
&= \frac{2^d-1}{2^d}\sqrt{1 - \exp\left(-\frac{(2^q+2)^2\left(2^{\frac{1}{q}}\cdot\epsilon_j\right)^{2q}}{2(2^d-1)}\sum_{s\leq T_{j-1}}\sum_{l\neq k} \mathbb{P}_{j,k,k}^{T_{j-1}}\left(\mathbf{x}_s \in S_l^{2^{\frac{1}{q}}\cdot\epsilon_j}\right)\right)} \\
&\leq \frac{2^d-1}{2^d}\sqrt{1 - \exp\left(-\frac{2(2^q+2)^2\epsilon_j^{2q}T_{j-1}}{2^d-1}\right)} \\
&\leq \frac{2^d-1}{2^d}\sqrt{1 - \exp\left(\frac{1}{16}\cdot\frac{1}{M^2}\right)} \\
&\leq \frac{1}{4}\cdot\frac{1}{M}\cdot\frac{2^d-1}{2^d}. \qquad (25)
\end{aligned}
$$

For the second term on the right side of (20), we have the same inequality by subtituting $\mathbb{P}_{j,k,l}^t$ (resp. $\mathbb{P}_{j,k,k}^t$) with $\mathbb{P}_{M,k,l}^t$ (resp. $\mathbb{P}_{M,k,2^d}^t$).

Combining (20), (22), (24) and (25), we have

$$
\begin{aligned}
&\sup_{I\in\{I_{j,k,l}\}_{j\in[M],k<2^d,l\in[2^d]}} \mathbb{E}\left[R^\pi(T)\right] \\
\geq\ & \frac{1}{3^q}\cdot\frac{1}{M}\left[\sum_{j=1}^{M-1}\epsilon_j^q T_j\frac{1}{2^d-1}\cdot\frac{1}{2^d}\sum_{k=1}^{2^d-1}\sum_{l\neq k}\left(\mathbb{P}_{j,k}(A_j)-\frac{3}{2}D_{TV}\left(\mathbb{P}_{j,k,k}^{T_{j-1}},\mathbb{P}_{j,k,l}^{T_{j-1}}\right)\right)\right. \\
&\left.+\epsilon_M^q T_M\frac{1}{2^d-1}\cdot\frac{1}{2^d}\sum_{k=1}^{2^d-1}\sum_{l\neq 2^d}\left(\mathbb{P}_{M,k}(A_M)-\frac{3}{2}D_{TV}\left(\mathbb{P}_{M,k,2^d}^{T_{M-1}},\mathbb{P}_{M,k,l}^{T_{M-1}}\right)\right)\right] \\
=\ & \frac{1}{3^q}\cdot\frac{1}{M}\left[\sum_{j=1}^{M-1}\epsilon_j^q T_j\frac{1}{2^d-1}\sum_{k=1}^{2^d-1}\left(\frac{1}{2^d}\sum_{l\neq k}\mathbb{P}_{j,k}(A_j)-\frac{3}{2}\cdot\frac{1}{2^d}\sum_{l\neq k}D_{TV}\left(\mathbb{P}_{j,k,k}^{T_{j-1}},\mathbb{P}_{j,k,l}^{T_{j-1}}\right)\right)\right. \\
&\left.+\epsilon_M^q T_M\frac{1}{2^d-1}\sum_{k=1}^{2^d-1}\left(\frac{1}{2^d}\sum_{l\neq 2^d}\mathbb{P}_{M,k}(A_M)-\frac{3}{2}\cdot\frac{1}{2^d}\sum_{l\neq 2^d}D_{TV}\left(\mathbb{P}_{M,k,2^d}^{T_{M-1}},\mathbb{P}_{M,k,l}^{T_{M-1}}\right)\right)\right] \\
\geq\ & \frac{1}{3^q}\cdot\frac{1}{M}\sum_{j=1}^{M}\epsilon_j^q T_j\left[\frac{1}{2^d-1}\sum_{k=1}^{2^d-1}\mathbb{P}_{j,k}\left(A_j\right)-\frac{3}{2}\cdot\frac{1}{4}\cdot\frac{1}{M}\right]\cdot\frac{2^d-1}{2^d} \\
=\ & \frac{1}{3^q}\cdot\frac{1}{M}\cdot\frac{2^d-1}{2^d}\sum_{j=1}^{M}\epsilon_j^q T_j\left(p_j-\frac{3}{8}\cdot\frac{1}{M}\right).
\end{aligned}
$$

