Optimal Sample Complexity Bounds for Non-convex Optimization under Kurdyka-Lojasiewicz Condition

Qian Yu
Princeton University

Yining Wang
University of Texas at Dallas

Baihe Huang
University of California, Berkeley

Qi Lei
New York University

Jason D. Lee
Princeton University

Abstract

Optimization of smooth reward functions under bandit feedback is a long-standing problem in online learning. This paper approaches this problem by studying the convergence under smoothness and Kurdyka-Lojasiewicz conditions. We design a search-based algorithm that achieves an improved rate compared to the standard gradient-based method. In conjunction with a matching lower bound, this algorithm shows optimality in the dependence on precision for the low-dimension regime.

1 Introduction

Zeroth-order optimization with bandit feedback pertains to a class of optimization problems in which one aims at searching over a set of candidate alternatives to minimize (or maximize) an unknown objective function, while only accessible to (noisy) values of the objective function at several one-point queries. This setting corresponds to a wide range of real-world applications, such as in healthcare (Yu et al., 2020), robotics (Kober et al., 2014), and education (Kizilcec et al., 2020), where the analytic forms of neither the objective function nor the directive at given points are available. Considering the cost of conducting experiments or trials in such applications, it is often desirable to amortize the sample cost (number of queries) by exploiting the geometry of the objective function.

In the past decades, this problem has attracted extensive research in both optimization community (Flaxman et al., 2005; Nesterov and Spokoiny, 2017; Lan, 2019) and machine learning community (Abernethy et al., 2008; Agarwal et al., 2011a; 2013; Hazan and Levy, 2014b; Lattimore and György, 2021). For various special cases, such as the noiseless feedback setting (Nesterov and Spokoiny, 2017), multi-arm bandits (Auer, 2002), linear bandits (Dani et al., 2008; Abbasi-Yadkori et al., 2011), and bandit optimization for quadratic functions (Shamir, 2013), the complexity of this problem has been well understood and matching upper and lower bounds have been established. However, little is known for zeroth-order optimization with bandit feedback for *non-convex* objective functions. In this paper, we consider the class of smooth functions that satisfy the Kurdyka-Lojasiewicz (KL) condition (Kurdyka, 1998). This class thus includes strongly convex functions and many other non-convex functions. Formally, we assume that the Lojasiewicz inequality holds for all points where the function values are sufficiently close to the global minimum (see condition A1). This property naturally extends the well-studied Polyak-Lojasiewicz condition, which is satisfied in cases of interest, such as support vector machines (Karimi et al., 2016) and tabular reinforcement learning (Agarwal et al., 2021). For this class of objective functions, we study the following fundamental question: *Can we design algorithms for zeroth-order optimization with bandit feedback whose sample cost depends favorably on the the exponent of KL conditions?*

Motivated by this fundamental question, we develop a provably sample efficient algorithm and establish information-theoretic lower bounds. Our main contributions are summarized as follows: We develop the first set of sample efficient algorithms where the convergence rates depend favorably on α. In the low dimension regime ($d \leq 5$), our algorithm achieves ϵ simple regret with the sample complexity of $\widetilde{O}\left(\epsilon^{-2-\frac{d(d+3)(\alpha-\frac{1}{2})}{d+5}}\right)$. We further prove a matching lower bound under the assumption of Gaussian noise, showing that $\widetilde{\Theta}\left(\epsilon^{-2-\frac{d(d+3)(\alpha-\frac{1}{2})}{d+5}}\right)$ is indeed the optimal rate.

Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, Valencia, Spain. PMLR: Volume 206.

Finding sample-efficient algorithms in the low dimensional regime has been of interest for various reasons. On one hand, both the optimal sample complexity and the required achievability algorithm could differ significantly (Chewi et al., 2022). On the other hand, algorithmic techniques developed for low dimensional regimes could potentially serve as building blocks to improve results in high dimension (Bubeck and Mikulincer, 2020). For the framework considered in this paper, our result implies that in high-dimensional regimes ($d > 5$) the minimax exponent of ϵ become independent of d modulo a constant factor, which is bounded between -4α and $-(6\alpha - 1)$.

Related works on linear and convex bandit. It is known from classical optimization literature (Karimi et al., 2016) that gradient descent-type algorithms are able to find the global optimum of smooth and strongly convex functions in the noiseless feedback setting, in linear convergence rates. However, optimization under zeroth-order bandit feedback faces more challenges due to the restriction of available information of the objective function and the existence of noise in the signal. For linear bandit problems, the attainable simple regret is well-known as exactly $O(\sqrt{d/T})$ (See reference therein [(Auer, 2002; Dani et al., 2008; Rusmevichientong and Tsitsiklis, 2010; Abbasi-Yadkori et al., 2011; Chu et al., 2011; Audibert et al., 2009). Here d is the dimension and T is the number of queries. For bandit optimization of smooth and strongly-convex functions, which corresponds to the KL condition with exponent $1/2$, the optimal simple regret is known to be $O(\sqrt{d^2/T})$ (Shamir, 2013). For more general convex settings, the dependence on d is less clear but it has been established the optimal dependence on T is $O(1/\sqrt{T})$ (Shamir, 2013; Lattimore and Szepesvári, 2020; Duchi et al., 2015; Bubeck et al., 2017; Hazan and Levy, 2014a; Agarwal et al., 2011b; Wang et al., 2018).

