The Annals of Statistics
2022, Vol. 50, No. 4, 2179–2204
https://doi.org/10.1214/22-AOS2182

STOCHASTIC CONTINUUM-ARMED BANDITS WITH ADDITIVE MODELS: MINIMAX REGRETS AND ADAPTIVE ALGORITHM

BY T. TONY CAI[a] AND HONGMING PU[b]

Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, [a]*tcai@wharton.upenn.edu,* [b]*hpu@wharton.upenn.edu*

We consider d-dimensional stochastic continuum-armed bandits with the expected reward function being additive β-Hölder with sparsity s for $0 < \beta < \infty$ and $1 \le s \le d$. The rate of convergence $\tilde{O}(s \cdot T^{\frac{\beta+1}{2\beta+1}})$ for the minimax regret is established where T is the number of rounds. In particular, the minimax regret does not depend on d and is linear in s. A novel algorithm is proposed and is shown to be rate-optimal, up to a logarithmic factor of T.

The problem of adaptivity is also studied. A lower bound on the cost of adaptation to the smoothness is obtained and the result implies that adaptation for free is impossible in general without further structural assumptions. We then consider adaptive additive SCAB under an additional self-similarity assumption. An adaptive procedure is constructed and is shown to simultaneously achieve the minimax regret for a range of smoothness levels.

1. Introduction. The multiarmed bandit problem, first formulated by Robbins (1952), has been studied for more than sixty years. It is an online learning framework that captures the trade-off between exploration of new information and exploitation of historical information. This framework has since been widely used in many fields, including computer science such as paging and cashing (e.g., Blum, Burch and Kalai (1999)) and recommendation systems (e.g., Bresler, Chen and Shah (2014)); statistics such as sequential experimental design (e.g., Lai and Robbins (1985)); economics such as optimization of seller's prices (e.g., den Boer (2015)) and crowdsourcing platform design (e.g., Slivkins and Vaughan (2014)); medical decision making such as clinical trial design (e.g., Pearson and Berry (1981)); and operation research such as assortment selection (e.g., Agrawal et al. (2019)).

Among the bandit problems, the ones with exponentially or infinitely large action sets have become the subject of intensive study in recent years, see Kleinberg (2004), Slivkins (2011), Banks and Sundaram (1992), McMahan and Blum (2004), Auer et al. (2002), Hazan and Megiddo (2007), Kakade, Kalai and Ligett (2009), Awerbuch and Kleinberg (2008), and Hazan and Kale (2011). Such problems, known as continuum-armed bandits (Agrawal (1995)), originate from applications such as online auctions, web advertising, and adaptive routing, where one must make a decision among an infinite number of choices. In the case where the outcomes are independent and identically distributed (i.i.d.) when the same arm is pulled, the problem is known as stochastic continuum-armed bandits (SCAB).

1.1. *Stochastic continuum-armed bandits* (*SCAB*). In SCAB, nature draws a sequence of independent random functions

$$\big(Y_t(\cdot) : [0,1]^d \to \mathbb{R}\big), \quad t = 1, 2, \ldots,$$

where $Y_t(x)$ denotes the random reward for arm x at round t. We assume for each $x \in [0,1]^d$, $Y_t(x), t = 1, 2, \ldots,$ are i.i.d. random variables with the same expectation $f(x)$. At each

Received August 2021; revised February 2022.
MSC2020 subject classifications. Primary 62G08; secondary 62L12.
Key words and phrases. Adaptivity, communication constraints, additive model, bandits, curse of dimensionality, minimax lower bound, optimal rate of convergence, regret, self-similarity.

round t, the decision maker (algorithm) pulls an arm $x \in [0,1]^d$ based on the observations strictly anterior to the round t and receives a reward $Y_t(x)$. In this problem, an algorithm $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_T\}$ is a sequence of possibly randomized maps $\mathcal{A}_t : ([0,1]^d)^{t-1} \times \mathbb{R}^{t-1} \to [0,1]^d$, $t = 2, \ldots, T$ with the exception of $\mathcal{A}_1 \in [0,1]^d$ being a possibly random number. This algorithm creates a sequence of arms $\{X_1, \ldots, X_T\} \in ([0,1]^d)^T$ and observations $\{Y_1(X_1), \ldots, Y_T(X_T)\} \in \mathbb{R}^T$ where $X_1 = \mathcal{A}_1$ and

$$X_t = \mathcal{A}_t\big(X_1, \ldots, X_{t-1}, Y_1(X_1), \ldots, Y_{t-1}(X_{t-1})\big).$$

The goal of the decision maker is to maximize the expected total reward $\mathbb{E}\sum_{t=1}^T Y_t(X_t)$ where T is the time horizon. Let $x_* = \arg\max_{x\in[0,1]^d} f(x)$ be any maximizer of f. Obviously, the optimal algorithm $\mathcal{A}^*$ would be always pulling x_* if the mean reward function $f(\cdot)$ were known. However, $\mathcal{A}^*$ is infeasible since $f(\cdot)$ is unknown in practice. Following the convention, we use $\mathcal{A}^*$ as the oracle rule to benchmark all feasible policies. Specifically, the performance of an algorithm $\mathcal{A}$ is measured by its regret,

$$R_T(\mathcal{A}) = \mathbb{E}\sum_{t=1}^T \big[f(x_*) - Y_t(X_t)\big],$$

which is the difference between its and the oracle's total expected rewards.

There is a body of literature on SCAB (Kleinberg (2004), Auer, Ortner and Szepesvári (2007), Maillard and Munos (2010), Pandey et al. (2007)). One line of related literature (Kleinberg (2004), Auer, Ortner and Szepesvári (2007), Kleinberg, Slivkins and Upfal (2008), Bubeck et al. (2011), Locatelli and Carpentier (2018), Kleinberg, Slivkins and Upfal (2019), Liu, Wang and Singh (2021), Singh (2021), Zhao and Lai (2021)) models nonparametrically by making continuity assumptions on the mean reward function. Kleinberg (2004) focuses on the case where the mean reward is a β-Hölder function with $\beta \le 1$ and obtains nearly tight upper and lower bounds $\tilde{O}(T^{\frac{\beta+1}{2\beta+1}})$ and $\tilde{\Omega}(T^{\frac{\beta+1}{2\beta+1}-o(1)})$, respectively for the minimax regret in the one-dimensional setting. Kleinberg, Slivkins and Upfal (2008) considers the multidimensional setting and assumes the mean reward to be Lipschitz. Liu, Wang and Singh (2021) extends Kleinberg (2004) and Kleinberg, Slivkins and Upfal (2008) to general Hölder smoothness in the multidimensional setting and proposes an algorithm that achieves $\tilde{O}(T^{\frac{\beta+d}{2\beta+d}})$ regret. For the case of general Hölder smoothness, a lower bound $\Omega(T^{\frac{\beta+d}{2\beta+d}})$ for regret can be deduced from Wang, Balakrishnan and Singh (2019), which studies a more general problem, as argued by Liu, Wang and Singh (2021). Another line of research (Auer (2002), Dani, Hayes and Kakade (2008), Abbasi-Yadkori, Pál and Szepesvári (2011), Rusmevichientong and Tsitsiklis (2010)) instead considers SCAB with mean reward being a linear function. In this case, the minimax regret is shown to be $\tilde{\Theta}(d \cdot \sqrt{T})$ (Dani, Hayes and Kakade (2008)). The above two lines of literature model the mean reward by a nonparametric regression and a linear regression, respectively.

As seen from the rate $T^{\frac{\beta+d}{2\beta+d}}$, multidimensional SCAB suffers from the curse of dimensionality and the optimal performance deteriorates significantly as the dimension d grows. As known in the nonparametric function estimation literature, one effective approach to circumvent the curse of dimensionality is to use the additive model, which assumes the regression function to be the sum of univariate functions of the individual variables (e.g., Stone (1985), Linton and Nielsen (1995), Yuan and Zhou (2016)). A natural and popular extension of the additive regression is the sparse additive regression, where the regression function is the sum of s univariate functions of the individual variables for some $s \ll d$. Yuan and Zhou (2016) considers the setting where each component function resides in a certain reproducing kernel Hilbert space and shows that in some regime the optimal rate of convergence coincides with that for estimating a univariate function.

In this paper, we study minimax regrets and adaptivity for additive SCAB with the expected reward function being additive β-Hölder for $0 < \beta < \infty$. That is, the expected reward function

$$f(x) = \sum_{j=1}^{d} f_j(x^{(j)}), \quad \text{for all } x \in [0,1]^d,$$

where f_j are univariate β-Hölder functions. Let $s = \mathrm{Card}(\{j : f_j \not\equiv 0\})$ be the number of functions f_j that are not identically zero. We shall call such a function *additive β-Hölder with sparsity s* and the corresponding problem *additive β-Hölder SCAB with sparsity s*. We will explore the full range of the smoothness level $0 < \beta < \infty$ and the sparsity level $1 \le s \le d$ with a particular interest in the high-dimensional sparse additive SCAB.

1.2. *Main results and our contribution.* A novel algorithm is proposed for the additive β-Hölder SCAB with sparsity s and smoothness level $\beta > 0$. It is shown that the algorithm achieves $O(s \cdot T^{\frac{\beta+1}{2\beta+1}} \cdot \ln^3(T))$ regret. A lower bound for the minimax regret of order $\Omega(s \cdot T^{\frac{\beta+1}{2\beta+1}})$ is also obtained. The two results together establish the minimax rate $s \cdot T^{\frac{\beta+1}{2\beta+1}}$, up to a logarithmic factor, and thus the proposed algorithm is nearly minimax optimal.

In comparison with the minimax regret $\tilde{\Theta}(T^{\frac{\beta+d}{2\beta+d}})$ for the nonparametric SCAB, the minimax rate $s \cdot T^{\frac{\beta+1}{2\beta+1}}$ has two major distinctions. Firstly, the dimension d has no influence at all. It thus avoids the curse of dimensionality. Secondly, the sparsity s does not affect the exponent of T. Therefore, it significantly improves the minimax regret of the nonparametric SCAB, especially in the high-dimensional setting. On the other hand, the minimax rate for sparse additive regression is $\Theta(s \cdot n^{\frac{2\beta}{2\beta+1}} + \frac{s \cdot \ln d}{n})$ (Raskutti, Wainwright and Yu (2012)), where s is the sparsity, d is the dimension, and n is the sample size. There is a logarithmic dependence on d. However, the minimax regret of the sparse additive SCAB is dimension-free. This highlights the difference between the two problems. In addition, the linear SCAB can be viewed as a special case of the additive SCAB with smoothness $\beta = \infty$. Plugging $\beta = \infty$ into the minimax regret $\tilde{\Theta}(d \cdot T^{\frac{\beta+1}{2\beta+1}})$ of the additive SCAB formally recovers the minimax regret $\tilde{\Theta}(d \cdot \sqrt{T})$ of the linear SCAB. In all, our results (roughly) bridge the gap between the nonparametric SCAB and the linear SCAB.

Besides establishing the minimax rate, we also consider adaptation to the two unknown parameters: the sparsity s and smoothness β. It is shown that adaptation to the unknown sparsity s is free in that an algorithm can achieve $s \cdot \tilde{O}(T^{\frac{\beta+1}{2\beta+1}})$ regret without prior knowledge of s. On the other hand, we prove that adaptation to the unknown smoothness is in general impossible. It is shown that for any two smoothness levels $\alpha > \beta > 0$, no algorithm can achieve near minimax regrets $s \cdot \tilde{O}(T^{\frac{\alpha+1}{2\alpha+1}})$ and $s \cdot \tilde{O}(T^{\frac{\beta+1}{2\beta+1}})$ simultaneously over the α-Hölder and β-Hölder additive classes. A significant penalty must be paid for not knowing the smoothness level in the additive SCAB.

For applications, it is critical to have a data-driven adaptive procedure that does not rely on the knowledge of the smoothness β. To this end, we consider the additive SCAB under an additional self-similarity assumption, which has been used in the literature to enable the construction of adaptive nonparametric confidence intervals (Giné and Nickl (2010), Picard and Tribouley (2000)). It is shown that adding this condition does not make the problem easier in the sense that the minimax regret is still $s \cdot \tilde{\Theta}(T^{\frac{\beta+1}{2\beta+1}})$. We then construct an adaptive procedure that simultaneously achieves the minimax regret $s \cdot \tilde{O}(T^{\frac{\beta+1}{2\beta+1}})$ for a range of smoothness levels β.

In related literature (Minsker (2013), Wang, Balakrishnan and Singh (2019), Locatelli and Carpentier (2018), Zhao and Lai (2021)), additional assumptions that restrict the measure of superlevel sets of f is often considered. These assumptions can describe the difficulty of finding the maximum region of a function. We extend the current problem setting by adding a new superlevel set assumption. Suppose the measure of superlevel set of depth ϵ_1 is bounded by $O(\epsilon_1^\gamma)$ for each nonzero component and $\beta\gamma \le 1$. We prove that the minimax regret with this additional assumption is $\tilde{\Theta}(s \cdot T^{\frac{\beta+1-\beta\gamma}{2\beta+1-\beta\gamma}})$. This regret is again dimension-free, which shows that the dimension-free phenomenon can be generalized to the setting under the additional superlevel set assumption. We also consider the problem of adaptation to γ and show that under mild conditions, adaptation to γ can be achieved without additional cost.

1.3. *Related literature.* SCAB has been studied under other structural assumptions beyond Hölder continuity. For example, Agarwal et al. (2013) assumes that the mean cost function is globally convex and views it as an online convex optimization problem. An upper bound of $\tilde{O}(T^{\frac{1}{2}})$ is given in this case, which is clearly not improvable up to a logarithmic factor. Cope (2009) proves an upper bound of $O(T^{\frac{1}{2}})$ for the asymptotic regret under the assumptions that the mean reward is unimodal, three times continuously differentiable and its derivative is well behaved at its maximizer. SCAB with additive models has been considered in the Bayesian setting with a given prior (Kandasamy, Schneider and Póczos (2015), Rolland et al. (2018), Delbridge, Bindel and Wilson (2020)). However, minimax regrets for the additive SCAB and the more general sparse additive SCAB are unknown.