By definition of ϵ_j and T_j in (9), we have $\epsilon_j^q T_j = \frac{\sqrt{2}}{8}\cdot\frac{\sqrt{2^d-1}}{2^q+2}\cdot\frac{1}{M}\cdot T^{\frac{1}{2}\cdot\frac{1}{1-2^{-M}}}$ for all $j\in[M]$. Therefore, we continue from the above inequalities and get

$$
\begin{aligned}
&\sup_{I\in\{I_{j,k,l}\}_{j\in[B],k<2^d,l\in[2^d]}} \mathbb{E}\left[R^\pi(T)\right] \\
\geq\ & \frac{1}{3^q}\cdot\frac{1}{M^2}\cdot\frac{\sqrt{2}}{8}\cdot\frac{1}{2^q+2}\cdot\frac{(2^d-1)^{\frac{3}{2}}}{2^d}\cdot T^{\frac{1}{2}\cdot\frac{1}{1-2^{-M}}}\left(\sum_{j=1}^{M}p_j-\frac{3}{8}\right) \\
\geq\ & \frac{\sqrt{2}}{16}\cdot\frac{1}{M^2}\cdot\frac{1}{3^q(2^q+2)}\cdot\frac{(2^d-1)^{\frac{3}{2}}}{2^d}\cdot T^{\frac{1}{2}\cdot\frac{1}{1-2^{-M}}}
\end{aligned}
$$

where the last inequality uses Lemma 4.

□

Proof of Corollary 2. From Theorem 3, the expected regret is lower bounded by

$$\mathbb{E}\left[R_T(\pi)\right]\geq\frac{\sqrt{2}}{16}\cdot\frac{1}{M^2}\cdot\frac{1}{3^q(2^q+2)}\cdot\frac{(2^d-1)^{\frac{3}{2}}}{2^d}\cdot T^{\frac{1}{2}\cdot\frac{1}{1-2^{-M}}}.$$

Here we seek for the minimum M such that

$$\frac{\frac{1}{M^2}\cdot T^{\frac{1}{2}\cdot\frac{1}{1-2^{-M}}}}{\sqrt{T}}\leq e. \tag{26}$$

Calculation shows that

$$\frac{\frac{1}{M^2}\cdot T^{\frac{1}{2}\cdot\frac{1}{1-2^{-M}}}}{\sqrt{T}}=\frac{1}{M^2}T^{\frac{1}{2}\cdot\frac{1}{2^M-1}}. \tag{27}$$

Substituting (27) to (26) and taking log on both sides yield that

$$\frac{1}{2} \cdot \frac{1}{2^M - 1} \log T \leq \log(M^2 e)$$

and thus

$$M \geq \log_2 \left(1 + \frac{\log T}{2 \log(M^2 e)} \right). \tag{28}$$

We use $M_{\min}$ to denote the minimum M such that inequality (28) holds. Calculation shows that (28) holds for

$$M_* := \log_2 \left(1 + \frac{\log T}{2} \right),$$

so we have $M_{\min} \leq M_*$. Then since the RHS of (28) decreases with M, we have

$$M_{\min} \geq \log_2 \left(1 + \frac{\log T}{2 \log(M_{\min}^2 e)} \right) \geq \log_2 \left(1 + \frac{\log T}{2 \log(M_*^2 e)} \right).$$

Therefore, $\Omega(\log \log T)$ rounds of communications are necessary for any algorithm to achieve a regret rate of order $K_- A_-^d \sqrt{T}$, where K_- depends only on q and A_- is an absolute constant.

□

4.3 Lower bound for nondegenerate bandits without communication constraints

Having established the lower bound with communication constraints in the previous section, it is worth noting that the existing literature lacks a standard lower bound result specifically tailored for nondegenerate bandits. To this end, we proceed to fill this gap by presenting a lower bound that does not incorporate any communication constraints.

To prove this result, we need a different set of problem instances, which we introduce now. For any fixed ϵ, we partition the space $\mathbb{R}^d$ again into 2^d disjoint parts $U_1^\epsilon, U_2^\epsilon, \cdots, U_{2^d}^\epsilon$. For $k = 1$, we define $U_1^\epsilon = O_1 \cup \mathbb{B}(0, \frac{\epsilon}{2})$. For $k = 2, \cdots, 2^d$, we define $U_k^\epsilon = O_k \backslash \mathbb{B}\left(0, \frac{\epsilon}{2}\right)$.