Related works on non-convex bandit optimization. Finding optimal bandit under non-convex reward function is in general intractable. Prior work on non-convex stochastic (zeroth-order) optimization mainly focuses on finding ϵ-stationary point instead of ϵ-optimal reward (see e.g., (Larson et al., 2019) for a review). Exceptions occur in very specific settings like homogeneous quadratic (Lattimore and Hao, 2021) or high-order polynomials (Huang et al., 2021), sum of univariate functions (Zhang et al., 2015) or non-parametric settings (Zhao and Lai, 2021; Moulines and Bach, 2011).

Our work resembles the setting of zeroth-order stochastic optimization with the form: $\min_x : \mathbb{E}_\zeta F(x, \zeta)$. A flurry of work ((Fang et al., 2018; Zhou et al., 2018; Wang et al., 2019b; Cutkosky and Orabona, 2019; Balasubramanian and Ghadimi, 2018)) studies zeroth or first-order stochastic optimization and their variance reduced variants (Liu et al., 2018b; Ji et al., 2019; Liu et al., 2018a), with a best known rate to be $O(1/T^{1/3})$. However, these problems are strictly easier than our considered setting, with "multi-point" queries of the function values for the same random seed, i.e., at $\{F(x_1, \zeta), F(x_2, \zeta), \cdots, F(x_n, \zeta)\}$ with the same ζ. A more detailed discussion and separation results on its distinction to our setting can be found in (Arjevani et al., 2019).

2 Optimal Bandit Optimization under Kurdyka-Lojasiewicz Conditions

We give a rigorous formulation of the stochastic zeroth-order optimization problem studied in this paper. Given a fixed dimension parameter d, let $f : \mathbb{R}^d \to \mathbb{R}$ be an unknown objective function defined on the entire Euclidean space $\mathbb{R}^d$, satisfying certain regularity conditions to be imposed later. At a time period t, an optimization algorithm $\mathcal{A}$ produces a query point $\boldsymbol{x}_t \in \mathbb{R}^d$ and receives feedback

$$y_t = f(\boldsymbol{x}_t) + w_t,$$

where $\{w_t\}_{t=1}^T$ are independently distributed random variables satisfying $\mathbb{E}[w_t|\boldsymbol{x}_t] = 0$ and the distributions of w_t are sub-Gaussian with parameter one[1]. The query points $\{\boldsymbol{x}_t\}_{t=1}^T$ produced by an optimization algorithm $\mathcal{A}$ can be *adaptively* chosen. More specifically, $\mathcal{A}$ can be parameterized as $\mathcal{A} = (\phi_1, \cdots, \phi_T)$, where $\phi_t(\cdot|\boldsymbol{x}_1, y_1, \cdots, \boldsymbol{x}_{t-1}, y_{t-1})$ is a conditional distribution measurable with respect to historical data $\{\boldsymbol{x}_\tau, y_\tau\}_{\tau<t}$ and $\boldsymbol{x}_t \sim \phi_t(\cdot|\boldsymbol{x}_1, y_1, \cdots, \boldsymbol{x}_{t-1}, y_{t-1})$.

We assume that the unknown objective function f is differentiable. Furthermore, we impose the following technical conditions.

(A1) (KL-inequality). There exist constants $\alpha \in (0.5, 1)$ and $C_1 < \infty$ such that for all $\boldsymbol{x} \in \mathbb{R}^d$ with $f(\boldsymbol{x}) \le f(\mathbf{0})$, it holds that $\|\nabla f(\boldsymbol{x})\|_2 \ge C_1|f(\boldsymbol{x}) - f^*|^\alpha$, where f^* is the infimum of f;

(A2) (Smoothness). There exists a constant $C_2 < \infty$ such that for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d$ with $f(\boldsymbol{x}), f(\boldsymbol{x}') \le f(\mathbf{0})$, it holds that $f(\boldsymbol{x}') - f(\boldsymbol{x}) \le (\boldsymbol{x}' - \boldsymbol{x}) \cdot \nabla f(\boldsymbol{x}) + C_2\|\boldsymbol{x}' - \boldsymbol{x}\|_2^2$

The KL-inequality (A1) is an important technical condition that generalizes classical strongly convexity of the objective functions.[2] The (strong) smoothness condition (A2) prevents the objective function (more precisely, the gradients of

[1] The sub-Gaussianity assumption can be relaxed to allow for general noise distributions with bounded variances, and the same sample complexity can be achieved by applying the truncation method in (Yu et al., 2023).

[2] The KL-inequality in Condition A1 implies that all local extremum points with function values no greater than $f(\mathbf{0})$ are global minimum. This may not hold true if the validity of Lojasiewicz inequality is restricted to smaller subsets.

the objective function) from changing too rapidly for neighboring sample points. It is a conventional condition imposed in many first-order and zeroth-order stochastic optimization problems (Nemirovskij and Yudin, 1983; Agarwal et al., 2010; Wang et al., 2019a).