A problem connected to SCAB is adversarial continuum-armed bandits where the payoff distribution is allowed to vary during the game. Kleinberg (2004) proves that the minimax regret is $\tilde{\Theta}(T^{\frac{\beta+1}{2\beta+1}})$ for adversarial continuum-armed bandits under the β-Hölder continuity assumption for $\beta \le 1$ with bandit feedback. Maillard and Munos (2010) further considers this problem in the full-feedback setting and shows that the minimax regret for the Lipschitz class (Hölder continuity with smoothness 1) is $\tilde{\Theta}(\sqrt{T})$.

Another related problem is SCAB with covariates, where in each round nature draws a context U and the decision maker chooses an action $\mathbf{X} \in \mathcal{X}$ not only based on the history but also on this context. The reward $Y(X, U)$ depends both on the action and context. Lu, Pál and Pál (2009) and Slivkins (2011) consider the problem under the Lipschitz assumption and prove a lower bound $\Omega(T^{\frac{d_1+d_2+1}{d_1+d_2+2}})$ for the minimax regret where d_1 and d_2 are the packing dimensions of the context and arms spaces. Lu, Pál and Pál (2009) also shows a nearly tight upper bound of $\tilde{O}(T^{\frac{\bar{d}_1+\bar{d}_2+1}{\bar{d}_1+\bar{d}_2+2}})$ where $\bar{d}_1$ and $\bar{d}_2$ are covering dimensions of the context and arms spaces.

In this paper, we prove that adaptation to smoothness is impossible. This phenomenon is connected to many other nonadaptivity results in various nonparametric bandits problems (Locatelli and Carpentier (2018), Liu, Wang and Singh (2021), Gur, Momeni and Wager (2021)). When adaptation for free is impossible, a common approach is to consider additional conditions that enable adaptivity (Combes and Proutiere (2014), Bull (2015), Locatelli and Carpentier (2018), Gur, Momeni and Wager (2021)). A novel approach has been recently proposed. Hadiji (2019) studies one-dimensional nonparametric SCAB, which is well known to be nonadaptive, and considers admissibility in the minimax sense. A class of algorithms are shown to be (minimax) admissible. Such (minimax) admissibility can be used as a general criterion to study nonadaptivity without additional assumptions. We give a further discussion on nonadaptivity in Section 6.

1.4. *Notation and definitions.* For two functions $g_1(T), g_2(T) > 0$, we write $g_1(T) = O(g_2(T))$ if $\limsup_{T\to\infty} \frac{g_1(T)}{g_2(T)} < \infty$; $g_1(T) = \tilde{O}(g_2(T))$ if there exists a constant $C > 0$ such that $g_1(T) = O(g_2(T) \cdot \ln^C(T))$; $g_1(T) = \Omega(g_2(T))$ if $\liminf_{T\to\infty} \frac{g_1(T)}{g_2(T)} > 0$; $g_1(T) = \Theta(g_2(T))$ if $g_1(T) = O(g_2(T))$ and $g_2(T) = O(g_1(T))$; and $g_1(T) = \tilde{\Theta}(g_2(T))$ if $g_1(T) = \tilde{O}(g_2(T))$ and $g_2(T) = \tilde{O}(g_1(T))$. For any vector $x \in \mathbb{R}^{d_0}$ where d_0 is some positive integer, let $\|x\|$ denote the L_2 norm of x. For any matrix M, let $\|M\|$ denote the L_2 operator norm of M, that is, $\|M\| = \sup_{x\neq 0} \frac{\|Mx\|}{\|x\|}$. For a positive number $\beta > 0$, $w(\beta)$ denotes the largest integer that is strictly smaller than β. For a finite-dimensional vector x, let $x^{(j)}$ denote the jth element of x. For a function g and a nonnegative integer k, let $g^{(k)}$ denote the kth derivative of g. For two vectors x and y of the same dimension, let $\langle x, y\rangle$ be the inner product of these two vectors. Let $\mathbb{Z}^+$ and $\mathbb{R}^+$ denote the collection of positive integers and the collection of positive real numbers.

DEFINITION 1. The Hölder class of functions $\mathcal{H}_0(\beta, L)$ is defined to be the set of $w(\beta)$ times continuously differentiable functions $g : [0, 1] \to \mathbb{R}$ such that for any $x, x' \in [0, 1]$,

$$\left|g^{(w(\beta))}(x) - g^{(w(\beta))}(x')\right| \le L \cdot |x - x'|^{\beta - w(\beta)}.$$

Let

$$\mathcal{H}(\beta, L) = \begin{cases} \mathcal{H}_0(\beta, L), & \beta < 1, \\ \mathcal{H}_0(\beta, L) \cap \mathcal{H}_0(1, L), & \beta \ge 1. \end{cases} \tag{1}$$

DEFINITION 2. A family of random variables $Y(x) \in \mathbb{R}$ indexed by $x \in [0, 1]^d$ is defined to be uniformly sub-Gaussian with positive constants u_1, u_2 if for any $x \in [0, 1]^d$ and $t > 0$ the following inequality holds:

$$\mathbb{P}\left(\left|Y(x)\right| \ge t\right) \le u_1 \cdot e^{-u_2 t^2}.$$

1.5. *Organization.* In Section 2, we consider the minimaxity of SCAB. A lower bound is presented in Section 2.1. An algorithm is proposed in Section 2.2 and an upper bound for the regret of this algorithm is obtained in Section 2.3. It is shown in Section 3 that adaptivity is in general impossible. In Section 4, we discuss the adaptivity under the self-similarity assumption. Specifically, Section 4.1 introduces the self-similarity assumption, a new algorithm is proposed in Section 4.2, and an upper bound is derived in Section 4.3 to show that this algorithm is adaptive under the self-similarity condition. Section 5 considers the additive SCAB under an additional superlevel set assumption and establishes the corresponding minimax regret. Adaptivity under this assumption is also considered. Finally, a discussion is provided in Section 6. The proof of a main theorem is presented in Section 7. For reasons of space, the proofs of other main and technical results are given in the Supplementary Material Cai and Pu (2022).

2. Minimax optimal rates for the regret. We begin by introducing some assumptions. The first is on the additive structure of the mean reward function.

ASSUMPTION 1. We assume the mean reward f can be represented as

$$f(x) = \sum_{j=1}^{d} f_j\left(x^{(j)}\right), \quad \text{for all } x \in [0, 1]^d,$$

where only s of these d functions $\{f_1, \ldots, f_d\}$ are nonzero.

Our second assumption addresses the hypothesis space containing $f_j(\cdot), j = 1, \ldots, d$. This paper aims to construct an algorithm achieving low regret without restricting each $f_j(\cdot)$ in a small parametric function space. Alternatively, we assume each $f_j(\cdot)$ to be in a Hölder class of functions.

ASSUMPTION 2. For any $j \in \{1, \ldots, d\}$, f_j belongs to the Hölder class $\mathcal{H}(\beta, L)$ for some $\beta > 0$ and $L > 0$.

Recall the definition of $\mathcal{H}(\beta, L)$ in (1). When $\beta \geq 1$, $f_j \in \mathcal{H}(\beta, L)$ means it is Lipschitz continuous.

The third assumption requires the conditional distribution of Y given X to be uniformly sub-Gaussian.

ASSUMPTION 3. $\mathbb{E}Y_1(x)$ exists for all $x \in [0, 1]^d$ and $\{Y_1(x) - \mathbb{E}Y_1(x) : x \in [0, 1]^d\}$ is uniformly sub-Gaussian with some constants $u_1 > 1$, $u_2 > 0$.

This assumption is strictly weaker than the usual assumption of bounded outcomes, which is often made in the literature (Rigollet and Zeevi (2010), Hu, Kallus and Mao (2020), Dudik et al. (2011)). It can also admit other common models such as outcomes being Gaussian with bounded variances. The requirement of $u_1 > 1$ is necessary. If $u_1 < 1$, then no distribution satisfies this condition.

With the above two assumptions we can formally define the minimax regret. Since the regret of any algorithm depends on the specific instance of the problem, we write $R_T(\mathcal{A})$ as $R_T(\mathcal{A}; \mathbb{P})$ where

$$\mathbb{P} = \{\mathbb{P}_x : x \in [0, 1]^d, \mathbb{P}_x \text{ is a distribution on } \mathbb{R}\}$$

denotes the distributions of $Y_1(x)$. We consider all instances that fit the assumptions. Let $\mathcal{P}(s, d, \beta, L, u_1, u_2)$ denote all of $\mathbb{P}$ that satisfy Assumptions 1, 2 and 3. A fundamental benchmark is the minimax regret defined by

$$\inf_{\mathcal{A}} \sup_{\mathbb{P} \in \mathcal{P}(s,d,\beta,L,u_1,u_2)} R_T(\mathcal{A}; \mathbb{P}).$$

The rate of minimax regret is established in two steps. We first prove a lower bound for the maximum regret of any algorithm and then develop an algorithm that attains this rate under the three assumptions.

2.1. *Minimax lower bound.* We begin with the minimax lower bound for the minimax regret over $\mathcal{P}(s, d, \beta, L, u_1, u_2)$.

THEOREM 1. *For any positive parameters β, L, u_1, u_2, and the number of rounds T, there exists a constant $C > 0$ depending on β, L, u_1, u_2 only and not on T, s, d such that*

$$\inf_{\mathcal{A}} \sup_{\mathbb{P} \in \mathcal{P}(s,d,\beta,L,u_1,u_2)} R_T(\mathcal{A}; \mathbb{P}) \geq C \cdot s \cdot T^{\frac{\beta+1}{2\beta+1}}.$$

This theorem establishes a general lower bound for the worst-case performance of any algorithm for the SCAB with additive models. Recall that the linear SCAB is a special case of the SCAB with additive models. In the linear SCAB, each f_j is a linear function and thus belongs to $\mathcal{H}(\beta, L)$ for any $\beta > 0$. Therefore, intuitively the linear SCAB can be roughly viewed as the additive SCAB with smoothness $\beta = \infty$. If we plug $\beta = \infty$ and $s = d$ into Theorem 1, then the sharp regret lower bound $\Theta(d \cdot T^{\frac{1}{2}})$ for the linear SCAB (Dani, Hayes

and Kakade (2008)) is recovered. Therefore Theorem 1 generalizes the lower bound of the regret for the linear SCAB to the additive SCAB.

We shall show this bound is nearly tight for all β by giving a novel algorithm and proving a matched upper bound in the next section.

2.2. *Algorithm.* In this section, we develop an algorithm for arbitrary smoothness. Specifically, we begin with a review of local polynomial regression in Section 2.2.1, which is utilized in our algorithm to estimate $f(\cdot)$ and construct confidence intervals. The proposed algorithm is then presented in detail in Section 2.2.2.

2.2.1. *Local polynomial regression.* Local polynomial regression is an offline nonparametric regression method. A classical result of local polynomial regression is that it can achieve minimax convergence rates in a Hölder ball with proper choices of tuning parameters (Györfi et al. (2006)). We briefly review local polynomial regression in this section, and present the details of how to use it in our algorithm in the next section.

Let $\mathbb{O} = \{(X_{(1)}, Y_{(1)}), \ldots, (X_{(m)}, Y_{(m)})\}$ be i.i.d. samples, where $X_{(1)}$ has support $\subset [0, 1]$. The goal is estimating $\mathbb{E}[Y_{(1)}|X_{(1)}]$ with these samples nonparametrically. Let $S = [a, b]$ be a subset of $[0, 1]$. We consider the observations $(X_{(i)}, Y_{(i)})$ such that $X_{(i)} \in S$. Without loss of generality, we let these observations be $\mathbb{O}_S = \{(X_{(1)}, Y_{(1)}), \ldots, (X_{(m_0)}, Y_{(m_0)})\}$. We estimate $\mathbb{E}[Y_{(1)}|X_{(1)}]$ on $[a, b]$ by fitting a polynomial regression with observations in $[a, b]$, that is, $\mathbb{O}_S$, as follows.

Let $t_k(x) = (\frac{1}{2} + \frac{x - \frac{a+b}{2}}{b-a})^k$ and $t^{(l)}(x) = (t_0(x), t_1(x), \ldots, t_l(x))^T$ for some integer l. Define

$$\hat{\theta} = \arg \min_{\theta \in \mathbb{R}^{l+1}} \sum_{k=1}^{m_0} \big(Y_{(k)} - \langle t(X_{(k)}), \theta \rangle\big)^2.$$

For concreteness, if the minimizer is not unique we define $\hat{\theta} = 0$. The local polynomial regression estimate on S is given by

$$\hat{f}(x; \mathbb{O}, l, S) := \langle t^{(l)}(x), \hat{\theta} \rangle.$$

In our algorithm, we only use a special case of local polynomial regression, where S is taken to be the support of $X_{(1)}$ and $X_{(1)}$'s distribution is uniform on $[a, b]$. In this case, the following proposition justifies the convergence of $\hat{f}$.