For any $k = 2, \cdots, 2^d$, and $\epsilon > 0$, define

$$f_k^\epsilon(\mathbf{x}) = \begin{cases} \|\mathbf{x} - \mathbf{x}_{k,\epsilon}^*\|_\infty^q - \|\mathbf{x}_{k,\epsilon}^*\|_\infty^q, & \text{if } \mathbf{x} \in \mathbb{B}(\mathbf{x}_{k,\epsilon}^*, \epsilon) \backslash \mathbb{B}(0, \frac{\epsilon}{2}), \\ \|\mathbf{x}\|_\infty^q, & \text{otherwise.} \end{cases} \tag{29}$$

In addition, we define the function f_1^ϵ as

$$f_1^\epsilon(\mathbf{x}) = \|\mathbf{x}\|_\infty^q, \tag{30}$$

and slightly overload the notations to define $\mathbf{x}_{1,\epsilon}^* := 0$. Note that $f_1^\epsilon(\mathbf{x})$ and $\mathbf{x}_{1,\epsilon}^*$ do not depend on ϵ. We keep the ϵ superscript for notational consistency.

Firstly, we observe that instances specified by $\{f_k^\epsilon\}_{k \in [2^d]}$ satisfy the properties stated in Proposition 7.

Proposition 7. *The functions f_k^ϵ satisfies*

1. *For each $k = 1, 2, \cdots, 2^d$, $\frac{1}{2^{q-1}} \|\mathbf{x} - \mathbf{x}_{k,\epsilon}^*\|_\infty^q \leq f_k^\epsilon(\mathbf{x}) - f_k^\epsilon(\mathbf{x}_{k,\epsilon}^*)$, for all $\mathbf{x} \in \mathbb{R}^d$.*

2. *For each $k = 2, 3, \cdots, 2^d$,*

$$\begin{cases} |f_k^\epsilon(\mathbf{x}) - f_1^\epsilon(\mathbf{x})| \leq (2^q + 2)\epsilon^q, & \forall \mathbf{x} \in U_k^\epsilon, \\ |f_k^\epsilon(\mathbf{x}) - f_1^\epsilon(\mathbf{x})| = 0, & \forall \mathbf{x} \notin U_k^\epsilon. \end{cases}$$

3. For each $k = 1, 2, \cdots, 2^d$, $f_k^\epsilon(\mathbf{x}) - f_k^\epsilon(\mathbf{x}_{k,\epsilon}^*) \le 3^{q+1}\|\mathbf{x} - \mathbf{x}_{k,\epsilon}^*\|_\infty^q$, *for all* $\mathbf{x} \in \mathbb{R}^d$.

Proof. Item 1 is clearly true when $\mathbf{x} \in U_k^\epsilon$, it remains to consider $\mathbf{x} \notin U_k^\epsilon$. For item 1, we use Jensen's inequality to get

$$\left\| \frac{\mathbf{x} - \mathbf{y}}{2} \right\|_\infty^q \le \frac{\|\mathbf{x}\|_\infty^q + \|\mathbf{y}\|_\infty^q}{2}, \quad \forall q \ge 1, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d.$$

Rearranging terms, and substituting $\mathbf{y} = \mathbf{x}_{k,\epsilon}^*$ in the above inequality gives that, for any $\mathbf{x} \notin U_k^\epsilon$,

$$\frac{1}{2^{q-1}} \|\mathbf{x} - \mathbf{x}_{k,\epsilon}^*\|_\infty^q \le \|\mathbf{x}\|_\infty^q + \|\mathbf{x}_{k,\epsilon}^*\|^q = f(\mathbf{x}) - f(\mathbf{x}_{k,\epsilon}^*).$$

For item 2, we have, for each k and $\mathbf{x} \in \mathbb{B}(\mathbf{x}_{k,\epsilon}^*, \epsilon) \backslash \mathbb{B}(0, \frac{\epsilon}{2})$,

$$\begin{aligned} &|f_k^\epsilon(\mathbf{x}) - f_1^\epsilon(\mathbf{x})| = |\|\mathbf{x} - \mathbf{x}^*\|_\infty^q - \|\mathbf{x}^*\|_\infty^q - \|\mathbf{x}\|_\infty^q| \\ \le& \epsilon^q + \epsilon^q + (2\epsilon)^q = (2^q + 2)\epsilon^q \end{aligned}$$

where the last inequality uses $\|\mathbf{x}\|_\infty \le 2\epsilon$ for all $\mathbf{x} \in \mathbb{B}(\mathbf{x}_{k,\epsilon}^*, \epsilon)$.