Throughout the rest of the paper we use $\mathcal{F}(\alpha, C_1, C_2)$ to denote all differentiable functions $f : \mathbb{R}^d \to \mathbb{R}$ satisfying assumptions (A1) and (A2) with parameters α, C_1 and C_2. For simplicity, we define $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^d \mid f(\boldsymbol{x}) \leq f(\mathbf{0})\}$. We adopt the $\widetilde{O}$ notation, i.e., $g = \widetilde{O}(f)$ implies that $g = O(f \log^c f)$ for some universal constant c, where the multiplicative factor could depend α, and polynomially on the logarithms of C_1, C_2.

We are interested in the minimax *simple* regret over function classes $\mathcal{F}$, where the *simple* regret of an algorithm $\mathcal{A}$ is simply defined as $f(\boldsymbol{x}_T) - f^*$, with $x_T \in \mathbb{R}^d$ being the query point selected at the last time period and f^* the infimum of f. More specifically, for any $\alpha \in (0.5, 1)$ and $C_1, C_2 < \infty$, the minimax regret function $\mathfrak{R}$ is defined as

$$\mathfrak{R}(T; \alpha, C_1, C_2) \triangleq \inf_{\mathcal{A}} \sup_{f \in \mathcal{F}(\alpha, C_1, C_2)} \mathbb{E}\left[f(\boldsymbol{x}_T) - f^*\right].$$

We formulate our results in terms of the smallest T such that $\mathfrak{R}(T; \alpha, C_1, C_2) \leq \varepsilon$ for sufficiently small accuracy parameters $\varepsilon > 0$. Our main results can be summarized into the following theorems.

Theorem 1. *For any $d \leq 5$, there is an algorithm that achieves an expected simple regret of $\mathbb{E}\left[f(\boldsymbol{x}_T) - f^*\right] \leq \epsilon$ for all $f \in \mathcal{F}(\alpha, C_1, C_2)$, with a sample complexity of $T = \widetilde{O}\left(\left(\frac{C_2}{C_1^2}\right)^{\frac{d(d+3)}{2(d+5)}} \epsilon^{-2-\frac{d(d+3)(\alpha-\frac{1}{2})}{d+5}}\right)$.*

Remark 2.1. *As a comparison, stochastic gradient based methods can achieve a sample complexitiy with an ϵ-dependency of $\widetilde{O}\left(\epsilon^{-(6\alpha-1)}\right)$ (see Appendix A). Theorem 1 strictly improves this sample complexity for all $d \leq 5$. Furthermore, a direct application of the algorithm presented in Section 3.3 also provides a strict improvement in $d = 6$, achieving an ϵ-dependency of $\widetilde{O}\left(\epsilon^{-\left(5\alpha-\frac{1}{2}\right)}\right)$. In Remark 3.2, we provide a detailed discussion on how the low-dimension requirement emerges in the analysis.*

The above theorem is proved by showing the following high-probability bound.

Theorem 2. *For any $d \leq 5$ and sufficiently small δ, there is an algorithm that achieves an error of $f(\boldsymbol{x}_T) - f^* \leq \epsilon$ w.p. $1 - \delta$ for all $f \in \mathcal{F}(\alpha, C_1, C_2)$, with a sample complexity of $T = \widetilde{O}\left(\left(\frac{C_2}{C_1^2}\right)^{\frac{d(d+3)}{2(d+5)}} \epsilon^{-2-\frac{d(d+3)(\alpha-\frac{1}{2})}{d+5}} \log(1/\delta)\right)$.*

We present the related algorithms in Section 3.1, 3.2 and 3.3. We also show a matching lower bound, stated as follows and proved in Section 4.

Theorem 3. *If $d \leq 5$ and the distributions of w_t are i.i.d. standard Gaussian, then any algorithm that achieves an expected error of ϵ for all $f \in \mathcal{F}(\alpha, C_1, C_2)$ requires a sample complexity of $T = \Omega\left(\left(\frac{C_2}{C_1^2}\right)^{\frac{d(d+3)}{2(d+5)}} \epsilon^{-2-\frac{d(d+3)(\alpha-\frac{1}{2})}{d+5}}\right)$.*

Remark 2.2. *The optimal sample complexity for any fixed ϵ is non-decreasing in d (see Appendix B). Hence, the lower bound stated in Theorem 3 for $d = 5$ also applies to any $d > 5$ cases, implying a lower bound of $\Omega\left(\epsilon^{-4\alpha}\right)$.*

Remark 2.3. *For completeness, we have presented a full characterization of sample complexities in our main theorems to include dependencies on all parameters. However, one should note that to obtain such a full characterization, it suffices to obtain the near optimal ϵ dependency and apply dimensional analysis. Specifically, if we know the optimal sample complexity is $\widetilde{\Theta}(\epsilon^{-\theta})$ for some $\theta > 0$ for any fixed α, d, C_1 and C_2, then the full characterization has to be in the form of $\widetilde{\Theta}\left(\left(\frac{C_2}{C_1^2}\right)^{\frac{\theta-2}{2\alpha-1}} \epsilon^{-\theta}\right)$ asymptotically.*

3 Proofs for Bandit Optimization under Kurdyka-Lojasiewicz conditions

3.1 A Gentle Example: The 1D Algorithm

The proposed algorithm operates alternatively in two phases, a global search phase and a local search phase. We presented an illustration using the $d = 1$ case, where the optimal complexity can be achieved using a simpler routine.