PROPOSITION 1. *Let X be a uniform random variable in an interval $[a, b] \subset [0, 1]$. Let $\mathbb{O} = \{(X_{(1)}, Y_{(1)}), \ldots, (X_{(n)}, Y_{(n)})\}$ be an i.i.d. sample of $(X, Y(X))$. If $g(x) := \mathbb{E}[Y|X = x] \in \mathcal{H}(\beta, L)$ and $Y(x)$ is a constant C_Y plus a uniformly sub-Gaussian random variable with constants u_1 and u_2. Suppose $u_1 \le \exp(u_1' \cdot n^{\nu})$ for some positive constants ν and u_1'. Then there exists positive constants C_1, C_2 and C_3 that depend on l, u_1', u_2 but not on a, b, n, u_1, C_Y such that with probability at least $1 - O(e^{-C_2 \ln^2(n)})$, for any $x \in [a, b]$, $n > C_3$ the following inequality holds*:

$$\big|\mathbb{E}[Y|X = x] - \hat{f}\big(x; \mathbb{O}, l, [a, b]\big)\big| < (b - a)^{\beta} \ln(n) + \ln^3(n) \cdot n^{-\frac{1}{2}(1-\nu)}.$$

If we know the value of β and ν, by this proposition, we can further construct interval estimates for $\mathbb{E}[Y|X = x]$ with high confidence. Let

$$\begin{aligned}
\hat{f}^{ub}\big(x; \mathbb{O}, l, [a, b], \beta, \nu\big) &:= \hat{f}\big(x; \mathbb{O}, l, [a, b]\big) + (b - a)^{\beta} \ln(n) + \ln^3(n) \cdot n^{-\frac{1}{2}(1-\nu)}, \\
\hat{f}^{lb}\big(x; \mathbb{O}, l, [a, b], \beta, \nu\big) &:= \hat{f}\big(x; \mathbb{O}, l, [a, b]\big) - (b - a)^{\beta} \ln(n) - \ln^3(n) \cdot n^{-\frac{1}{2}(1-\nu)}, \\
\hat{f}^{ub}\big(x; \mathbb{O}, l, [a, b], \beta\big) &= \hat{f}^{ub}\big(x; \mathbb{O}, l, [a, b], \beta, 0\big), \\
\hat{f}^{lb}\big(x; \mathbb{O}, l, [a, b], \beta\big) &= \hat{f}^{lb}\big(x; \mathbb{O}, l, [a, b], \beta, 0\big).
\end{aligned}$$

By Proposition 1, $[\hat{f}^{lb}(x;\mathbb{O},l,[a,b],\beta,\nu),\ \hat{f}^{ub}(x;\mathbb{O},l,[a,b],\beta,\nu)]$ is a $1-O(e^{-C_2\ln^2(n)})$ confidence interval for $\mathbb{E}[Y|X=x]$ if n is large enough and ν is known. Especially, if the conditional distribution of Y given X is fixed then we can let $\theta=0$ and use $[\hat{f}^{lb}(x;\mathbb{O},l,[a,b],\beta),\ \hat{f}^{ub}(x;\mathbb{O},l,[a,b],\beta)]$ as a $1-O(e^{-C_2\ln^2(n)})$ confidence interval for $\mathbb{E}[Y|X=x]$. In our algorithm, the local polynomial regression as an offline regression method is used as a basic tool to estimate and construct confidence intervals for the mean reward function.

2.2.2. *Our algorithm.* In this section, we present our new procedure for the additive SCAB, which is summarized in Algorithm 1 at the end of this section. The algorithm has three input parameters: the total number of rounds T, Hölder smoothness level β_0, and the polynomial degree l. We begin by introducing the main ideas behind the construction.

The algorithm proceeds in epochs and maintains a feasible region $\subset[0,1]$ for each $j\in\{1,\dots,d\}$ where, with high confidence, the value of f_j is close to the maximal mean reward $f_j(x_*^{(j)})$. Let $G_{i,j}$ denote the feasible region in the ith epoch for the jth dimension. In each epoch, the algorithm only pulls arms whose jth elements are in the current corresponding feasible region $G_{i,j}$. Each feasible region begins with $[0,1]$ and is narrowed down as more data is collected and confidence intervals become narrower. In each epoch the algorithm decomposes current feasible region into a set of nonoverlapping intervals. Then it pulls arms from a distribution depending on the structure of these intervals and collects the observations. After that, the algorithm fits local polynomial regression with these observations to construct confidence intervals for each f_j and utilizes them to narrow the feasible region.

Specifically, in each epoch the algorithm has four steps: *reallocating*, *pulling*, *fitting*, and *eliminating*. In the reallocating step, the algorithm breaks each current feasible region into a set of intervals; in the pulling step, the algorithm pulls arms from a distribution depending on the structure of these intervals; in the fitting step, the algorithm fits a local polynomial regression with samples obtained in the pulling step and constructs confidence intervals; in the eliminating step, the algorithm eliminates the points with too small confidence upper bounds in each current feasible region and makes the remaining ones the new feasible region.

The algorithm recursively does these four steps for

$$K=\left\lfloor \ln_{2l+2}\left(\left(\frac{T}{5}\right)^{\frac{1}{2\beta_0+1}}\right)\right\rfloor-\left\lfloor \ln_{2l+2}\left(\left(\frac{T}{5}\right)^{\frac{\beta_0+1}{(2\beta_0+1)(4\beta_0+1)}}\right)\right\rfloor$$

epochs. After K epochs, all the points in the final feasible region $G_{K+1,j}$ are "sufficiently good" with high probability for each $j\in\{1,\dots,d\}$. Then it simply pulls arms randomly in the product of the d final feasible regions recursively till the final round. We describe below the four steps in detail.

Reallocating step. Note that each feasible region is always a union of intervals (see Lemma 1). For a feasible region $G_{i,j}$, let the natural decomposition of $G_{i,j}$ be $G_{i,j}=\bigcup_{k=1}^{N_0}J_k$ where $\{J_k:k=1,\dots,N_0\}$ denote intervals in the ascending order s.t.:

$$\sup_{x\in J_k}x<\inf_{x\in J_{k+1}}x,\quad \forall 1\le k<k+1\le N_0.$$

For example, if $G_{i,j}=[0,0.1)\cup[0.2,0.45)\cup[0.5,0.55)$ then $J_1=[0,0.1)$, $J_2=[0.2,0.45)$, $J_3=[0.5,0.55)$ and $N_0=3$.

For each J_k, the algorithm further decomposes it into shorter intervals. With a length parameter c_i, each J_k is broken into a collection of intervals that all have length c_i except for the last one. Specifically, let $x_{\min}=\inf_{x\in J_k}x$, $x_{\max}=\sup_{x\in J_k}x$ and $m=\lceil\frac{x_{\max}-x_{\min}}{c_i}\rceil$, then J_k is decomposed into m intervals $[x_{\min},x_{\min}+c_i),[x_{\min}+c_i,x_{\min}+2c_i),\dots,[x_{\min}+(m-$

$2)c_i, x_{\min} + (m-1)c_i), [x_{\min} + (m-1)c_i, x_{\max}] \cap J_k$. For example, if $J_k = [0.1, 0.45)$ and $c_i = 0.1$ then J_k is decomposed into four intervals $[0.1, 0.2)$, $[0.2, 0.3)$, $[0.3, 0.4)$, $[0.4, 0.45)$.

Finally, we collect all these shorter intervals into a set as $\{\mathcal{I}_{i,j,1}, \ldots, \mathcal{I}_{i,j,g_{i,j}}\}$ where $g_{i,j}$ is the capacity of this set. Define $F(\cdot,\cdot)$ by

$$F(G_{i,j}, c_i) = \{\mathcal{I}_{i,j,1}, \ldots, \mathcal{I}_{i,j,g_{i,j}}\} \tag{2}$$

For example, if $G_{i,j} = [0, 0.1) \cup [0.2, 0.45) \cup [0.5, 0.55)$ and $c_i = 0.1$ then $\mathcal{I}_{i,j,1} = [0, 0.1)$, $\mathcal{I}_{i,j,2} = [0.2, 0.3)$, $\mathcal{I}_{i,j,3} = [0.3, 0.4)$, $\mathcal{I}_{i,j,4} = [0.4, 0.45)$, $\mathcal{I}_{i,j,5} = [0.5, 0.55)$ and $g_{i,j} = 5$.

Pulling step. In this step, the algorithm pulls arm independently T_i times from a distribution $\mathcal{D}_i$. Let $\mathcal{D}_i$ be the distribution of a d-dimensional random variable $Z = (Z_1, \ldots, Z_d)$ that is sampled as follows. Let $Z_1, \ldots, Z_d$ be independent. Each Z_j is sampled by first randomly picking m from $\{1, \ldots, g_{i,j}\}$ with equal probability and then taking a random number from the uniform distribution on $\mathcal{I}_{i,j,m}$. After pulling arms, the algorithm logs the samples observed. This process is separate for different elements of the covariates and different intervals in the feasible regions. That is for each $j = 1, \ldots, d$ and $m = 1, \ldots, g_{i,j}$, the algorithm creates a set of observations $\mathbb{O}_{i,j,m}$, which only logs the samples whose jth covariates are in $\mathcal{I}_{i,j,m}$ and only stores the information of outcomes and the jth covariates. The algorithm will fit one-dimensional local polynomial regression on each $\mathbb{O}_{i,j,m}$ separately.

Fitting step. Recall the algorithm collects observations $\mathbb{O}_{i,j,m}$ in the ith epoch, for each $j = 1, \ldots, d$ and $m = 1, \ldots, g_{i,j}$. With these observations, we use local polynomial regression with order l to construct confidence lower and upper bounds of f_j for $x \in \mathcal{I}_{i,j,m}$:

$$\hat{f}^{lb}_{i,j}(x; \beta_0, l) = \left\{ \begin{array}{ll} \hat{f}^{lb}(x; \mathbb{O}_{i,j,m}, l, \mathcal{I}_{i,j,m}, \beta_0), & i > 1, \\ \hat{f}^{lb}\left(x; \mathbb{O}_{i,j,m}, l, \mathcal{I}_{i,j,m}, \beta_0, \dfrac{1}{2}\right), & i = 1, \end{array} \right\}, \tag{3}$$

$$\hat{f}^{ub}_{i,j}(x; \beta_0, l) = \left\{ \begin{array}{ll} \hat{f}^{ub}(x; \mathbb{O}_{i,j,m}, l, \mathcal{I}_{i,j,m}, \beta_0), & i > 1, \\ \hat{f}^{ub}\left(x; \mathbb{O}_{i,j,m}, l, \mathcal{I}_{i,j,m}, \beta_0, \dfrac{1}{2}\right), & i = 1. \end{array} \right\}. \tag{4}$$

Eliminating step. The algorithm already obtains upper and lower bounds of f_j for each $j = 1, \ldots, d$ on the jth feasible region $G_{i,j}$ in the last step. In this step, the algorithm narrows the feasible region for each covariate by simply eliminating the points x such that the upper bound estimate of $f_j(x)$ is smaller than the largest lower bound of f_j.

The whole algorithm is summarized in Algorithm 1 below.

We should note that it is not obvious that each $G_{i,j}$ can always be decomposed into intervals in the reallocating step. The following lemma ensures this is the case with probability one.

LEMMA 1. *In Algorithm 1, for all $1 \le i \le K$ we have with probability* 1:

1. *The Lebesgue measure of $G_{i,j}$ is strictly greater than* 0,
2. *the feasible region $G_{i,j}$ is a union of nonoverlapping intervals*.

The number of rounds in each epoch is predetermined by the algorithm. Therefore, by simple calculations we can prove that after all the K epochs the total number of rounds is smaller than T.

LEMMA 2. *In Algorithm 1, we have $r_K < T$ with probability* 1.

Hence, with probability 1 the algorithms does not stop after all K epochs. The points left in each final feasible region $G_{K+1,j}$ are sufficiently good. The algorithm then randomly pulls arms in $\prod_{j=1}^{d} G_{K+1,j}$ for another $T - r_K$ rounds till the game ends.

Algorithm 1: $\mathcal{A}^1(T, \beta_0, l)$

Input: Total number of rounds T, Hölder smoothness β_0, and polynomial degree l.

0 Initialize round counter $r_0 = 0$, feasible regions $G_{1,j} = [0,1]$, $\forall j = 1, \ldots, d$.

0 Set $K_0 = \lfloor \ln_{2l+2}((\frac{T}{5})^{\frac{\beta_0+1}{(2\beta_0+1)\cdot(4\beta_0+1)}}) \rfloor$ and epoch number $K = \lfloor \ln_{2l+2}((\frac{T}{5})^{\frac{1}{2\beta_0+1}}) \rfloor - K_0$, $\zeta = \frac{2\beta_0(\beta_0+1)}{(2\beta_0+1)\cdot(4\beta_0+1)}$.

for $i \in \{1, \ldots, K\}$ **do**

0 Set $c_i = (2l+2)^{-(i+K_0)}$, $b_i = \lfloor (2l+2)^{2\beta_0(i+K_0)} \rfloor$

0 **Reallocating Step:**
Break $G_{i,j}$ into a set of intervals: $\{\mathcal{I}_{i,j,1}, \ldots, \mathcal{I}_{i,j,g_{i,j}}\} = F(G_{i,j}, c_i)$ for $j = 1, \ldots, d$, where $F(\cdot,\cdot)$ is defined in (2) and $g_{i,j}$ is defined in Section 2.2.2.

0 **Pulling Step:**
Set constant $T_i = 2 \cdot b_i \cdot \frac{1}{c_i} \cdot \mathbb{1}\{i > 1\} + 2 \cdot b_i \cdot \frac{1}{c_i} \cdot T^{\zeta} \mathbb{1}\{i = 1\}$.
Pull arms T_i times independently from a distribution $\mathcal{D}_i$, where $\mathcal{D}_i$ is defined in Section 2.2.2.
For each $j = 1, \ldots, d$ and $m = 1, \ldots, g_{i,j}$, let the samples where the jth element of covariates is in $\mathcal{I}_{i,j,m}$ be $\mathbb{O}_{i,j,m,0} = \{(X_t, Y_t) : r_{i-1} < t \le r_{i-1} + T_i, X_t^{(j)} \in \mathcal{I}_{i,j,m}\}$.
Log these samples with outcomes and only the jth elements of covariates $\mathbb{O}_{i,j,m} = \{(X_t^{(j)}, Y_t) : (X_t, Y_t) \in \mathbb{O}_{i,j,m,0}\}$.
Update the round counter $r_i = r_{i-1} + T_i$.

0 **Fitting Step:**
For $j = 1, \ldots, d$, and $m = 1, \ldots, g_{i,j}$, fit local polynomial regression on $\mathcal{I}_{i,j,m}$ with $\mathbb{O}_{i,j,m}$ and construct confidence lower bound $\hat{f}_{i,j}^{lb}(x; \beta_0, l)$ and confidence upper bound $\hat{f}_{i,j}^{ub}(x; \beta_0, l)$ of $f_j(x)$ for each $x \in \mathcal{I}_{i,j,m}$ as in (3) and (4).

0 **Eliminating Step:**
Set $\hat{f}_{i,j}^{\max,lb} = \sup_{x \in G_{i,j}} \hat{f}_{i,j}^{lb}(x)$ for each $j = 1, \ldots, d$.
Update $G_{i+1,j} = \{x \in G_{i,j} : \hat{f}_{i,j}^{ub}(x) > \hat{f}_{i,j}^{\max,lb}\}$ for each $j = 1, \ldots, d$.