Next we proof item 3. Fix any $r \in (\frac{\epsilon}{2}, \infty)$. For any $\mathbf{x} \in \mathbb{S}(\mathbf{x}_{k,\epsilon}^*, r)$, we have $\|\mathbf{x}\|_\infty \le r + \epsilon$, and thus

$$3^q \|\mathbf{x} - \mathbf{x}_{k,\epsilon}^*\|_\infty^q = 3^q r^q \ge (r + \epsilon)^q \ge \|\mathbf{x}\|_\infty^q.$$

The above inequality gives,

$$3^{q+1}\|\mathbf{x} - \mathbf{x}_{k,\epsilon}^*\|_\infty^q \ge (2^q + 3^q)\|\mathbf{x} - \mathbf{x}_{k,\epsilon}^*\|_\infty^q \ge \|\mathbf{x}\|_\infty^q + \|\mathbf{x}_{k,\epsilon}^*\|_\infty^q = f(\mathbf{x}) - f(\mathbf{x}^*), \quad \forall \mathbf{x} \notin \mathbb{B}(\mathbf{x}_{k,\epsilon}^*, \epsilon) \backslash \mathbb{B}(0, \frac{\epsilon}{2}).$$

We conclude the proof by noticing that item 3 is clearly true when $\mathbf{x} \in \mathbb{B}(\mathbf{x}_{k,\epsilon}^*, \epsilon) \backslash \mathbb{B}(0, \frac{\epsilon}{2})$. □

Proof of Theorem 2. Fix any policy π. Let $\mathbb{P}_{k,\epsilon}$ be the probability of running π on f_k^ϵ. Let $\mathbb{E}_{k,\epsilon}$ be the expectation with respect to $\mathbb{P}_{k,\epsilon}$.

Firstly, we note that $\{\mathbf{x}_t \notin U_k^\epsilon\} \implies \{f_k^\epsilon(\mathbf{x}_t) - f_k^\epsilon(\mathbf{x}_k^*) \ge 2^{-2q+1}\epsilon^q\}$. Thus we have

$$\begin{aligned} \frac{1}{2^d} \sum_{k=1}^{2^d} \mathbb{E}_{k,\epsilon}\left[R_T(\pi)\right] &\ge \frac{1}{2^d} \sum_{k=1}^{2^d} \sum_{t=1}^{T} \mathbb{E}_{k,\epsilon}^t \left[f_k^\epsilon(\mathbf{x}_t) - f_k^\epsilon(\mathbf{x}_{k,\epsilon}^*) \right] \\ &\ge \frac{2^{-2q+1}\epsilon^q}{2^d} \sum_{k=1}^{2^d} \sum_{t=1}^{T} \mathbb{P}_{k,\epsilon}\left(f_k^\epsilon(\mathbf{x}_t) - f_k^\epsilon(\mathbf{x}_{k,\epsilon}^*) \ge 2^{-2q+1}\epsilon^q \right) \\ &\ge \frac{2^{-2q+1}\epsilon^q}{2^d} \sum_{k=1}^{2^d} \sum_{t=1}^{T} \mathbb{P}_{k,\epsilon}\left(\mathbf{x}_t \notin U_k^\epsilon \right). \end{aligned} \tag{31}$$

We continue the above derivation, and obtain

$$
\begin{aligned}
&\frac{2^{-2q+1}\epsilon^q}{2^d}\sum_{k=1}^{2^d}\sum_{t=1}^{T}\mathbb{P}_{k,\epsilon}\left(\mathbf{x}_t\notin U_k^\epsilon\right)\\
&\geq 2^{-2q+1}\epsilon^q\frac{1}{2^d}\sum_{k=1}^{2^d}\sum_{t=1}^{T}\left(1-\mathbb{P}_{k,\epsilon}\left(\mathbf{x}_t\in U_k^\epsilon\right)\right)\\
&\geq 2^{-2q+1}\epsilon^q\frac{1}{2^d}\sum_{k=1}^{2^d}\sum_{t=1}^{T}\left(1-\mathbb{P}_{1,\epsilon}^t\left(\mathbf{x}_t\in U_k^\epsilon\right)-D_{TV}\left(\mathbb{P}_{1,\epsilon},\mathbb{P}_{k,\epsilon}\right)\right)\\
&= 2^{-2q+1}\epsilon^q\left(1-\frac{1}{2^d}\right)T-2^{-2q+1}\epsilon^q\frac{1}{2^d}\sum_{k=2}^{2^d}\sum_{t=1}^{T}D_{TV}\left(\mathbb{P}_{1,\epsilon},\mathbb{P}_{k,\epsilon}\right)\\
&\geq 2^{-2q+1}\epsilon^q\left(1-\frac{1}{2^d}\right)T-2^{-2q+1}\epsilon^q\frac{1}{2^d}\sum_{k=2}^{2^d}\sum_{t=1}^{T}\left(1-\frac{1}{2}\exp\left(-D_{kl}\left(\mathbb{P}_{1,\epsilon}^t\|\mathbb{P}_{k,\epsilon}\right)\right)\right)\\
&= 2^{-2q}\frac{\epsilon^q}{2^d}\sum_{k=2}^{2^d}\sum_{t=1}^{T}\exp\left(-D_{kl}\left(\mathbb{P}_{1,\epsilon}^t\|\mathbb{P}_{k,\epsilon}\right)\right)\\
&\geq 2^{-2q}\frac{2^d-1}{2^d}\sum_{t=1}^{T}\epsilon^q\exp\left(-\frac{1}{2^d-1}\sum_{k=2}^{2^d}D_{kl}\left(\mathbb{P}_{1,\epsilon}\|\mathbb{P}_{k,\epsilon}^t\right)\right),
\end{aligned}
\tag{32}
$$