Proposition 1. *For $d = 1$ and sufficiently small δ, there is an algorithm that achieves an error of $f(x_T) - f^* \leq \epsilon$ w.p. $1 - \delta$ for all $f \in \mathcal{F}(\alpha, C_1, C_2)$ with a sample complexity of $\widetilde{O}\left(\left(\frac{C_2}{C_1^2}\right)^{\frac{1}{3}} \epsilon^{-\frac{2\alpha+5}{3}} \log(1/\delta)\right)$.*

Proof. As described in Algorithm 1, the proposed routine maintains a variable x_{c} over T iterations. We ensure that the gap between $f(x_{\mathrm{c}})$ and f^* is upper bounded by a quantity g with high probability, so that if g is reduced by a constant factor each iteration, then the stated learning error ϵ can be achieved within logarithmic number of iterations.

Each iteration starts with a global search phase, where the algorithm maintains an interval $[x_{\mathrm{c}} - L, x_{\mathrm{c}} + L]$, of which the length is to be reduced by a factor of $\frac{1}{2}$ through each search step. The main idea of the algorithm is to ensure that the minimum of f within the interval approximates f^* with high probability, so that once L is small enough, one can switch to a local search phase, and the approximated minim can be found using a grid search. The transition threshold L_{th} is selected to minimize the overall sampling cost (see the analysis at the end of the proof).

The design and the choice of parameter values in the algorithm are based on the following building blocks. First, a

Algorithm 1 1D Search

Initialize $x_{\text{c}} = 0$, $g = \left(\frac{4C_2}{C_1^2}\right)^{\frac{1}{2\alpha-1}}$
Let $T = \lceil 2\log_2(g/\epsilon) \rceil$.
for $t \leftarrow 1$ to T **do**
  Global Search Phase:
  Let $C_0 = \frac{1}{2}C_1(1-\alpha)(g/2)^{\alpha}$, $L_{\text{th}} = \left(\frac{1}{C_0^2}\sqrt{\frac{g^5}{C_2}}\right)^{\frac{1}{3}}$
  Let $\delta_t = \frac{\delta}{2^{T-t}}$. Initialize $L = \frac{g}{C_0}$.
  while $L > L_{\text{th}}$ **do**
    Let $\Delta_C = C_0\left(L_{\text{th}}/2L\right)^{\frac{2}{3}}$, $\epsilon_L = \frac{1}{8}\Delta_C \cdot L$
    Let $\delta_L = \delta_t \cdot \frac{L_{\text{th}}}{24L}$, $\mathcal{S} = \{x_{\text{c}} - \frac{L}{2}, x_{\text{c}}, x_{\text{c}} + \frac{L}{2}\}$
    Let $x_{\text{c}} = \text{argmin}_{y \in \mathcal{S}}(C_0 + \Delta_C) \cdot |y - x_{\text{c}}|$
      $+\text{Sample}(y, \epsilon_L, \delta_L)$
    Let $L = \frac{1}{2}L$.
  end while
  Local Search Phase:
  Let $N = 2\left\lceil L/\sqrt{\frac{g}{C_2}}\right\rceil$, $\epsilon_{\text{s}} = \frac{g}{16}$, $\delta_{\text{s}} = \delta_t \cdot \frac{1}{2(N+1)}$
  Let $\mathcal{D} = \{x_{\text{c}} - L, x_{\text{c}} - L + \frac{2L}{N}, ..., x_{\text{c}} + L\}$
  Let $x_{\text{c}} = \text{argmin}_{y \in \mathcal{D}}$ $\text{Sample}(y, \epsilon_{\text{s}}, \delta_{\text{s}})$
    $-2\epsilon_{\text{s}} \cdot \mathbb{1}(y = x_{\text{c}})$
  Let $g = \frac{g}{2}$.
end for
  return x_c

procedure SAMPLE($\boldsymbol{x}, \epsilon_{\text{s}}, \delta_{\text{s}}$)
  return the average of $\lceil 2\epsilon_{\text{s}}^{-2} \ln \frac{1}{\delta_{\text{s}}} \rceil$ samples of f at $\boldsymbol{x}_t = \boldsymbol{x}$.
end procedure

subroutine *Sample* is to be used throughout this work, which repetitively samples f at the same point to provide an accurate estimation. Formally, we have the following guarantee, which directly follows from the Hoeffding inequality of sub-Gaussian random variables.

Proposition 2. *The Sample function returns a value that is within* $(f(\boldsymbol{x}) - \epsilon_{\text{s}}, f(\boldsymbol{x}) + \epsilon_{\text{s}})$ *w.p.* $1 - \delta_{\text{s}}$.

We shall always consider the high-probability regime where all sampled results fall into the intervals $(f(\boldsymbol{x}) - \epsilon_{\text{s}}, f(\boldsymbol{x}) + \epsilon_{\text{s}})$. The overall error probability is controlled by the union bound.

Besides, we use the following property to set the initialization of g, so that we have $f(x_{\text{c}}) - f^* \leq g$ at the beginning of the first iteration. The proof of Proposition 3 can be found in appendix C.