0 Pull arms $T - r_K$ times uniformly at random in $\prod_{j=1}^{d} G_{K+1,j}$.

2.3. *Minimax upper bound.* We now show that with proper choice of parameters, Algorithm 1 achieves the optimal rate of regret.

THEOREM 2. *Let $\beta_0 = \beta$, $l \ge \lfloor \beta \rfloor$ and $s < T^{\frac{\beta(\beta+1)}{(4\beta+1)^2}}$, then there exists a constant $C > 0$ such that for any $\mathbb{P} \in \mathcal{P}(s, d, \beta, L, u_1, u_2)$,*

$$R_T(\mathcal{A}^1; \mathbb{P}) \le C \cdot s \cdot \ln^3(T) \cdot T^{\frac{\beta+1}{2\beta+1}},$$

where the constant C depends on parameters β, L, u_1, u_2 and not on $\mathbb{P}$, d, s, T.

Together with the lower bound given in Theorem 1, this result establishes the minimax optimal rate $s \cdot \tilde{\Theta}(T^{\frac{\beta+1}{2\beta+1}})$, where the poly-logarithmic terms in $\tilde{\Theta}(\cdot)$ here depends on T and not on s or d, under the condition $s < T^{\frac{\beta(\beta+1)}{(4\beta+1)^2}}$. It shows Algorithm 1 is rate-optimal up to a logarithmic factor if the Hölder smoothness level β_0 is correctly chosen as β. Recall that the

minimax regret of the nonparametric SCAB is $\tilde{\Theta}(T^{\frac{\beta+d}{2\beta+d}})$ (Liu, Wang and Singh (2021)). In comparison, the minimax regret of the additive SCAB is much improved in that the sparsity s does not affect the exponent of T and the dimension d has no influence on the regret. When $s \geq T^{\frac{\beta(\beta+1)}{(4\beta+1)^2}}$, it remains unknown whether the general lower bound $\Omega(s \cdot T^{\frac{\beta+1}{2\beta+1}})$ is (nearly) tight. In this case, searching for the maximizer can be difficult since not only the natural dimension but also the effective dimension is large compared to T.

2.4. *Discussion of minimax regret under the additive model.* In the current problem, the minimax regret is a function of β, s, d, T and can have different forms if the relative magnitudes of these four parameters vary. In this paper, we consider β to be an arbitrary fixed number and allow s, d, T to vary. We prove the minimax regret is always $s \cdot \tilde{\Theta}(T^{\frac{\beta+1}{2\beta+1}})$ as long as $s < T^{\frac{\beta(\beta+1)}{(4\beta+1)^2}}$. This general setting covers several different detailed settings including: low-dimensional setting, where s equals d and does not vary with T; nonsparse mild high-dimensional setting, where s equals d and tends to infinity as T grows but is much smaller than T; sparse ultra high-dimensional setting, where $s \ll d$ and $d \gg T$.

Compared with the sparse additive regression, the minimax optimal rate $s \cdot \tilde{\Theta}(T^{\frac{\beta+1}{2\beta+1}})$ for the additive SCAB sheds new insights. The minimax rate of the L_2 risk in the sparse additive regression is $\tilde{\Theta}(s \cdot n^{-\frac{2\beta}{2\beta+1}} + s \cdot \ln(d)/n)$ (Raskutti, Wainwright and Yu (2012)). Therefore, in the sparse additive regression, the estimation risk has an inevitable logarithmic dependence on the dimension d while in the sparse additive SCAB the regret has no dependence on d. Specifically, if $\ln(d) \gg n$ then no consistent estimator exists in the sparse additive regression but in the sparse additive SCAB one can still achieve low regret no matter how high the dimension is. This highlights the difference between the sparse additive regression and the sparse additive SCAB.

In fact, in high-dimensional estimation problems, at least a $\ln(d)$ dependence in the risk is generally unavoidable (Raskutti, Wainwright and Yu (2011), Cai, Zhang and Zhou (2010), Cai and Guo (2017)). The dimension-free phenomenon here shows the intrinsic difference between high-dimensional bandits and high-dimensional estimation problems. The key reason for the dimension-free phenomenon here is that in the current setting, the overall regret can be decomposed as the sum of regrets in individual components. This gives a possibility of decomposing the whole task as d independent tasks. The algorithm here treats each dimension independently and find the maximizer of each f_j independently. To be free of $\ln(d)$, our Algorithm 1 neither tries to estimate f as a whole nor tries to estimate the effective dimensions. As a result, the algorithm does not learn which s of them are really effective. However this does not matter, since the remaning $d - s$ tasks yield no regret no matter how the algorithm does in them.

A related problem is sparse high-dimensional linear bandits. In this problem, the mean payoff f is modeled as a sparse linear function on the action X_t. In this case the problem can be reduced to a discrete problem since the maximizer of the mean payoff must be located in the vertex set $\{0, 1\}^d$ by linearity. Lattimore and Szepesvári (2020) shows a $\tilde{O}(s \cdot \sqrt{T})$ regret upper bound in this case. Our results can be viewed as extending this problem to the general additive case. Although both results show a dimension-free upper bound, our results give a deeper insight. This dimension-free phenomenon only depends on the additivity structure and not on linearity. With additivity, even though the problem cannot be reduced to a discrete problem and the maximizer can be intrinsically hard to find, one can still have a low regret algorithm no matter how large the dimension is.

In the current approach, the dimension-free phenomenon relies on the fact that the overall regret can be decomposed as the sum of regret of each component. Unfortunately, this depends not only on the additivity structure of f but also on the shape of the action set. The

action set here $[0,1]^d$ can be written as a product of sets on each dimension. This enables the decomposition of the whole task as independent tasks in each dimension. It is easy to see that the dimension-free phenomenon here can be generalized to any hyperrectangle action sets but can hardly be generalized to all action sets. In sparse high-dimensional linear bandits, it has been shown that action sets can play an important role in the regret (Lattimore and Szepesvári (2020), Hao, Lattimore and Wang (2020)). For example, although in high-dimensional linear bandits with hypercube action sets one can get a dimension-free regret, Lattimore and Szepesvári (2020) proves a $\Omega(\sqrt{sdT})$ regret lower bound on some specific action set, which implies a polynomial dependence on the dimension is generally not avoidable. Such results imply in high-dimensional SCAB, the regret is heavily affected by the action set and on a general action set dimension-free regret is typically impossible. However, a complete theory of how the action sets affect the regret remains to be developed.

The proposed algorithm requires no information on s and achieves the minimax optimal regret for all $s < T^{\frac{\beta(\beta+1)}{(4\beta+1)^2}}$. Therefore, this algorithm is adaptive to the sparsity s with no additional cost.

3. Impossibility of adaptation to smoothness. In this section, we turn to the important question of adaptation to the smoothness β. Although the proposed Algorithm 1 achieves the optimal rate of regret, the method relies on the value of the smoothness β, which is typically unknown in practice. This naturally raises the question whether it is possible to construct an algorithm that achieves the minimax optimal regret adaptively over a range of degrees of smoothness without prior knowledge of the true degree β.

Formally, an adaptive algorithm $\mathcal{A}$ satisfies

$$\sup_{\mathbb{P}\in\mathcal{P}(s,d,\beta,L,u_1,u_2)} R_T(\mathcal{A};\mathbb{P}) = \tilde{O}\big(s\cdot T^{\frac{\beta+1}{2\beta+1}}\big),$$

simultaneously for all $\beta\in\Omega_0$, where $\Omega_0\subset\mathbb{R}^+$ denotes the range of smoothness to which $\mathcal{A}$ can adapt. The following theorem shows adaptivity is impossible to achieve even if Ω_0 only contains two elements.

THEOREM 3. *Fix any two positive Hölder smoothness parameters $\alpha>\beta>0$, parameters $L_\alpha, L_\beta, u_1(\alpha), u_1(\beta), u_2(\alpha), u_2(\beta)>0$ and any T that is larger than some constant and s such that $s < T^{\frac{\alpha-\beta}{8(2\alpha+1)(2\beta+1)}}/\ln(T)$. Suppose an algorithm $\mathcal{A}$ achieves the near optimal regret $\tilde{O}(s\cdot T^{\frac{\alpha+1}{2\alpha+1}})$ over $\mathcal{P}(s,d,\alpha,L_\alpha,u_1(\alpha),u_2(\alpha))$, then there exists a constant $C>0$ independent of T, s and $\mathcal{A}$ such that*

$$\sup_{\mathbb{P}\in\mathcal{P}(s,d,\beta,L_\beta,u_1(\beta),u_2(\beta))} \mathbb{R}_T(\mathcal{A};\mathbb{P}) \ge C\cdot s\cdot T^{\frac{\beta+1}{2\beta+1}+\frac{\beta(\alpha-\beta)}{2(2\beta+1)^2(2\alpha+1)}}.$$

For an algorithm that can nearly achieve the minimax optimal regret over a class of smoother functions, this theorem establishes a lower bound on the maximum regret over a class of less smooth payoff functions. This lower bound can be decomposed as a product of two terms $\Theta(s\cdot T^{\frac{\beta+1}{2\beta+1}})$ and $\Theta(T^{\frac{\beta(\alpha-\beta)}{2(2\beta+1)^2(2\alpha+1)}})$. The first term $\Theta(s\cdot T^{\frac{\beta+1}{2\beta+1}})$ is the minimax regret over β-Hölder payoff functions and the second term can be viewed as the cost of smoothness misspecification. This cost is at least $\Theta(T^{\frac{\beta(\alpha-\beta)}{2(2\beta+1)^2(2\alpha+1)}})$, where the power of T is alway positive for $\alpha>\beta>0$. Therefore this theorem implies that without further restrictions on the function class, adaptivity is impossible to achieve over any range of smoothness parameters. This phenomenon is connected to the lack of adaptivity for the construction of confidence intervals in nonparametric regression (Cai and Low (2004)).

The proof of Theorem 3 follows the nonadaptivity theory for nonparametric SCAB (Locatelli and Carpentier (2018), Liu, Wang and Singh (2021)). Such a proof is generally based on proving two functions f of different smoothness can hardly be distinguished. In the additive setting, one cannot directly use the same calculations for the one-dimensional case as in Locatelli and Carpentier (2018) and Liu, Wang and Singh (2021). This is because for two different additive functions, even though each component of these two functions cannot be distinguished, these two functions themselves may still be distinguished, since they are the sum of s different component functions. In all, our proof follows the idea of Locatelli and Carpentier (2018) and generalizes its technique to the additive setting. Adaptivity in bandits has attracted much recent attention (Locatelli and Carpentier (2018), Hadiji (2019), Gur, Momeni and Wager (2021)). We shall give a more detailed review in the discussion section.

4. Adaptivity under self-similarity. The fact that the smoothness parameter is typically unknown makes adaptivity critically important for an algorithm to be practically useful. We now consider adaptivity under an additional structural assumption – self-similarity. We first introduce the concept of self-similarity in Section 4.1 and then present a new data-driven algorithm in Section 4.2 and show in Section 4.3 that, under the self-similar assumption, it adapts to smoothness with considerable generality.

4.1. *Self-similarity.* Before formally introducing the self-similarity assumption, we first introduce some notations. For any positive integer p and a function $g(\cdot)$, let $\mathrm{Poly}(p)$ denote the set of all polynomials of degree less than or equal to p and define $\Gamma_p^U g(\cdot)$ to be the L_2-projection of the function $g(\cdot)$ onto $\mathrm{Poly}(p)$ over some interval U:

$$\Gamma_p^U g(x) := q(x), \quad \text{s.t. } q = \arg\min_{q \in \mathrm{Poly}(p)} \int_U |g(u) - q(u)|^2 \, du.$$

For any positive integer c, let $\mathcal{V}_c = \{[\frac{i}{2^c}, \frac{i+1}{2^c}] : i = 0, 1, \ldots, 2^c - 1\}$. We introduce the self-similarity definition as follows.

DEFINITION 3. A function $g : [0, 1] \to R$ is self-similar with parameters β, $p \in Z^+$, $M_1 \in R^+$, $M_2 \in \mathbb{R}^+$ if for any integer $c > M_1$,

$$\max_{V \in \mathcal{V}_c} \sup_{x \in V} |\Gamma_p^V g(x) - g(x)| \geq M_2 2^{-c\beta}.$$

Self-similar condition gives a lower bound on the maximum error for the approximation of a function g by piecewise polynomial functions. Recall that the Hölder condition requires a function to be well approximated by piecewise polynomial functions. Hence, self-similar condition can be viewed as a dual condition of the Hölder condition.

The following is an example of self-similar functions. Let $C_1, , C_2 > 0$ and $0 < \beta < 1$ be three constants. Define $\mathcal{S}_0 = \{a \cdot x^\beta + g : a \in \mathbb{R}, |a| \geq C_1, g \in \mathcal{C}^1, \|g'\|_\infty \leq C_2\}$. Then $\mathcal{S}_0$ is the function class of a nondiminishing β-Hölder function $a \cdot x^\beta$ plus an arbitrary smoother function g. The following lemma proves that $\mathcal{S}_0$ is a class of self-similar functions.

LEMMA 3. *$\mathcal{S}_0$ is self-similar with parameters β, $p = 0$ and some constants $M_1, M_2 > 0$.*

ASSUMPTION 4. We assume there exists $j \in \{1, \ldots, d\}$ such that f_j, the jth component of mean reward function, is a self-similar function with some parameters β, p, M_1, M_2.

Self-similarity has been used in nonparametric regression to enable adaptivity for the construction of confidence intervals (Picard and Tribouley (2000), Giné and Nickl (2010)). Here

we assume at least one of the component of the mean payoff is self-similar and do not require it to be known. We shall show that this assumption enables adaptivity with considerable generality. Let $\mathcal{P}_0(\beta, L, s, d, u_1, u_2, p, M_1, M_2)$ denote all of $\mathbb{P}$ that satisfy assumptions 1, 2, 3 and 4. The minimax regret over this function class is then defined as

$$\inf_{\mathcal{A}} \sup_{\mathbb{P}\in\mathcal{P}_0(\beta,L,s,d,u_1,u_2,p,M_1,M_2)} R_T(\mathcal{A};\mathbb{P}),$$

where $\mathcal{A}$ denotes an arbitrary algorithm.