where the fourth line uses $\sum_{k=1}^{2^d}\mathbb{P}_{1,\epsilon}\left(\mathbf{x}_t\in U_k^\epsilon\right)=1$, the fifth line uses Lemma 3, and the last line uses Jensen's inequality.

By the chain rule of KL-divergence, we have

$$
\begin{aligned}
&D_{kl}\left(\mathbb{P}_{1,\epsilon}\|\mathbb{P}_{k,\epsilon}\right)\\
&= D_{kl}\left(\mathbb{P}_{1,\epsilon}\left(\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_T,y_T\right)\|\mathbb{P}_{k,\epsilon}\left(\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_T,y_T\right)\right)\\
&= D_{kl}\left(\mathbb{P}_{1,\epsilon}\left(\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_{T-1},y_{T-1}\right)\|\mathbb{P}_{k,\epsilon}\left(\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_{T-1},y_{T-1}\right)\right)\\
&\quad+\mathbb{E}_{\mathbb{P}_{1,\epsilon}}\left[D_{kl}\left(\mathcal{N}\left(f_1^\epsilon(\mathbf{x}_T),1\right)\|\mathcal{N}\left(f_k^\epsilon(\mathbf{x}_T),1\right)\right)\right]\\
&\quad+D_{kl}\left(\mathbb{P}_{1,\epsilon}\left(\mathbf{x}_T|\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_{T-1},y_{T-1}\right)\|\mathbb{P}_{k,\epsilon}\left(\mathbf{x}_T|\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_{T-1},y_{T-1}\right)\right)
\end{aligned}
\tag{33}
$$

where $\mathcal{N}(\mu,1)$ is the Gaussian random variable of mean μ and variance 1. Under the fixed policy π, $\mathbf{x}_T$ is fully determined by choices and observations before it. Thus

$$
D_{kl}\left(\mathbb{P}_{1,\epsilon}\left(\mathbf{x}_T|\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_{T-1},y_{T-1}\right)\|\mathbb{P}_{k,\epsilon}\left(\mathbf{x}_T|\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_{T-1},y_{T-1}\right)\right)=0.
$$

Also, $D_{kl}\left(\mathcal{N}\left(f_1^\epsilon(\mathbf{x}_T),1\right)\|\mathcal{N}\left(f_k^\epsilon(\mathbf{x}_T),1\right)\right)=\frac{1}{2}\left(f_1^\epsilon(\mathbf{x}_T)-f_k^\epsilon(\mathbf{x}_T)\right)^2$. We plug the above results into (33) and get, for any $k\geq 2$,