Proposition 3 (Boundedness)**.** *For any differentiable function f that satisfies smoothness and KL inequality for all $\boldsymbol{x} \in \mathcal{X}$, i.e., $f(\boldsymbol{x}) \leq f(\mathbf{0})$, we have*

$$f(\mathbf{0}) - f^* \leq \left(\frac{4C_2}{C_1^2}\right)^{\frac{1}{2\alpha-1}}.$$

Moreover, for any $\boldsymbol{x} \in \mathcal{X}$ there is a global minimum of f within an L_2-distance of $L_{\boldsymbol{x}} \triangleq \frac{(f(\boldsymbol{x})-f^*)^{1-\alpha}}{C_1(1-\alpha)}$. For any $\boldsymbol{x} \in \mathcal{X}$ and $L \leq L_{\boldsymbol{x}}$, let $\mathcal{B}$ be the L_2-ball centered at $\boldsymbol{x}$ with a radius of L. If $L_{\boldsymbol{y}} > L_{\boldsymbol{x}} - L$ for all $\boldsymbol{y} \in \partial B$, then there is a global minimum of f within $\mathcal{B}$.

Given these facts, our algorithm is designed to ensure that for any fixed iteration, if $f(x_{\text{c}}) - f^* \leq g$ holds at the beginning, the same inequality holds with high probability at the end. To avoid confusion, we denote the value of g at the beginning of the iteration by g_t, so at the end of of the iteration we have $f(x_{\text{c}}) - f^* \leq g_t/2$.

The gist of the global search phase is to ensure that throughout the entire process, either we have $f(x_{\text{c}}) - f^* \leq g_t/2$, or a global minimum of f exists in $(x_{\text{c}} - L, x_{\text{c}} + L)$. By the second and third statements of Proposition 3, this condition can be implied if $\boldsymbol{x}_{\text{c}} \in \mathcal{X}$ and both $f(x_{\text{c}} - L)$ and $f(x_{\text{c}} + L)$ are greater than $f(x_{\text{c}}) - 2C_0 \cdot L$. In the ideal case where the function value of f can be exactly obtained, one can apply a binary search to maintain the above property while decreasing the interval length L. However, due to the stochastic nature of the optimization problem, we need to consider an alternative list of gradually relaxed constraints to reduce the sample complexity. Hence, we instead ensure the following stronger condition to hold at the end of each search step (i.e., each inner iteration).

$$f(x_{\text{c}} \pm L) \geq f(x_{\text{c}}) - (C_0 + \Delta_C)\,L. \tag{1}$$

As shown in the detailed analysis in Appendix D, the presented global search guarantees the above condition, with the number of samples required for each search step being inversely proportional to the square of the increment of Δ_C times L. Therefore, we design the relaxation parameter Δ_C to be exponentially dependent on the number of inner iterations to compensate the exponential decay of L, to minimize the overall sample complexity.

Note that in both global search and local search phases, x_{c} is only updated when $f(x_{\text{c}})$ is known to be decreased with high probability. Hence, we have $x_{\text{c}} \in \mathcal{X}$ throughout the entire process, which ensures the correctly of the global search algorithm.

Now we consider the local search phase. Because $f(x_{\text{c}})$ is non-increasing with high probability. It remains to focus on the case where $x_{\text{c}} > g_t/2$. According to the earlier conclusion, a global minimum exists within $(x_{\text{c}} - L, x_{\text{c}} + L)$. From the smoothness condition, any point within a distance of $O\left(\sqrt{\frac{g_t}{C_2}}\right)$ to this global minimum has a function value no greater than $f^* + O(g_t)$. Based on this observation, we perform the grid search by letting adjacent sample points having the same distance, i.e., setting the number of sample points N to be $\Theta\left(\left\lceil L_{\text{th}}/\sqrt{\frac{g_t}{C_2}}\right\rceil\right)$. Formally, let $\mathcal{D}$ denote the set of all $N+1$ sample points in the local search and

f^*_{disc} denote the minimum of f within this discrete set. The smoothness property at x^* implies the following quantization error bound.

$$f^*_{\text{disc}} - f^* \leq C_2 \left(\frac{L}{N}\right)^2 \leq \frac{g_t}{4}.$$

Therefore, by Proposition 2, x_{c} is updated to a different point with high probability due to $f(x_{\text{c}}) - f^*_{\text{disc}} > 4\epsilon_{\text{s}}$, and the updated function value is upper bounded by $f^*_{\text{disc}} + 2\epsilon_{\text{s}} < f^* + \frac{g_t}{2}$.

As mentioned earlier, the overall error probability can be controlled by the union bound. For each iteration t, any global search step calls the sample function three times, incurring an error probability of at most $3\delta_L = \delta_t \cdot \frac{L_{\text{th}}}{8L}$. Recall that L decays exponentially and is lower bounded by $\frac{1}{2}L_{\text{th}}$. The total error probability due to global search within this iteration is at most $\frac{1}{2}\delta_t$. Then for the local search, the sample function is called $N+1$ times, and the overall error probability is bounded by $(N+1)\delta_{\text{s}} = \frac{1}{2}\delta_t$ as well. Taking all iterations into account, the algorithm returns an x_{c} with $f(x_{\text{c}}) - f^* \leq \epsilon$ with an error probability of at most $\sum_{t=1}^{T} \delta_t < \delta$.