4.2. *Adaptive algorithm.* We now present a new algorithm devised for self-similar payoff functions. The algorithm has three input parameters: the total number of rounds T, the lower bound $\beta_{\min}$ and upper bound $\beta_{\max}$ for the Hölder smoothness so $(\beta_{\min}, \beta_{\max})$ is the range for β. This procedure can be divided into two steps. The first step estimates the true smoothness β and then the second step simply calls Algorithm 1 with the estimated Hölder smoothness $\hat{\beta}$. The procedure is summarized in Algorithm 2.

In the first step, Algorithm 2 aims at estimating β with the help of self-similarity condition. Note the self-similarity condition together with the Hölder assumption gives a tight lower bound and upper bound for the bias of the local polynomial regression. To utilize this property, in Algorithm 2 we fit local polynomial regressions to estimate each f_j two times. In both times we divide $[0, 1]$ into many intervals and use local polynomial regression to estimate each f_j in each interval separately. The first step of Algorithm 2 thus has two epochs. In the first epoch, $[0, 1]$ is partitioned into more intervals and the estimation bias is thus smaller. In the second epoch, $[0, 1]$ is partitioned into much fewer intervals and the estimation bias is thus much larger. We pull arms enough times so that the standard deviation is dominated

Algorithm 2: $\mathcal{A}^2(T, \beta_{\min}, \beta_{\max})$

Input: Total number of rounds T, minimal Hölder smoothness $\beta_{\min}$ and maximal Hölder smoothness $\beta_{\max}$

0 Set local polynomial regression degree $l = w(\beta_{\max})$.

0 Set $k_1 = \frac{1}{10\beta_{\max}+5}$, $K_1 = 2^{\lfloor k_1 \cdot \ln_2^T \rfloor}$, $k_2 = \frac{1}{10\beta_{\max}+10}$, $K_2 = 2^{\lfloor k_2 \cdot \ln_2^T \rfloor}$, $r_0 = 0$.

for $i \in \{1, 2\}$ **do**

0 Set constant $T_i = \lfloor T^{\frac{1}{2}+k_i} \rfloor$.
Pull arms T_i times independently from the uniform distribution on $[0, 1]^d$.
For each $j = 1, \ldots, d$ and $m = 1, \ldots, K_i$, let the samples where the jth element of covariates is in $[\frac{m-1}{K_i}, \frac{m}{K_i})$ be
$\mathbb{O}'_{i,j,m,0} = \{(X_t, Y_t) : r_{i-1} < t \le r_{i-1} + T'_i, X_t^{(j)} \in [\frac{m-1}{K_i}, \frac{m}{K_i})\}$. Log these samples with outcomes and only the jth elements of covariates
$\mathbb{O}'_{i,j,m} = \{(X_t^{(j)}, Y_t) : (X_t, Y_t) \in \mathbb{O}_{i,j,m,0}\}$.
Update the round counter $r_i = r_{i-1} + T'_i$.

0 For $j = 1, \ldots, d$, and $m = 1, \ldots, K_i$, fit local polynomial regression on $[\frac{m-1}{K_i}, \frac{m}{K_i})$ with $\mathbb{O}'_{i,j,m}$ and construct estimate $\hat{f}_{i,j}(x)$ of $f_j(x)$ for each $x \in [\frac{m-1}{K_i}, \frac{m}{K_i})$ as in (5).

0 Let $\hat{\beta} = -\frac{\ln(\max_{1\le j\le d} \|\hat{f}_{2,j} - \hat{f}_{1,j}\|_\infty)}{k_2 \cdot \ln(T)} - \frac{\ln(\ln(T))}{\ln(T)}$.

0 Call $\mathcal{A}^1(T - T'_1 - T'_2, \hat{\beta}, l)$.

by the bias. Then the maximal difference between the two polynomial regression estimates should be roughly of the same order as the larger bias. Then we can compare this difference with the bound provided by the self-similarity condition to estimate the smoothness parameter β.

Specifically, let $k_1 = \frac{1}{10\beta_{\max}+5}$, $K_1 = 2^{\lfloor k_1 \cdot \ln_2^T \rfloor}$, $k_2 = \frac{1}{10\beta_{\max}+10}$, $K_2 = 2^{\lfloor k_2 \cdot \ln_2^T \rfloor}$, and $T_i' = \lfloor T^{\frac{1}{2}+k_i} \rfloor$, $i = 1, 2$. For each $i \in \{1, 2\}$, the algorithm pulls arms independently T_i' times from the uniform distribution on $[0, 1]^d$ in the ith epoch. Then the algorithm records the observations whose jth covariates falling into $[\frac{m-1}{K_i}, \frac{m}{K_i})$ in the ith epoch as $\mathbb{O}'_{i,j,m,0}$. Let the dataset only containing the outcomes and the jth covariates of $\mathbb{O}'_{i,j,m,0}$ be $\mathbb{O}'_{i,j,m}$. Then each component $\hat{f}_j$ can be estimated by local polynomial regression:

$$\hat{f}_{i,j}(x) = \hat{f}\left(x; \mathbb{O}'_{i,j,m}, w(\beta_{\max}), \left[\frac{m-1}{K_i}, \frac{m}{K_i}\right]\right), \tag{5}$$

for each $i \in \{1, 2\}$, $j \in \{1, \ldots, d\}$, $m \in \{1, \ldots, K_i\}$ and $x \in [\frac{m-1}{K_i}, \frac{m}{K_i})$. Finally, we estimate the smoothness by

$$\hat{\beta} = -\frac{(10\beta_{\max} + 10) \cdot \ln(\max_{1 \le j \le d} \| \hat{f}_{2,j} - \hat{f}_{1,j} \|_\infty)}{\ln(T)} - \frac{\ln(\ln(T))}{\ln(T)}.$$

The following proposition shows $\hat{\beta}$ converges to β with the rate $O_p(\frac{\ln(\ln(T))}{\ln(T)})$.

PROPOSITION 2. *Suppose $s \le T^{\frac{1}{20}}$ and $\ln^8(d) < C \cdot T$ for some sufficiently small constant $C > 0$. There exists a constant $C_3 > 0$ such that with probability at least $1 - O(e^{-C_3 \ln^2 T})$,*

$$\hat{\beta} \in \left[\beta - \frac{(10\beta_{\max} + 12)\ln(\ln(T))}{\ln(T)}, \beta\right].$$

4.3. *Theoretical results.* We now turn to the theoretical properties of the adaptive algorithm given in Section 4.2. Note that the function class with the self-similarity assumption $\mathcal{P}_0(\beta, L, s, d, u_1, u_2, p, M_1, M_2)$ is a subset of $\mathcal{P}(s, d, \beta, L, u_1, u_2)$, therefore the minimax regret under the self-similar condition is smaller than or equal to the general minimax regret.

LEMMA 4. *If $s < T^{\frac{\beta(\beta+1)}{(4\beta+1)^2}}$,*

$$\begin{aligned}\inf_{\mathcal{A}} \sup_{\mathbb{P} \in \mathcal{P}_0(\beta,L,s,d,u_1,u_2,p,M_1,M_2)} R_T(\mathcal{A}; \mathbb{P}) &\le \inf_{\mathcal{A}} \sup_{\mathbb{P} \in \mathcal{P}(s,d,\beta,L,u_1,u_2)} R_T(\mathcal{A}; \mathbb{P}) \\ &\le O\big(s \cdot \ln^3(T) \cdot T^{\frac{\beta+1}{2\beta+1}}\big).\end{aligned}$$

The following lower bound shows that restricting the payoff function class by the self-similarity condition does not make the problem easier.

THEOREM 4. *For any positive parameters $\beta, u_1, u_2, p = w(\beta), M_1, L > 0$, there exists a constant $M_2 > 0$ that only depends on β, u_1, u_2, L satisfying*

$$\inf_{\mathcal{A}} \sup_{\mathbb{P} \in \mathcal{P}_0(\beta,L,s,d,u_1,u_2,p,M_1,M_2)} R_T(\mathcal{A}; \mathbb{P}) \ge C \cdot s \cdot T^{\frac{\beta+1}{2\beta+1}},$$

where $C > 0$ is a constant depending only on the parameters of the class $\mathcal{P}_0$ and not on $\mathcal{A}$, T, d.

This lower bound together with Lemma 4 shows that the minimax regret with the self-similar assumption is $\tilde{\Theta}(s \cdot T^{\frac{\beta+1}{2\beta+1}})$, the same as the general minimax regret. This lower bound only applies to the case when M_2 is sufficiently small. This is unavoidable because this function class is empty if L is too small compared to M_2.

THEOREM 5. *Let* $0 < \beta_{\min} < \beta_{\max}$, $s \le T^{\frac{1}{20}} \wedge T^{\frac{\beta(\beta+1)}{(4\beta+2)^2}}$ *and* $\ln^8(d) < C \cdot T$ *for some small enough constant* $C > 0$. *Algorithm 2 satisfies, for all* $\beta \in (\beta_{\min}, \beta_{\max})$, *there exist constants* $C_1, C_2 > 0$ *only depending on the parameters of* $\mathcal{P}_0$ *such that*

$$\sup_{\mathbb{P} \in \mathcal{P}_0(\beta, L, s, d, u_1, u_2, w(\beta_{\max}), M_1, M_2)} R_T\big(\mathcal{A}^2; \mathbb{P}\big) \le C_1 \cdot s \cdot \ln^{3+C_2}(T) \cdot T^{\frac{\beta+1}{2\beta+1}}.$$

This theorem shows that Algorithm 2 adaptively achieves the minimax regret $\tilde{O}(s \cdot T^{\frac{\beta+1}{2\beta+1}})$ simultaneously for all $\beta \in (\beta_{\min}, \beta_{\max})$, where the poly-logarithmic terms in $\tilde{\Theta}(\cdot)$ here depends on T and not on s, d. Compared to Theorem 2, the upper bound of the regret for Algorithm 2 is further multiplied by a poly-logarithmic term $\ln^{C_2}(T)$. This extra term comes from using estimated smoothness $\hat{\beta}$ instead of the true β and thus can be viewed as the cost of smoothness adaptation.

5. Minimax regret under additional superlevel-set assumption. In the related literature, additional assumptions that restrict the measure of superlevel sets of f (or equivalently the sublevel set in optimization) is often considered (Minsker (2013), Wang, Balakrishnan and Singh (2019), Locatelli and Carpentier (2018), Zhao and Lai (2021)). These assumptions capture the difficulty of finding the maximizer of a function. If the superlevel set is restricted to be smaller, then the maximizer should have less potential competitors and one can expect a good algorithm to perform better. In this section, we study the problem under an additional superlevel-set assumption.

We first introduce some notation. For any measurable set $V \subset [0, 1]$, let $m(V)$ be the Lebesgue measure of V. For a set $V \subset [0, 1]$ that equals a union of intervals and any $\epsilon > 0$, define the "interval packing number" of V to be $\mathcal{N}(V, \epsilon) = \min K$ such that there exist intervals $V_{(1)}, \ldots, V_{(K)}$ with

$$V = \bigcup_{i=1}^{K} V_{(i)} \quad \text{and} \quad m(V_{(i)}) \le \epsilon, \quad i = 1, \ldots, K.$$

For any function g and its maximizer x^*, The ϵ-superlevel set of g is $\{x \in [0, 1] : g(x^*) - g(x) \le \epsilon\}$. We denote this set by $\mathcal{L}(f, \epsilon)$.

ASSUMPTION 5. There exists a constant $C_l > 0$ such that for any nonzero f_j, any $\epsilon_1, \epsilon_2 \in (0, 1)$

$$\mathcal{N}\big(\mathcal{L}(f_j, \epsilon_1), \epsilon_2\big) \le C_l \cdot \epsilon_1^{\gamma} / \epsilon_2 + 1.$$

Let $\mathcal{P}(s, d, \beta, \gamma, L, C_l, u_1, u_2)$ denote all of $\mathbb{P}$ that satisfy Assumptions 1, 2, 3 and 5. Then the minimax regret is defined by

$$\inf_{\mathcal{A}} \sup_{\mathbb{P} \in \mathcal{P}(s, d, \beta, \gamma, L, C_l, u_1, u_2)} R_T(\mathcal{A}; \mathbb{P}).$$

The following two theorems establish the lower and the upper bounds for the minimax regret under the new assumption.

THEOREM 6. *For any positive parameters β, L, u_1, u_2, γ, C_l that satisfy $\gamma\beta \le 1$ and C_l is greater than some constant that is only determined by β, L, γ, and the number of rounds T, there exists a constant $C > 0$ depending on β, L, u_1, u_2, γ, C_l only and not on T, s, d such that*

$$\inf_{\mathcal{A}} \sup_{\mathbb{P}\in\mathcal{P}(s,d,\beta,\gamma,L,C_l,u_1,u_2)} R_T(\mathcal{A};\mathbb{P}) \ge C \cdot s \cdot T^{\frac{\beta+1-\beta\gamma}{2\beta+1-\beta\gamma}}.$$

This theorem establishes a lower bound for the minimax regret. It requires C_l to be greater than a constant. Such requirement is unavoidable because the function space $\mathcal{P}(s, d, \beta, \gamma, L, C_l, u_1, u_2)$ is empty if C_l is too small.

We develop an algorithm that achieves (near) minimax optimal regret. Similar to Algorithm 1, this algorithm also has four steps in each epoch: reallocating, pulling, fitting, and eliminating, and maintains feasible regions recursively. The main difference is in the design of T_i, the number of rounds pulled in each epoch. In Algorithm 1, T_i is proportion to the upper bound of $g_{i,j}$, the number of intervals in each feasible region. Under the additional assumption the upper bound of $g_{i,j}$ is reduced. Therefore, T_i is changed correspondingly. For reasons of space, the details of this algorithm are given in the Supplementary Material (Cai and Pu (2022))

THEOREM 7. *Suppose $\beta\gamma \le 1$. Then there exists an algorithm $\mathcal{A}$ such that for any $s \le T^{\frac{\beta(\beta+1-\beta\gamma)}{(4\beta+1-\beta\gamma)(4\beta+2)}}$ and any $\mathbb{P} \in \mathcal{P}(s, d, \beta, \gamma, L, C_l, u_1, u_2)$,*

$$R_T(\mathcal{A};\mathbb{P}) \le C \cdot s \cdot \ln^{3+\frac{3\beta\gamma}{2\beta+1-\beta\gamma}}(T) \cdot T^{\frac{\beta+1-\beta\gamma}{2\beta+1-\beta\gamma}},$$

where the constant C depends on β, L, u_1, u_2, l, γ, C_l but not on $\mathbb{P}$, d, s, T.