$$
\begin{aligned}
D_{kl}\left(\mathbb{P}_{1,\epsilon}\|\mathbb{P}_{k,\epsilon}\right) &= D_{kl}\left(\mathbb{P}_{1,\epsilon}\left(\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_{T-1},y_{T-1}\right)\|\mathbb{P}_{k,\epsilon}\left(\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_{T-1},y_{T-1}\right)\right)\\
&\quad+\mathbb{E}_{\mathbb{P}_{1,\epsilon}^t}\left[\frac{1}{2}\left(f_1^\epsilon(\mathbf{x}_T)-f_k^\epsilon(\mathbf{x}_T)\right)^2\right]\\
&\leq D_{kl}\left(\mathbb{P}_{1,\epsilon}\left(\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_{T-1},y_{T-1}\right)\|\mathbb{P}_{k,\epsilon}\left(\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_{T-1},y_{T-1}\right)\right)\\
&\quad+\frac{(2^q+2)^2}{2}\mathbb{E}_{\mathbb{P}_{1,\epsilon}}\left[\epsilon^{2q}\mathbb{I}_{\left\{\mathbf{x}_T\in U_k^\epsilon\right\}}\right]\\
&= D_{kl}\left(\mathbb{P}_{1,\epsilon}\left(\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_{T-1},y_{T-1}\right)\|\mathbb{P}_{k,\epsilon}\left(\mathbf{x}_1,y_1,\mathbf{x}_2,y_2,\cdots,\mathbf{x}_{T-1},y_{T-1}\right)\right)\\
&\quad+\frac{(2^q+2)^2\epsilon^{2q}}{2}\mathbb{P}_{1,\epsilon}\left(\mathbf{x}_T\in U_k^\epsilon\right).
\end{aligned}
$$

We can then recursively apply chain rule and the above calculation, and obtain

$$D_{kl}\left(\mathbb{P}_{1,\epsilon}\|\mathbb{P}_{k,\epsilon}\right) \leq \frac{(2^q+2)^2\epsilon^{2q}}{2}\sum_{s=1}^{T}\mathbb{P}_{1,\epsilon}\left(\mathbf{x}_s \in U_k^\epsilon\right).$$

Combining the above inequality with (31) and (32) gives

$$\begin{aligned}
\frac{1}{2^d}\sum_{k=1}^{2^d}\mathbb{E}_{k,\epsilon}\left[R_T(\pi)\right] &\geq 2^{-2q}\frac{2^d-1}{2^d}\sum_{t=1}^{T}\epsilon^q\exp\left(-\frac{1}{2^d-1}\sum_{k=2}^{2^d}D_{kl}\left(\mathbb{P}_{1,\epsilon}\|\mathbb{P}_{k,\epsilon}\right)\right) \\
&\geq 2^{-2q}\frac{2^d-1}{2^d}\sum_{t=1}^{T}\epsilon^q\exp\left(-\frac{1}{2^d-1}\sum_{k=2}^{2^d}\frac{(2^q+2)^2\epsilon^{2q}}{2}\sum_{s=1}^{T}\mathbb{P}_{1,\epsilon}\left(\mathbf{x}_s \in U_k^\epsilon\right)\right) \\
&\geq 2^{-2q}\frac{2^d-1}{2^d}\sum_{j=1}^{T}\epsilon^q\exp\left(-\frac{1}{2^d-1}\cdot\frac{(2^q+2)^2\epsilon^{2q}}{2}T\right),
\end{aligned}$$

where the last line uses $\sum_{k=2}^{2^d}\mathbb{P}_{1,\epsilon}\left(\mathbf{x}_s \in U_k^\epsilon\right) \leq 1$, since U_k^ϵ are disjoint.

By picking $\epsilon^q = \frac{\sqrt{2(2^d-1)}}{2^q+2}\cdot\sqrt{\frac{1}{T}}$, we have

$$\frac{1}{2^d}\sum_{k=1}^{2^d}\mathbb{E}_{k,\epsilon}\left[R_T(\pi)\right] \geq \frac{2^d-1}{2^d}\cdot\frac{\sqrt{2(2^d-1)}}{(2^q+2)2^{2q}}e^{-1}\sqrt{T}.$$

□

5 Conclusion

This paper studies the nondegenerate bandit problem with communication constraints. The nondegenerate bandit problem is important in that it encapsulates important problem classes, ranging from dynamic pricing to Riemannian optimization. We introduce the Geometric Narrowing (GN) algorithm that solves such problems in a near-optimal way. We establish that, when compared to GN, there is little room for improvement in terms of regret order or communication complexity.

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24.

Absil, P.-A., Mahony, R., and Sepulchre, R. (2008). *Optimization algorithms on matrix manifolds*. Princeton University Press.

Agarwal, A., Agarwal, S., Assadi, S., and Khanna, S. (2017). Learning with limited rounds of adaptivity: coin tossing, multi-armed bandits, and ranking from pairwise comparisons. In *Conference on Learning Theory*, pages 39–75. PMLR.

Agarwal, A., Ghuge, R., and Nagarajan, V. (2022). Batched dueling bandits. *arXiv preprint arXiv:2202.10660*.

Agrawal, R. (1995). The continuum-armed bandit problem. *SIAM Journal on Control and Optimization*, 33(6):1926–1951.