The sample complexity bound can be analyzed similarly. Consider any iteration t. The sampling cost of any global search step is at most $O(\epsilon_L^{-2} \log(1/\delta_L))$, and the total cost within the iteration is at most $O\left(C_0^{-2} L_{\text{th}}^{-2} \log\left(\frac{1}{\delta_t}\right)\right) = O\left(\left(\frac{C_2}{C_1^2}\right)^{\frac{1}{3}} g_t^{-\frac{2\alpha+5}{3}} \log\left(\frac{1}{\delta_t}\right)\right)$ for sufficiently small δ. Then for the local search, the same Sample function is called $O(N)$ times, each requiring $O(\epsilon_{\text{s}}^{-2} \log(1/\delta_{\text{s}}))$ samples. Hence, the total sampling cost for local search is upper bounded by $O\left(N g_t^{-2} \log\left(\frac{1}{g_t \delta_t}\right)\right) = \widetilde{O}\left(\left(\frac{C_2}{C_1^2}\right)^{\frac{1}{3}} g_t^{-\frac{2\alpha+5}{3}} \log\left(\frac{1}{\delta_t}\right)\right)$. Using the fact that g_t exponentially depends on t, the overall sample complexity is bounded by $\widetilde{O}\left(\left(\frac{C_2}{C_1^2}\right)^{\frac{1}{3}} \epsilon^{-\frac{2\alpha+5}{3}} \log(1/\delta)\right)$. □

3.2 Dominant Point and Iterative Reduction

To present the general search algorithm, we first state some useful concepts and procedures. For brevity, we shall reuse the Sample function defined in Algorithm 1. We will also use Proposition 2 and Proposition 3, which hold for any d.

Definition 1 (Dominant Point). *Given any set $\mathcal{B}$, function f, and fixed $C > 0$, we say $\boldsymbol{x} \in \mathcal{B}$ is a* dominant point *of $\mathcal{B}$ if for any $\boldsymbol{y} \in \partial\mathcal{B}$ we have*

$$f(\boldsymbol{x}) < f(\boldsymbol{y}) + C||\boldsymbol{y} - \boldsymbol{x}||_2.$$

We prove the following property (see Appendix E), which ensures that any dominant point $\boldsymbol{x}$ on a set $\mathcal{B}$ imposes guarantees on either the location of a global minimum of f, or the function value of $\boldsymbol{x}$.

Proposition 4. *For any given $f \in \mathcal{F}$ and $C > 0$, if $\boldsymbol{x} \in \mathcal{X}$ is a dominant point of a closed set $\mathcal{B}$, then either a global minimum of f can be found within $\mathcal{B}$, or $f(\boldsymbol{x}) - f^* \leq \left(\frac{C}{C_1(1-\alpha)}\right)^{\frac{1}{\alpha}}$.*

A key building block of the proposed algorithm is a function that runs iteratively and each time returns a dominant point from a smaller set (see Algorithm 2). More specifically, given any hypercuboid $\mathcal{B}$ and a dominant point $\boldsymbol{x}$, they returns a dominant point for a smaller hypercuboid, where the length of the longest edge is reduced by at least a factor of $4/5$.

Algorithm 2

procedure REDUCE($\mathcal{B} = \prod_{j=1}^{d}[a_j, b_j]$, $\boldsymbol{x} = (x_1, ..., x_d)$, C_0, Δ, δ_{r})
  Let $k = \text{argmax}_j (b_j - a_j)$, $\epsilon_{\text{r}} = \Delta(b_k - a_k)/20$
  Let $(p, q) = \text{Balanced Partition}(a_k, b_k, x_k)$
    $\mathcal{S}_p = \{\boldsymbol{y} = (y_1, ..., y_d) \in \mathcal{B} \mid y_k = p\}$
    $\mathcal{S}_q = \{\boldsymbol{y} = (y_1, ..., y_d) \in \mathcal{B} \mid y_k = q\}$
  Let $\mathcal{S}_{\text{net},p}$, $\mathcal{S}_{\text{net},q}$ be $\sqrt{\epsilon_{\text{r}}/C_2}$-nets of $\mathcal{S}_p$ and $\mathcal{S}_q$, respectively, and $\mathcal{S} = \mathcal{S}_{\text{net},p} \cup \mathcal{S}_{\text{net},q} \cup \{\boldsymbol{x}\}$
  Let $\boldsymbol{y}^* = \text{argmin}_{\boldsymbol{y} \in \mathcal{S}} (C_0 + \Delta)||\boldsymbol{y} - \boldsymbol{x}||_2 + \text{Sample}(\boldsymbol{y}, \epsilon_{\text{r}}, \frac{\delta_{\text{r}}}{|\mathcal{S}|})$
  if $\boldsymbol{y}^* \notin \mathcal{S}_p$ **then**
    $a_k = p$
  end if
  if $\boldsymbol{y}^* \notin \mathcal{S}_q$ **then**
    $b_k = q$
  end if
  return $(\mathcal{B}^* = \prod_{j=1}^{d}[a_j, b_j], \boldsymbol{y}^*)$.
end procedure

procedure BALANCED PARTITION(a, b, x)
  return $(p, q) =$
$$\begin{cases} (\frac{2x+b}{3}, \frac{x+2b}{3}) & \text{if } x \in [a, \frac{3a+2b}{5}) \\ (\frac{a+x}{2}, \frac{x+b}{2}) & \text{if } x \in [\frac{3a+2b}{5}, \frac{2a+3b}{5}) \\ (\frac{2a+x}{3}, \frac{a+2x}{3}) & \text{if } x \in [\frac{2a+3b}{5}, b] \end{cases}$$
end procedure

Formally, we have the following guarantees on the reduce function, which is proved in Appendix F.