These two theorems together yield the minimax regret $\tilde{\Theta}(s \cdot T^{\frac{\beta+1-\beta\gamma}{2\beta+1-\beta\gamma}})$. If $\gamma = 0$, then the superlevel-set assumption is almost a null assumption and in this case the regret here recovers the minimax regret $\tilde{\Theta}(s \cdot T^{\frac{\beta+1}{2\beta+1}})$ given in Section 2. The theorems also show that the larger the γ the smaller the regret. This is due to the fact that larger γ corresponds to smaller superlevel sets and consequently it is easier for an algorithm to find the maximum region. The minimax regret here is also dimension-free. So the dimension-free phenomenon only relies on the additive structure and the action set and remains true for the function classes satisfying Assumption 5.

In both theorems, we only consider the case where $\beta\gamma \le 1$. However the full possible range of β, γ should be $\{\beta\gamma \le 1\} \cup \{\beta > 1, \frac{1}{\beta} < \gamma \le 1\}$ (Audibert and Tsybakov (2007)). It would be interesting to explore the problem in the case where $\beta > 1$, $\frac{1}{\beta} < \gamma \le 1$. Unfortunately, the technique we use in the lower bound can not be easily extended to this case due the difficulty of constructing examples that satisfy the assumptions. We leave this interesting problem for future study.

5.1. *Adaptivity to γ.* The algorithm considered in Theorem 7 requires the value of γ, which is typically unknown in practice. We now turn to the question of adaptivity under Assumption 5. The following theorem considers the nonsparse setting ($s = d$) and justifies the adaptivity to γ in this case.

THEOREM 8. *Let $s = d \le T^{\frac{\beta}{4(4\beta+2)}}$ and $\gamma \le \frac{1}{\beta}$. Then there exists an adaptive algorithm $\mathcal{A}$ that does not depend on γ and a constant $C > 0$ depending on β, L, u_1, u_2, l, γ, C_l but not on d, s, T such that*

$$\sup_{\mathbb{P}\in\mathcal{P}(s,d,\beta,\gamma,L,C_l,u_1,u_2)} R_T(\mathcal{A};\mathbb{P}) \le C \cdot s \cdot \ln^{3+\frac{3\beta\gamma}{2\beta+1-\beta\gamma}}(T) \cdot T^{\frac{\beta+1-\beta\gamma}{2\beta+1-\beta\gamma}}.$$

We now consider the sparse case $s \le d$. Note that even with the knowledge of β, there are two intrinsic adaptivity problems: adaptivity to s and to γ. The following theorem shows the simultaneous adaptivity to both parameters is possible under mild conditions.

THEOREM 9. *Let* $s \le \frac{1}{2}(T^{\frac{1}{14}} \wedge T^{\frac{\beta}{4(4\beta+2)}})$ *and* $\ln(d) \le \frac{T^{\frac{1}{5}}}{\ln(T)}$. *Then there exists an adaptive algorithm* $\mathcal{A}$ *that depends on* β *and not on* γ, s *such that for any* $\gamma \in (0, \frac{1}{\beta}]$,

$$\sup_{\mathbb{P}\in\mathcal{P}(s,d,\beta,\gamma,L,C_l,u_1,u_2)} R_T(\mathcal{A};\mathbb{P}) \le C \cdot s \cdot \ln^{3+\frac{3\beta\gamma}{2\beta+1-\beta\gamma}}(T) \cdot T^{\frac{\beta+1-\beta\gamma}{2\beta+1-\beta\gamma}},$$

where $C > 0$ *is a constant depending on* β, L, u_1, u_2, l, γ, γ_0, C_l *and not on* d, s, T.

This theorem shows one can adapt to s and γ simultaneously under the additional restriction of $\ln(d) \le \frac{T^{\frac{1}{5}}}{\ln(T)}$. This restriction comes from the step for estimating the effective dimensions in the adaptive algorithm, which is needed for choosing the correct number of rounds to pull in each epoch. Although we have this restriction on d, the regret is still dimension-free. This still highlights the substantial difference between high-dimensional bandits and high-dimensional regression problems.

Theorem 8 shows that it is possible to adapt to all values of γ except 0. This is because for any $\gamma > 0$, Assumption 5 requires the superlevel set of each nonzero component to shrink as the depth of the superlevel set goes to zero, which helps estimate the effective dimensions. When $\gamma = 0$, the superlevel set does not shrink, which fails to enable the estimation of effective dimensions.

6. Discussion. We studied in the present paper the minimax optimality for the d-dimensional additive β-Hölder SCAB with sparsity s for the full range of smoothness levels $0 < \beta < \infty$ and sparsity levels $1 \le s \le d$. We establish the minimax regret to be $\tilde{\Theta}(s \cdot T^{\frac{\beta+1}{2\beta+1}})$. The problem of adaptivity is also investigated. It is shown that adaptation to the sparsity is free but adaptation to the smoothness is in general impossible. Under the additional self-similarity assumption, a data-driven algorithm is introduced and shown to achieve the minimax rate adaptively up to a logarithmic factor over a range of smoothness levels.

In this paper it is shown that adaptivity to smoothness is general impossible. Such nonadaptivity phenomenon is common in related literature. For example, Locatelli and Carpentier (2018) and Liu, Wang and Singh (2021) show nonadaptivity in nonparametric continuum-armed bandits. Gur, Momeni and Wager (2021) proves nonadaptivity in nonparametric contextual bandits. Our proof is related to Locatelli and Carpentier (2018) but is different. The proof of Locatelli and Carpentier (2018) for the one-dimensional case cannot be directly applied here. This is because compared to two one-dimensional functions with different smoothness, in the current setting, two additive functions with different smoothness are harder to distinguish.

In this paper, we show adaptivity can be achieved under self-similarity. Some other conditions have also been considered in continuum-armed bandits. Combes and Proutiere (2014) consider one-dimensional nonparametric SCAB and prove adaptivity is possible if $\beta\gamma = 1$ and the functions is unimodal. In general d-dimensional SCAB, the case $\beta\gamma = d$, which can be thought of as an additional condition, has been partially treated in Bull (2015) for the special class of zooming continuous functions. In this setting, Bull (2015) introduced an adaptive strategy such that its expected cumulative regret is parametric regret.

Some other structural conditions such as monotonicity, convexity, and concavity also seem to be worth considering since each of them can enable the construction of adaptive confidence intervals in one-dimensional nonparametric regression (Cai, Low and Xia (2013)).

However, imposing either one of these three structural assumptions in the current SCAB problem changes the minimax regret completely and significantly reduces the complexity of the problem. Under either the monotonicity or convexity assumption in the one-dimensional case, the maximizer must be 0 or 1, which reduces the problem into a two-armed bandit problem. In this case the minimax regret is known to be $\tilde{\Theta}(\sqrt{T})$ (Audibert and Bubeck (2010); Auer et al. (2002); Auer, Cesa-Bianchi and Fischer (2002)). On the other hand, under the concavity assumption, Agarwal et al. (2013) shows that the minimax regret is also reduced to $\tilde{\Theta}(\sqrt{T})$. Therefore, under either one of these three assumptions the minimax regret is the parametric rate and thus independent of smoothness.

A novel approach to deal with nonadaptive bandit problems has been proposed by Hadiji (2019). Instead of imposing stronger conditions, Hadiji (2019) considers admissibility in the minimax sense for one-dimensional nonparametric SCAB. For each algorithm $\mathcal{A}$, suppose the maximum regret of this algorithm over all β-Hölder functions is bounded by $O(T^{\theta(\beta)})$, then $\theta(\cdot)$ can be viewed as its rate function. A rate function is admissible if it does not dominate any other different rate functions. Hadiji (2019) not only provides an admissible algorithm but also determines the set of all admissible rate functions. Based on our knowledge this is the first time one considers (minimax) admissibility in nonadaptive problems. This (minimax) admissibility approach has obvious advantages in weaker assumptions. It is interesting to generalize this new criterion to other nonadaptive problems.

Finally, we outline several future directions. First and foremost, it would be interesting to consider the problem in the adversarial setting. Whether it is possible to construct a nontrivial algorithm in the adversarial additive SCAB is still an open question. Second, it is also interesting to consider the contextual additive SCAB. For the nonparametric SCAB, the results in Lu, Pál and Pál (2009) and Slivkins (2011) show that in the nonsmooth case the minimax regret heavily depends on the dimensions of the context and action spaces and it can be close to linear if both dimensions are high. The rate of minimax regret may be improved under the additive models. Third, the additive SCAB under general smoothness assumptions can be interesting. In this paper we consider the setting that all the components of the mean reward have the same smoothness β. This can be extended to the case where different components have different smoothness levels. Suppose the jth component f_j is a β_j-Hölder function. We conjecture that with a modification of the techniques developed in the present paper it can be shown that the minimax regret in that more general setting is $\tilde{\Theta}(\sum_{j=1}^{d} T^{\frac{\beta_j+1}{2\beta_j+1}})$.

7. Proofs. In this section, we present the proof for Theorem 4. Note that Theorem 1 is strictly weaker than Theorem 4 and so it follows from Theorem 4. For reasons of space, the proofs of other main and technical results are given in the Supplementary Material Cai and Pu (2022).

We first consider the case where $s = d$ and then prove the general case ($d \geq s$) is at least as hard as this special case.

First Step: In this step, we only consider the case where $s = d$. Before giving the proof, we first state three useful lemmas.

LEMMA 5. *Let $Y_{[1]}$, $Y_{[2]}$ be two random variables with distributions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$. Then the KL divergence of Y_1, Y_2 is $\frac{(\mu_1-\mu_2)^2}{2\sigma^2}$.*

LEMMA 6. *Let $\mathcal{Q}_1$ and $\mathcal{Q}_2$ be two probability measures on the same σ-algebra Σ. Then for any event $A \in \Sigma$, we have*

$$KL(\mathcal{Q}_1 \| \mathcal{Q}_2) \geq 2\big(\mathcal{Q}_1(A) - \mathcal{Q}_2(A)\big)^2.$$

LEMMA 7. *For each $L, \beta > 0$, there exists a function $g : [0,1] \to [0,1]$ satisfying that $g \in \mathcal{H}(\beta, L)$; g is self-similar with constants β, $p = w(\beta)$, $M_1 = 0$ and some positive number $M_2 > 0$; g has a unique maximizer at $\frac{1}{2}$; and for any $x \in \{0,1\}$, $t \in \{0,1,\ldots,w(\beta)\}$ it holds that $g^{(t)}(x) = 0$.*

Let

$$Y_x = f_1(x^{(1)}) + \cdots + f_d(x^{(d)}) + N(0,\sigma^2). \tag{6}$$

where $N(0,\sigma^2)$ denotes a zero-mean normal random variable with variance σ^2. Let σ be small enough such that assumption 3 holds. We introduce some notations and definitions. Let

$$\phi(x) := \begin{Bmatrix} C_\sigma \cdot g, & x \in [0,1], \\ 0, & \text{otherwise}, \end{Bmatrix},$$

where $C_\sigma = \frac{\sigma}{4} \wedge \frac{1}{2}$ is a constant and g is a function given by Lemma 7. Let positive integer $k = \lfloor 10 \cdot T^{\frac{1}{2\beta+1}} \rfloor$. Let $\phi_{k,-1}(x) = -\phi(2x-1)$. Define a function $\phi_{k,i}(\cdot) : [0,1] \to [-1,1]$ by $\phi_{k,i}(x) = (2k)^{-\beta}\phi(2kx - i) + \phi_{k,-1}(x) \in \mathcal{H}(\beta, L)$, $\forall x \in [0,1]$ for each $i = 0,\ldots,k-1$. Then $\phi_{k,i}$ is self-similar with parameters β, $p = w(\beta)$, $M_1 = 1$ and some constant $M_2 > 0$.

For functions $f_1,\ldots,f_d$, constant σ, and algorithm $\mathcal{A}$, let the $\mathbb{P}(\mathcal{A}; f_1,\ldots,f_d,\sigma^2)$ denote the probability measure of $\{(X_t, Y_t), t = 1,\ldots,T\}$, provided Y_x is given by (6). We consider two cases $\mathbb{P}(\mathcal{A}; f_1,\ldots,f_{j-1},\phi_{k,-1},f_{j+1},\ldots,f_d,\sigma^2)$ and $\mathbb{P}(\mathcal{A}; f_1,\ldots,f_{j-1},\phi_{k,i},f_{j+1},\ldots,f_d,\sigma^2)$ where all component functions are the same except for f_j for any $j = 1,\ldots,d$ and $i = 0,\ldots,k-1$. Let $\mathbb{E}_{\mathbb{P}}Z$ denote the expectation of Z where the probability measure is given by $\mathbb{P}$. For simplicity, we let $\mathbb{P}_1 = \mathbb{P}(\mathcal{A}; f_1,\ldots,f_{j-1},-\phi(2x-1),f_{j+1},\ldots,f_d,\sigma^2)$ and $\mathbb{P}_2 = \mathbb{P}(\mathcal{A}; f_1,\ldots,f_{j-1},\phi_{k,i},f_{j+1},\ldots,f_d,\sigma^2)$. Define $Z_{k,i,j} = \sum_{t=1}^{T} \mathbb{1}\{X_t^{(j)} \in [\frac{i}{2k}, \frac{i+1}{2k})\}$. We first prove for each $i = 0,\ldots,k-1$ and $j = 1,\ldots,d$,

$$\mathbb{E}_{\mathbb{P}_2} \sum_{t=1}^{T} \left[f_j(x_*^{(j)}) - f_j(X_t^{(j)}) \right] \geq C_\sigma \cdot g\left(\frac{1}{2}\right) \cdot \frac{T^{\frac{\beta+1}{2\beta+1}}}{20^{\beta+1}} \cdot \mathbb{1}\{\mathbb{E}_{\mathbb{P}_1} Z_{k,i,j} \leq 2T/k\}, \tag{7}$$

where $f_j = \phi_{k,i}$ under $\mathbb{P}_2$.