Agrawal, S. and Goyal, N. (2012). Analysis of thompson sampling for the multi-armed bandit problem. In *Conference on learning theory*, pages 39–1. JMLR Workshop and Conference Proceedings.

Arora, S., Hazan, E., and Kale, S. (2012). The multiplicative weights update method: a meta-algorithm and applications. *Theory of computing*, 8(1):121–164.

Auer, P. (2002). Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422.

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002a). Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. (2002b). The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77.

Auer, P., Ortner, R., and Szepesvári, C. (2007). Improved rates for the stochastic continuum-armed bandit problem. In *Conference on Computational Learning Theory*, pages 454–468. Springer.

Berry, D. A. and Fristedt, B. (1985). Bandit problems: sequential allocation of experiments (monographs on statistics and applied probability). *London: Chapman and Hall*, 5(71-87):7–7.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Boumal, N. (2023). *An introduction to optimization on smooth manifolds*. Cambridge University Press.

Bretagnolle, J. and Huber, C. (1978). Estimation des densités : Risque minimax. In Dellacherie, C., Meyer, P. A., and Weil, M., editors, *Séminaire de Probabilités XII*, pages 342–363, Berlin, Heidelberg. Springer Berlin Heidelberg.

Bubeck, S. and Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122.

Bubeck, S., Munos, R., Stoltz, G., and Szepesvári, C. (2009). Online optimization in $\mathcal{X}$-armed bandits. *Advances in Neural Information Processing Systems*, 22:201–208.

Bubeck, S., Munos, R., Stoltz, G., and Szepesvári, C. (2011a). $\mathcal{X}$-armed bandits. *Journal of Machine Learning Research*, 12(5):1655–1695.

Bubeck, S., Stoltz, G., and Yu, J. Y. (2011b). Lipschitz bandits without the Lipschitz constant. In *International Conference on Algorithmic Learning Theory*, pages 144–158. Springer.

Cesa-Bianchi, N., Dekel, O., and Shamir, O. (2013). Online learning with switching costs and other adaptive adversaries. *Advances in Neural Information Processing Systems*, 26:1160–1168.

Chen, S., Ma, S., Man-Cho So, A., and Zhang, T. (2020). Proximal gradient method for nonsmooth optimization over the stiefel manifold. *SIAM Journal on Optimization*, 30(1):210–239.

Chen, X. and Wang, Y. (2023). Robust dynamic pricing with demand learning in the presence of outlier customers. *Operations Research*, 71(4):1362–1386.

Chu, W., Li, L., Reyzin, L., and Schapire, R. (2011). Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 208–214. JMLR Workshop and Conference Proceedings.

Contal, E., Perchet, V., and Vayatis, N. (2014). Gaussian process optimization with mutual information. In *International Conference on Machine Learning*, pages 253–261. PMLR.

Cope, E. W. (2009). Regret and convergence bounds for a class of continuum-armed bandit problems. *IEEE Transactions on Automatic Control*, 54(6):1243–1253.

Dani, V., Kakade, S. M., and Hayes, T. (2007). The price of bandit information for online optimization. *Advances in Neural Information Processing Systems*, 20.

Esfandiari, H., Karbasi, A., Mehrabian, A., and Mirrokni, V. (2021). Regret bounds for batched bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7340–7348.

Feng, Y., Huang, Z., and Wang, T. (2023). Lipschitz bandits with batched feedback. *IEEE Transactions on Information Theory*.

Gao, B., Liu, X., Chen, X., and Yuan, Y.-x. (2018). A new first-order algorithmic framework for optimization problems with orthogonality constraints. *SIAM Journal on Optimization*, 28(1):302–332.

Gao, B., Son, N. T., Absil, P.-A., and Stykel, T. (2021). Riemannian optimization on the symplectic stiefel manifold. *SIAM Journal on Optimization*, 31(2):1546–1575.

Gao, Z., Han, Y., Ren, Z., and Zhou, Z. (2019). Batched multi-armed bandits problem. *Advances in Neural Information Processing Systems*, 32:503–513.

Gemp, I., Marris, L., and Piliouras, G. (2024). Approximating nash equilibria in normal-form games via stochastic optimization. In *The Twelfth International Conference on Learning Representations*.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 41(2):148–164.

Han, Y., Zhou, Z., Zhou, Z., Blanchet, J., Glynn, P. W., and Ye, Y. (2020). Sequential batch learning in finite-action linear contextual bandits. *arXiv preprint arXiv:2004.06321*.

Huang, W., Gallivan, K. A., and Absil, P.-A. (2015). A broyden class of quasi-newton methods for riemannian optimization. *SIAM Journal on Optimization*, 25(3):1660–1685.