Proposition 5. *For any input $\boldsymbol{x} \in \mathcal{X}$ being a dominant point of the input hypercuboid $\mathcal{B}$ with parameter $C = C_0$, if f satisfies the smoothness condition on $\mathcal{B} \cap \mathcal{X}$ and $C_0 + 2\Delta \leq \frac{2}{5}C_2(b_k - a_k)$, then w.p. $1 - \delta_{\text{r}}$, the Reduce function returns a pair such that $\boldsymbol{y}^*$ is a dominant point of $\mathcal{B}^*$ with parameter $C = C_0 + 2\Delta$. Moreover, if $\frac{b_j - a_j}{b_k - a_k} = \Theta(1)$ for each $j \in [d]$ and $\Delta = O(b_k - a_k) = O(1)$, then for any sufficiently small δ, the overall sample complexity is*

$O\left(\frac{C_2^{\frac{d-1}{2}}(b_k-a_k)^{\frac{d-5}{2}}}{\Delta^{\frac{d+3}{2}}}\log(\frac{C_2(b_k-a_k)}{\Delta\delta})\right)$ *for any fixed d.*

Remark 3.1. *In the Reduce function, the choice of using ϵ-nets to construct the set of sample points is merely for brevity. In practice, one can exploit the fact that the sets $\mathcal{S}_p$ and $\mathcal{S}_q$ to be sampled are always hypercuboid. Thus, any ϵ-net used in the algorithm can be replaced by a uniform grid with an interval length of $\epsilon/\sqrt{d}$ while still ensuring the correctness of Proposition 5. This replacement also ensures that the required computational complexity is at most linear with respect to the sample complexity.*

3.3 Proof of Achievability Theorems

Similar to the 1D case, an expected simple regret of $O(\epsilon)$ is guaranteed once it can be achieved with high probability, i.e., with $\delta = O(\epsilon)$.[3] Therefore, we focus on the algorithm needed for Theorem 2, which is presented as Algorithm 3.

Algorithm 3 General Search Algorithm

```
procedure KL SEARCH(ε, δ)
    Initialize x = 0, g = (4C_2/C_1^2)^{1/(2α−1)}
    Let T_1 = ⌈2 log_{3/2}(g/ε)⌉
    for t ← 1 to T_1 do
        Initialize L = 3g^{1−α}/(C_1(1−α)), B = [−L, L]^d
        Initialize C = C_0 = g/L
        Let L_th = g^{(d+4)/(d+5)} / (C_2^{1/(d+5)} C_0^{(d+3)/(d+5)})
        Let T_2 = d ⌈log_{5/4}(L/L_th)⌉^+
        Let Δ = C_0/(2T_2), δ_r = δ/(2T_1T_2).
        while L ≥ L_th do
            (B, x) =Reduce(B, x, C, Δ, δ_r).
            Let L be the longest edge length in B
            Let C = C + 2Δ
        end while
        x =Uniform search(B, x, g/3, δ/(2T_1)), g = (2/3)^α g
    end for
    return x
end procedure

procedure UNIFORM SEARCH(B, x, ε_u, δ_u)
    Let Y be an √(ε_u/C_2)-net of B
    return y* = argmin_{y∈Y∪{x}}
                Sample(y, ε_u/4, δ_u/(|Y|+1)) − (1/2)ε_u · 𝟙(y = x)
end procedure
```

Within each iteration of the outer loop, the algorithm first executes a global search using the Reduce function to locate a dominant point within a hypercuboid. Formally, recall that the Reduce function reduces the longest edge length of $\mathcal{B}$ by a factor of at least $\frac{4}{5}$. The inner loop halts within T_2 iterations. By Proposition 5, the end result of the inner loop is a dominant point of the returned set $\mathcal{B}$ with parameter $C = 2C_0$ w.p. $1 - \frac{\delta}{2T_1}$, if $\boldsymbol{x}$ is initially a dominant point of $[-L, L]^d$ with $C = C_0$. From proposition 3, this condition is satisfied if $f(\boldsymbol{x}) - f^* \leq g$.

A local search follows the inner loop to return a point that is close to the global minimum.