The inequality (7) obviously holds if $E_{\mathbb{P}_1} Z_{k,i,j} > 2T/k$. Therefore, we only need to consider the case where $E_{\mathbb{P}_1} Z_{k,i,j} \leq 2T/k$.

The KL divergence between $\mathbb{P}_1$ and $\mathbb{P}_2$ can be decomposed as

$$KL(\mathbb{P}_1 \| \mathbb{P}_2) = \mathbb{E}_{\mathbb{P}_1} \sum_{t=1}^{T} KL\big(\mathbb{P}_1(Y_t|X_t) \| \mathbb{P}_2(Y_t|X_t)\big),$$

where $\mathbb{P}_1(Y_t|X_t)$ and $\mathbb{P}_2(Y_t|X_t)$ denote the conditional distribution of Y_t given X_t under $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively. Note $\mathbb{P}_1(Y_t|X_t)$ and $\mathbb{P}_2(Y_t|X_t)$ are normal distributions with the same variance. Therefore, we have $KL(\mathbb{P}_1(Y_t|X_t) \| \mathbb{P}_2(Y_t|X_t)) = \frac{[\phi_{k,i}(X_t^{(j)}) - \phi_{k,-1}(X_t^{(j)})]^2}{2\sigma^2}$ by Lemma 5.

Then we have

$$\begin{aligned} KL(\mathbb{P}_1 \| \mathbb{P}_2) &= \mathbb{E}_{\mathbb{P}_1} \sum_{t=1}^{T} KL\big(\mathbb{P}_1(Y_t|X_t) \| \mathbb{P}_2(Y_t|X_t)\big) \\ &= \mathbb{E}_{\mathbb{P}_1} \left\{ \sum_{t=1}^{T} KL\big(\mathbb{P}_1(Y_t|X_t) \| \mathbb{P}_2(Y_t|X_t)\big) \mathbb{1}\left\{ X_t^{(j)} \in [\frac{i}{2k}, \frac{i+1}{2k}) \right\} \right. \end{aligned}$$

$$
\begin{aligned}
&+\sum_{t=1}^{T} KL\big(\mathbb{P}_1(Y_t|X_t)\|\mathbb{P}_2(Y_t|X_t)\big)\mathbb{1}\left\{X_t^{(j)} \notin [\frac{i}{2k}, \frac{i+1}{2k})\right\}\Bigg\} \\
&\le \mathbb{E}_{\mathbb{P}_1} \sum_{t=1}^{T} \frac{C_\sigma^2 \cdot k^{-2\beta}}{2\sigma^2} \cdot \mathbb{1}\left\{X_t^{(j)} \in [\frac{i}{2k}, \frac{i+1}{2k})\right\} \\
&= \mathbb{E}_{\mathbb{P}_1} \frac{C_\sigma^2 \cdot k^{-2\beta}}{2\sigma^2} \cdot \mathbb{E}_{\mathbb{P}_1} Z_{k,i,j} \\
&\le \mathbb{E}_{\mathbb{P}_1} \frac{C_\sigma^2 \cdot k^{-(2\beta+1)} \cdot T}{\sigma^2} < \frac{1}{5}.
\end{aligned}
$$

By Lemma 6, we have for any event A: $2(\mathbb{P}_1(A) - \mathbb{P}_2(A))^2 \le \frac{1}{5}$, which further implies $|\mathbb{P}_1(A) - \mathbb{P}_2(A)| \le \frac{1}{3}$. Let $A = \{Z_{k,i,j} > 4T/k\}$. Since $\mathbb{E}_{\mathbb{P}_1} Z_{k,i,j} \le 2T/k$, we have $\mathbb{P}_1(A) \le \frac{1}{2}$. If follows that $\mathbb{P}_2(A) \le \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$. We have

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}_2} \sum_{t=1}^{T} [f_j(x_*^{(j)}) - f_j(X_t^{(j)})] &\ge \mathbb{E}_{\mathbb{P}_2} \sum_{t=1}^{T} [f_j(x_*^{(j)}) - f_j(X_t^{(j)})] \mathbb{1}\left\{X_t^{(j)} \notin [\frac{i}{2k}, \frac{i+1}{2k})\right\} \\
&\ge \mathbb{E}_{\mathbb{P}_2} \sum_{t=1}^{T} f_j(x_*^{(j)}) \mathbb{1}\left\{X_t^{(j)} \notin [\frac{i}{2k}, \frac{i+1}{2k})\right\} \\
&= C_\sigma \cdot g\left(\frac{1}{2}\right) \cdot (2k)^{-\beta} \cdot \mathbb{E}_{\mathbb{P}_2}(T - Z_{k,i,j}) \\
&\ge \mathbb{E}_{\mathbb{P}_2} C_\sigma \cdot g\left(\frac{1}{2}\right) \cdot (T - Z_{k,i,j}) \frac{T^{-\frac{\beta}{2\beta+1}}}{20^\beta} \\
&\ge \mathbb{E}_{\mathbb{P}_2} C_\sigma \cdot g\left(\frac{1}{2}\right) \cdot (T - Z_{k,i,j}) \frac{T^{-\frac{\beta}{2\beta+1}}}{20^\beta} \mathbb{1}_{A^c} \\
&\ge C_\sigma \cdot g\left(\frac{1}{2}\right) \frac{T^{\frac{\beta+1}{2\beta+1}}}{2 \cdot 20^\beta} \mathbb{P}_2(A^c) \\
&\ge C_\sigma \cdot g\left(\frac{1}{2}\right) \cdot \frac{T^{\frac{\beta+1}{2\beta+1}}}{20^{\beta+1}}.
\end{aligned}
$$

This proves inequality (7).

With inequality (7), we are ready to prove the lower bound. Specifically, we define a prior of $f_1, \ldots, f_d$ and then prove a lower bound for the average performance of any algorithm under this prior. After that, the worst-case lower bound naturally follows. The prior is defined as follows.

Let the prior of f_j for each $j = 1, \ldots, d$ be

$$
f_j := \left\{ \begin{array}{cc} \phi_{k,0} & \text{with probability} \frac{1}{k} \\ \ldots & \ldots \\ \phi_{k,k-1} & \text{with probability} \frac{1}{k} \end{array} \right\},
$$

and the joint prior of $\{f_1, \ldots, f_d\}$ be the product of each f_j's prior. Let $\mathcal{D}_d$ denote this prior.

For simplicity, let $\mathbb{E}_{v_1,\ldots,v_d}$ denote taking expectation with respect to the probability measure $\mathbb{P}(\mathcal{A}; \phi_{k,v_1}, \ldots, \phi_{k,v_d}, \sigma^2)$ for any $(v_1, \ldots, v_d) \in \{0, \ldots, k-1\}^d$. For each $j = 1, \ldots, d$,

$v_j = -1$ and $v_i \in \{0, \ldots, k-1\}$ for $i \neq j$, let $\mathbb{E}_{v_1,\ldots,v_d}$ denote taking expectation with respect to the probability measure $\mathbb{P}(\mathcal{A}; \phi_{k,v_1}, \ldots, \phi_{k,v_{j-1}}, \phi_{k,-1}, \phi_{j+1}, \ldots, \phi_{k,v_d}, \sigma^2)$. The expected regret of $\mathcal{A}$ on the prior $\mathcal{D}_d$ is then given by

$$\begin{aligned}\mathbb{E}_{\mathbb{P}\sim\mathcal{D}_d} R_T(\mathcal{A}; \mathbb{P}) &= \frac{1}{k^d} \sum_{v_1=0}^{k-1} \cdots \sum_{v_d=0}^{k-1} \sum_{t=1}^{T} \sum_{j=1}^{d} \mathbb{E}_{v_1,\ldots,v_d}\big[f_j\big(x_*^{(j)}\big) - f_j\big(X_t^{(j)}\big)\big] \\ &\geq \frac{1}{k^d} \sum_{v_1=0}^{k-1} \cdots \sum_{v_d=0}^{k-1} \sum_{j=1}^{d} C_\sigma \cdot g\left(\frac{1}{2}\right) \cdot \frac{T^{\frac{\beta+1}{2\beta+1}}}{20^{\beta+1}} \\ &\quad \cdot \mathbb{1}\{\mathbb{E}_{v_1,\ldots,v_{j-1},-1,v_{j+1},\ldots,v_d} Z_{k,v_j,j} \leq 2T/k\},\end{aligned} \tag{8}$$

where the last inequality follows from (7). Note for each $v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_d$, we have

$$\begin{aligned}T &\geq \sum_{v_j=0}^{k-1} \mathbb{E}_{v_1,\ldots,v_{j-1},-1,v_{j+1},\ldots,v_d} Z_{k,v_j,j} \\ &\geq \frac{2T}{k} \sum_{v_j=0}^{k-1} \mathbb{1}\left\{\mathbb{E}_{v_1,\ldots,v_{j-1},-1,v_{j+1},\ldots,v_d} Z_{k,v_j,j} > \frac{2T}{k}\right\}.\end{aligned}$$

It follows that

$$\sum_{v_j=0}^{k-1} \mathbb{1}\left\{\mathbb{E}_{v_1,\ldots,v_{j-1},-1,v_{j+1},\ldots,v_d} Z_{k,v_j,j} > \frac{2T}{k}\right\} \leq \frac{k}{2},$$

which further implies

$$\sum_{v_j=0}^{k-1} \mathbb{1}\left\{\mathbb{E}_{v_1,\ldots,v_{j-1},-1,v_{j+1},\ldots,v_d} Z_{k,v_j,j} \leq \frac{2T}{k}\right\} \geq \frac{k}{2}.$$

Plugging this into (8) yields

$$\begin{aligned}\mathbb{E}_{\mathbb{P}\sim\mathcal{D}_d} R_T(\mathcal{A}; \mathbb{P}) &\geq \frac{1}{k^d} \sum_{j=1}^{d} \left\{\sum_{v_1=0}^{k-1} \cdots \sum_{v_{j-1}=0}^{k-1} \sum_{v_{j+1}=0}^{k-1} \cdots \sum_{v_d=0}^{k-1} C_\sigma \cdot g\left(\frac{1}{2}\right) \cdot \frac{T^{\frac{\beta+1}{2\beta+1}}}{20^{\beta+1}} \cdot \frac{k}{2}\right\} \\ &= \frac{1}{k^d} \cdot d \cdot k^{d-1} \cdot C_\sigma \cdot g\left(\frac{1}{2}\right) \cdot \frac{T^{\frac{\beta+1}{2\beta+1}}}{20^{\beta+1}} \cdot \frac{k}{2} \\ &= d \cdot C_\sigma \cdot g\left(\frac{1}{2}\right) \cdot \frac{T^{\frac{\beta+1}{2\beta+1}}}{2 \cdot 20^{\beta+1}}.\end{aligned}$$

Note in this case $s = d$, therefore we have $\mathbb{E}_{\mathbb{P}\sim\mathcal{D}_d} R_T(\mathcal{A}; \mathbb{P}) \geq s \cdot T^{\frac{\beta+1}{2\beta+1}}$. By definition we have

$$\sup_{\mathbb{P}\in\mathcal{P}(s,d,\beta,L,u_1,u_2)} R_T(\mathcal{A}; \mathbb{P}) \geq \mathbb{E}_{\mathbb{P}\sim\mathcal{D}_d} R_T(\mathcal{A}; \mathbb{P}).$$

Therefore, we complete the proof by combining the above two inequalities.

Second Step: Now we consider the general case. We shall prove for any $d \geq s$, we have

$$\inf_{\mathcal{A}} \sup_{\mathbb{P}\in\mathcal{P}(s,d,\beta,L,u_1,u_2)} R_T(\mathcal{A}; \mathbb{P}) \geq \inf_{\mathcal{A}} \sup_{\mathbb{P}\in\mathcal{P}(s,s,\beta,L,u_1,u_2)} R_T(\mathcal{A}; \mathbb{P}).$$

For any algorithm $\mathcal{A}_d$ for the d-dimensional case, we construct a corresponding algorithm $\mathcal{A}_s$ for the s-dimensional case. Note in each round, $\mathcal{A}_d$ outputs a d-dimensional vector. $\mathcal{A}_s$ takes the first s-dimensional vector as the arm to be pulled and keeps the whole vector as a fake "arm." $\mathcal{A}_s$ pulls the s-dimensional arm and observes Y_t. Then A_s uses the d-dimensional fake arms and the observed outcomes as the input of $\mathcal{A}_d$ in the next round to get a new d-dimensional vector as an output. Intuitively, A_s is constructed by restricting $\mathcal{A}_d$ to the s-dimensional space. For each $\mathbb{P}_s \in \mathcal{P}(s, s, \beta, L, u_1, u_2)$, we construct a corresponding example $\mathbb{P}_d$ in $\mathcal{P}(s, d, \beta, L, u_1, u_2)$ as follows. In $\mathbb{P}_d$, let the distribution of the outcome Y_1 given the arm X only depends on the first s elements of X, $X^{[s]} = (X^{(1)}, \ldots, X^{(s)})$, and is the same as the distribution of $Y_1(X^{[s]})$ in $\mathbb{P}_s$. This means only the first s components f_j, $j = 1, \ldots, s$ are nonzero. Then we have the joint distribution of $X_1^{[s]}, \ldots, X_T^{[s]}, Y_1, \ldots, Y_T$ created by $\mathcal{A}_d$ and $\mathbb{P}_d$ is the same as that of $X_1, \ldots, X_T, Y_1, \ldots, Y_T$ created by $\mathcal{A}_s$ and $\mathbb{P}_s$. We then have

$$R_T(\mathcal{A}_d; \mathbb{P}_d) = R_T(\mathcal{A}_s; \mathbb{P}_s).$$

Therefore, we have

$$\sup_{\mathbb{P} \in \mathcal{P}(s,d,\beta,L,u_1,u_2)} R_T(\mathcal{A}_d; \mathbb{P}) \geq \sup_{\mathbb{P} \in \mathcal{P}(s,s,\beta,L,u_1,u_2)} R_T(\mathcal{A}_s; \mathbb{P}).$$

Since for each $\mathcal{A}_d$ we can have such a $\mathcal{A}_s$, it follows that

$$\inf_{\mathcal{A}} \sup_{\mathbb{P} \in \mathcal{P}(s,d,\beta,L,u_1,u_2)} R_T(\mathcal{A}; \mathbb{P}) \geq \inf_{\mathcal{A}} \sup_{\mathbb{P} \in \mathcal{P}(s,s,\beta,L,u_1,u_2)} R_T(\mathcal{A}; \mathbb{P}).$$

Recall in the first step we prove that for any algorithm $\mathcal{A}$,

$$\sup_{\mathbb{P} \in \mathcal{P}(s,s,\beta,L,u_1,u_2)} R_T(\mathcal{A}; \mathbb{P}) \geq \Omega\big(s \cdot T^{\frac{\beta+1}{2\beta+1}}\big).$$

Therefore, we conclude that

$$\inf_{\mathcal{A}} \sup_{\mathbb{P} \in \mathcal{P}(s,d,\beta,L,u_1,u_2)} R_T(\mathcal{A}; \mathbb{P}) \geq \Omega\big(s \cdot T^{\frac{\beta+1}{2\beta+1}}\big).$$

Acknowledgments. We would like to thank the Associate Editor and the referees for their detailed and constructive comments which have helped to improve the presentation of the paper.