Jun, K.-S., Jamieson, K., Nowak, R., and Zhu, X. (2016). Top arm identification in multi-armed bandits with batch arm pulls. In *Artificial Intelligence and Statistics*, pages 139–148. PMLR.

Karpov, N., Zhang, Q., and Zhou, Y. (2020). Collaborative top distribution identifications with limited interaction. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 160–171. IEEE.

Kleinberg, R. (2005). Nearly tight bounds for the continuum-armed bandit problem. *Advances in Neural Information Processing Systems*, 18:697–704.

Kleinberg, R., Slivkins, A., and Upfal, E. (2008). Multi-armed bandits in metric spaces. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 681–690.

Krishnamurthy, A., Langford, J., Slivkins, A., and Zhang, C. (2020). Contextual bandits with continuous actions: Smoothing, zooming, and adapting. *The Journal of Machine Learning Research*, 21(1):5402–5446.

Lai, T. L. and Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22.

Lattimore, T. and Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.

Li, J., Balasubramanian, K., and Ma, S. (2023a). Stochastic zeroth-order riemannian derivative estimation and optimization. *Mathematics of Operations Research*, 48(2):1183–1211.

Li, J., Balasubramanian, K., and Ma, S. (2023b). Zeroth-order riemannian averaging stochastic approximation algorithms. *arXiv preprint arXiv:2309.14506*.

Li, Z. and Scarlett, J. (2022). Gaussian process bandit optimization with few batches. In *International Conference on Artificial Intelligence and Statistics*, pages 92–107. PMLR.

Lu, S., Wang, G., Hu, Y., and Zhang, L. (2019). Optimal algorithms for Lipschitz bandits with heavy-tailed rewards. In *International Conference on Machine Learning*, pages 4154–4163.

Magureanu, S., Combes, R., and Proutiere, A. (2014). Lipschitz bandits: Regret lower bound and optimal algorithms. In *Conference on Learning Theory*, pages 975–999. PMLR.

Majzoubi, M., Zhang, C., Chari, R., Krishnamurthy, A., Langford, J., and Slivkins, A. (2020). Efficient contextual bandits with continuous actions. *Advances in Neural Information Processing Systems*, 33:349–360.

Perakis, G. and Singhvi, D. (2023). Dynamic pricing with unknown nonparametric demand and limited price changes. *Operations Research*.

Perchet, V., Rigollet, P., Chassang, S., and Snowberg, E. (2016). Batched bandit problems. *The Annals of Statistics*, 44(2):660–681.

Petersen, P. (2006). *Riemannian geometry*, volume 171. Springer.

Podimata, C. and Slivkins, A. (2021). Adaptive discretization for adversarial lipschitz bandits. In Belkin, M. and Kpotufe, S., editors, *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 3788–3805. PMLR.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.*, 58:527–535.

Ruan, Y., Yang, J., and Zhou, Y. (2021). Linear bandits with limited adaptivity and learning distributional optimal design. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 74–87.

Ruszczyński, A. (2021). A stochastic subgradient method for nonsmooth nonconvex multilevel composition optimization. *SIAM Journal on Control and Optimization*, 59(3):2301–2320.

Sato, H., Kasai, H., and Mishra, B. (2019). Riemannian stochastic variance reduced gradient algorithm with retraction and vector transport. *SIAM Journal on Optimization*, 29(2):1444–1472.

Shamir, O. (2013). On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on Learning Theory*, pages 3–24. PMLR.

Slivkins, A. (2014). Contextual bandits with similarity information. *Journal of Machine Learning Research*, 15(1):2533–2568.

Slivkins, A. (2019). Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286.

Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. W. (2012). Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE transactions on information theory*, 58(5):3250–3265.

Tao, C., Zhang, Q., and Zhou, Y. (2019). Collaborative learning with limited interaction: tight bounds for distributed exploration in multi-armed bandits. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 126–146. IEEE.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294.

Valko, M., Carpentier, A., and Munos, R. (2013). Stochastic simultaneous optimistic optimization. In *International Conference on Machine Learning*, pages 19–27. PMLR.

Yau, S.-T. (1974). Non-existence of continuous convex functions on certain riemannian manifolds. *Mathematische Annalen*, 207(4):269–270.

Zhang, L., Yangt, T., Yi, J., Jin, R., and Zhou, Z.-H. (2017). Improved dynamic regret for non-degenerate functions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 732–741, Red Hook, NY, USA. Curran Associates Inc.