Funding. The research was supported in part by NSF Grant DMS-2015259 and NIH Grants R01-GM129781 and R01-GM123056.

SUPPLEMENTARY MATERIAL

Supplement to "Stochastic continuum-armed bandits with additive models: Minimax regrets and adaptive algorithm" (DOI: 10.1214/22-AOS2182SUPP; .pdf). The supplement is divided into four parts. Supplementary Material A contains the proofs of Theorems 2, 3, and 5, and Propositions 1 and 2. Supplementary Material B gives the proofs of Lemmas 1–7 and A.1–A.3. Supplementary Material C presents the proofs of Lemmas A.5–A.7. Finally, Supplementary Material D presents all the algorithms, results, and proofs related to the superlevel-set assumption.

REFERENCES

ABBASI-YADKORI, Y., PÁL, D. and SZEPESVÁRI, C. (2011). Improved algorithms for linear stochastic bandits. In *NIPS* **11** 2312–2320.

AGARWAL, A., FOSTER, D. P., HSU, D., KAKADE, S. M. and RAKHLIN, A. (2013). Stochastic convex optimization with bandit feedback. *SIAM J. Optim.* **23** 213–240. MR3033105 https://doi.org/10.1137/110850827

AGRAWAL, R. (1995). The continuum-armed bandit problem. *SIAM J. Control Optim.* **33** 1926–1951. MR1358102 https://doi.org/10.1137/S0363012992237273

AGRAWAL, S., AVADHANULA, V., GOYAL, V. and ZEEVI, A. (2019). MNL-Bandit: A dynamic learning approach to assortment selection. *Oper. Res.* **67** 1453–1485. MR4014580 https://doi.org/10.1287/opre.2018.1832

AUDIBERT, J.-Y. and BUBECK, S. (2010). Regret bounds and minimax policies under partial monitoring. *J. Mach. Learn. Res.* **11** 2785–2836. MR2738783

AUDIBERT, J.-Y. and TSYBAKOV, A. B. (2007). Fast learning rates for plug-in classifiers. *Ann. Statist.* **35** 608–633. MR2336861 https://doi.org/10.1214/009053606000001217

AUER, P. (2002). Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.* **3** 397–422. MR1984023 https://doi.org/10.1162/153244303321897663

AUER, P., CESA-BIANCHI, N. and FISCHER, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* **47** 235–256.

AUER, P., ORTNER, R. and SZEPESVÁRI, C. (2007). Improved rates for the stochastic continuum-armed bandit problem. In *Learning Theory. Lecture Notes in Computer Science* **4539** 454–468. Springer, Berlin. MR2397605 https://doi.org/10.1007/978-3-540-72927-3_33

AUER, P., CESA-BIANCHI, N., FREUND, Y. and SCHAPIRE, R. E. (2002). The nonstochastic multiarmed bandit problem. *SIAM J. Comput.* **32** 48–77.

AWERBUCH, B. and KLEINBERG, R. (2008). Online linear optimization and adaptive routing. *J. Comput. System Sci.* **74** 97–114. MR2364184 https://doi.org/10.1016/j.jcss.2007.04.016

BANKS, J. S. and SUNDARAM, R. K. (1992). Denumerable-armed bandits. *Econometrica* **60** 1071–1096. MR1180234 https://doi.org/10.2307/2951539

BLUM, A., BURCH, C. and KALAI, A. (1999). Finely-competitive paging. In 40*th Annual Symposium on Foundations of Computer Science* (*New York*, 1999) 450–457. IEEE Computer Soc., Los Alamitos, CA. MR1917584 https://doi.org/10.1109/SFFCS.1999.814617

BRESLER, G., CHEN, G. H. and SHAH, D. (2014). A latent source model for online collaborative filtering. *Adv. Neural Inf. Process. Syst.* **27** 3347–3355.

BUBECK, S., MUNOS, R., STOLTZ, G. and SZEPESVÁRI, C. (2011). $\mathcal{X}$-armed bandits. *J. Mach. Learn. Res.* **12** 1655–1695. MR2813150

BULL, A. D. (2015). Adaptive-treed bandits. *Bernoulli* **21** 2289–2307. MR3378467 https://doi.org/10.3150/14-BEJ644

CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395 https://doi.org/10.1214/16-AOS1461

CAI, T. T. and LOW, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805–1840. MR2102494 https://doi.org/10.1214/009053604000000049

CAI, T. T., LOW, M. G. and XIA, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *Ann. Statist.* **41** 722–750. MR3099119 https://doi.org/10.1214/12-AOS1068

CAI, T. T and PU, H. (2022). Supplement to "Stochastic continuum-armed bandits with additive models: Minimax regrets and adaptive algorithm." https://doi.org/10.1214/22-AOS2182SUPP

CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885 https://doi.org/10.1214/09-AOS752

COMBES, R. and PROUTIERE, A. (2014). Unimodal bandits: Regret lower bounds and optimal algorithms. In *International Conference on Machine Learning* 521–529. PMLR.

COPE, E. W. (2009). Regret and convergence bounds for a class of continuum-armed bandit problems. *IEEE Trans. Automat. Control* **54** 1243–1253. MR2532613 https://doi.org/10.1109/TAC.2009.2019797

DANI, V., HAYES, T. P. and KAKADE, S. M. (2008). Stochastic linear optimization under bandit feedback.

DELBRIDGE, I., BINDEL, D. and WILSON, A. G. (2020). Randomly projected additive Gaussian processes for regression. In *Proceedings of the* 37*th International Conference on Machine Learning* (H. D. III and A. Singh, eds.). *Proceedings of Machine Learning Research* **119** 2453–2463. PMLR.

DEN BOER, A. V. (2015). Dynamic pricing and learning: Historical origins, current research, and new directions. *Surv. Oper. Res. Manag. Sci.* **20** 1–18. MR3352577 https://doi.org/10.1016/j.sorms.2015.03.001

DUDIK, M., HSU, D., KALE, S., KARAMPATZIAKIS, N., LANGFORD, J., REYZIN, L. and ZHANG, T. (2011). Efficient optimal learning for contextual bandits. arXiv preprint. Available at arXiv:1106.2369.

GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707 https://doi.org/10.1214/09-AOS738

GUR, Y., MOMENI, A. and WAGER, S. (2021). Smoothness-adaptive contextual bandits. Available at SSRN.
GYÖRFI, L., KOHLER, M., KRZYZAK, A. and WALK, H. (2006). *A Distribution-Free Theory of Nonparametric Regression*. Springer, Berlin.
HADIJI, H. (2019). Polynomial cost of adaptation for x-armed bandits. In *Proceedings of the* 33*rd International Conference on Neural Information Processing Systems* 1029–1038.
HAO, B., LATTIMORE, T. and WANG, M. (2020). High-dimensional sparse linear bandits. arXiv preprint. Available at arXiv:2011.04020.
HAZAN, E. and KALE, S. (2011). Better algorithms for benign bandits. *J. Mach. Learn. Res.* **12** 1287–1311. MR2804601
HAZAN, E. and MEGIDDO, N. (2007). Online learning with prior knowledge. In *Learning Theory. Lecture Notes in Computer Science* **4539** 499–513. Springer, Berlin. MR2397608 https://doi.org/10.1007/978-3-540-72927-3_36
HU, Y., KALLUS, N. and MAO, X. (2020). Smooth contextual bandits: Bridging the parametric and non-differentiable regret regimes. In *Conference on Learning Theory* 2007–2010. PMLR.
KAKADE, S. M., KALAI, A. T. and LIGETT, K. (2009). Playing games with approximation algorithms. *SIAM J. Comput.* **39** 1088–1106. MR2538851 https://doi.org/10.1137/070701704
KANDASAMY, K., SCHNEIDER, J. and PÓCZOS, B. (2015). High dimensional Bayesian optimisation and bandits via additive models. In *International Conference on Machine Learning* 295–304. PMLR.
KLEINBERG, R. (2004). Nearly tight bounds for the continuum-armed bandit problem. *Adv. Neural Inf. Process. Syst.* **17** 697–704.
KLEINBERG, R., SLIVKINS, A. and UPFAL, E. (2008). Multi-armed bandits in metric spaces. In *STOC'*08 681–690. ACM, New York. MR2582691 https://doi.org/10.1145/1374376.1374475
KLEINBERG, R., SLIVKINS, A. and UPFAL, E. (2019). Bandits and experts in metric spaces. *J. ACM* **66** 30. MR3962354 https://doi.org/10.1145/3299873
LAI, T. L. and ROBBINS, H. (1985). Asymptotically efficient adaptive allocation rules. *Adv. in Appl. Math.* **6** 4–22. MR0776826 https://doi.org/10.1016/0196-8858(85)90002-8
LATTIMORE, T. and SZEPESVÁRI, C. (2020). *Bandit Algorithms*. Cambridge Univ. Press, Cambridge.
LINTON, O. and NIELSEN, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82** 93–100. MR1332841 https://doi.org/10.1093/biomet/82.1.93
LIU, Y., WANG, Y. and SINGH, A. (2021). Smooth bandit optimization: Generalization to Hölder space. In *International Conference on Artificial Intelligence and Statistics* 2206–2214. PMLR.
LOCATELLI, A. and CARPENTIER, A. (2018). Adaptivity to smoothness in x-armed bandits. In *Conference on Learning Theory* 1463–1492. PMLR.
LU, T., PÁL, D. and PÁL, M. (2009). Showing relevant ads via context multi-armed bandits Technical report.
MAILLARD, O.-A. and MUNOS, R. (2010). Online learning in adversarial Lipschitz environments. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 305–320. Springer, Berlin.
MCMAHAN, H. B. and BLUM, A. (2004). Online geometric optimization in the bandit setting against an adaptive adversary. In *Learning Theory. Lecture Notes in Computer Science* **3120** 109–123. Springer, Berlin. MR2177904 https://doi.org/10.1007/978-3-540-27819-1_8
MINSKER, S. (2013). Estimation of extreme values and associated level sets of a regression function via selective sampling. In *Conference on Learning Theory* 105–121. PMLR.
PANDEY, S., AGARWAL, D., CHAKRABARTI, D. and JOSIFOVSKI, V. (2007). Bandits for taxonomies: A model-based approach. In *Proceedings of the* 2007 *SIAM International Conference on Data Mining* 216–227. SIAM, Philadelphia.
PEARSON, L. M. and BERRY, D. A. (1981). Optimal designs for two-stage clinical trials with dichotomous responses Technical report, Univ. Minnesota.
PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28** 298–335. MR1762913 https://doi.org/10.1214/aos/1016120374
RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE Trans. Inf. Theory* **57** 6976–6994. MR2882274 https://doi.org/10.1109/TIT.2011.2165799
RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2012). Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *J. Mach. Learn. Res.* **13** 389–427. MR2913704
RIGOLLET, P. and ZEEVI, A. (2010). Nonparametric bandits with covariates. arXiv preprint. Available at arXiv:1003.1630.
ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 527–535. MR0050246 https://doi.org/10.1090/S0002-9904-1952-09620-8
ROLLAND, P., SCARLETT, J., BOGUNOVIC, I. and CEVHER, V. (2018). High-dimensional Bayesian optimization via additive models with overlapping groups. In *Proceedings of the Twenty-First International Conference on*

Artificial Intelligence and Statistics (A. Storkey and F. Perez-Cruz, eds.). *Proceedings of Machine Learning Research* **84** 298–307. PMLR.

RUSMEVICHIENTONG, P. and TSITSIKLIS, J. N. (2010). Linearly parameterized bandits. *Math. Oper. Res.* **35** 395–411. MR2674726 https://doi.org/10.1287/moor.1100.0446

SINGH, S. (2021). Continuum-armed bandits: A function space perspective. In *International Conference on Artificial Intelligence and Statistics* 2620–2628. PMLR.

SLIVKINS, A. (2011). Contextual bandits with similarity information. In *Proceedings of the* 24*th Annual Conference on Learning Theory*. *JMLR Workshop and Conference Proceedings*. 679–702.

SLIVKINS, A. and VAUGHAN, J. W. (2014). Online decision making in crowdsourcing markets: Theoretical challenges. *ACM SIGecom Exch.* **12** 4–23.

STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705. MR0790566 https://doi.org/10.1214/aos/1176349548

WANG, Y., BALAKRISHNAN, S. and SINGH, A. (2019). Optimization of smooth functions with noisy observations: Local minimax rates. *IEEE Trans. Inf. Theory* **65** 7350–7366. MR4030889 https://doi.org/10.1109/TIT.2019.2921985

YUAN, M. and ZHOU, D.-X. (2016). Minimax optimal rates of estimation in high dimensional additive models. *Ann. Statist.* **44** 2564–2593. MR3576554 https://doi.org/10.1214/15-AOS1422

ZHAO, P. and LAI, L. (2021). Optimal stochastic nonconvex optimization with bandit feedback. arXiv preprint. Available at arXiv:2103.16082.