Proposition 6. *Given any inputs $\mathcal{B}$, $\boldsymbol{x}$, the function Uniform Search returns a point y^* satisfying the following conditions w.p. $1-\delta_{\rm u}$. If $\mathcal{B}$ contains a global minimum of f, we have $f(\boldsymbol{y}^*) - f^* \leq 2\epsilon_{\rm u}$; otherwise, we have $f(\boldsymbol{y}^*) - f(\boldsymbol{x}) \leq 0$.*

Combining Proposition 4 and Proposition 6, the point $\boldsymbol{x}$ returned from the Uniform Search satisfies $f(\boldsymbol{x}) - f^* \leq g$ for the updated g, w.p. $1 - \frac{\delta}{2T_1}$. Therefore, by induction, we always have $f(\boldsymbol{x}) - f^* \leq g$ at the beginning of each outer-loop iteration, and the initial condition is provided by the boundedness of f. After T_1 iterations, we have $f(\boldsymbol{x}) - f^* \leq \left(\frac{2}{3}\right)^{\alpha T_1}\left(\frac{4C_2}{C_1^2}\right)^{\frac{1}{2\alpha-1}} \leq \epsilon$, and the overall error probability from the union bound is at most δ.

The overall sample complexity can be bounded by counting the costs from the Reduce and Uniform Search functions. Within each outer loop, the sample complexity of each call of the Reduce function almost exponentially depends on the number of inner-loop iterations. Hence the total cost from the reduce function for each outer-loop iteration is $\widetilde{O}\left(\frac{C_2^{\frac{d-1}{2}}L_{\rm th}^{\frac{d-5}{2}}}{\Delta^{\frac{d+3}{2}}}\log(\frac{C_2L_{\rm th}}{\Delta\delta_{\rm r}})\right)$ for $d \leq 5$.[4] On the other hand, each uniform search costs $\widetilde{O}\left(\frac{C_2^{\frac{d}{2}}L_{\rm th}^d}{g^{\frac{d+4}{2}}}\log\left(\frac{2T_1}{\delta}\right)\right)$. Both can be written as $\widetilde{O}\left(\left(\frac{C_2}{C_1^2}\right)^{\frac{d(d+3)}{2(d+5)}} g^{-2-d\frac{(d+3)(\alpha-1/2)}{d+5}}\log\left(\frac{2T_1}{\delta}\right)\right)$. As a consequence, the overall sample complexity is given by $\widetilde{O}\left(\left(\frac{C_2}{C_1^2}\right)^{\frac{d(d+3)}{2(d+5)}}\epsilon^{-2-d\frac{(d+3)(\alpha-1/2)}{d+5}}\right)$.

Remark 3.2. *The requirement of $d \leq 5$ in the above analysis is introduced when taking the summation of individual costs from Reduce functions. Although the complexity bound in Proposition 5 holds for any d, the dominating term for $d > 5$ is instead from the first Reduce call, which has*

[3] Rigorously, although f could be unbounded, one can reserve $O(\epsilon^{-2})$ samples at the end of the algorithm to compare the function value at the returned point and $f(\mathbf{0})$. The needed simple regret can be achieved in expectation by choosing the point with the lower estimated function value.

[4] Here we presented a straightforward version of the search algorithm for readability. The polylog factor can be improved by letting Δ to be a function of L, which brings in exponential dependency on the number of inner loop iterations. Similar improvements can be achieved for the uniform search.

the largest $b_k - a_k$. This leads to an upper bound in a different form: $\widetilde{O}\left(\left(\frac{C_2}{C_1^2}\right)^{\frac{d-1}{2}} g^{-2-(d-1)(\alpha-\frac{1}{2})}\log\left(\frac{2T_1}{\delta}\right)\right)$. Since the proposed algorithm and the remaining analysis remain valid, one can simply take $L_{\text{th}} = \sqrt{\frac{g}{dC_2}}$ and achieve an overall complexity of $\widetilde{O}\left(\left(\frac{C_2}{C_1^2}\right)^{\frac{5}{2}} g^{-5\alpha+\frac{1}{2}}\log\left(\frac{1}{\delta}\right)\right)$ for $d = 6$, which matches the bound stated in Remark 2.1.

4 Lower Bounds for Optimization Under KL Conditions

The proof is based on the following framework, which can be derived from a standard Kullback–Leibler (KL) divergence argument (see Appendix G for a proof). For any function class $\mathcal{F}_{\rm H}$ and any distribution p defined on $\mathcal{F}_{\rm H}$, we define the *uniform sampling error* to be

$$P_\epsilon \triangleq \inf_{\boldsymbol{x}} \mathbb{P}_{f\sim p}[f(\boldsymbol{x}) - \inf f \geq \epsilon].$$

We also define the *maximum local variance* to be

$$V \triangleq \sup_{\boldsymbol{x}} \mathrm{Var}_{f\sim p}[f(x)].$$

Proposition 7. *For any sampling algorithm to achieve an expected learning error of $\epsilon > 0$ over a function class $\mathcal{F}_\epsilon$, if $P_{2\epsilon/c} \geq c$ for some universal constant $c \in (0, 1)$, and the observation noises are standard Gaussian, then the required sample complexity to achieve a minimax regret of ϵ is at least*

$$T \geq \Omega(1/V).$$

To apply the above proposition, we construct a subclass of $\mathcal{F}$ using the following functions. For any fixed C_1, C_2, and $\alpha \in (\frac{1}{2}, 1)$, let

$$f(x) = \begin{cases} 2C_2 x & \text{if } |x| \in \left(0, 2\sqrt{\frac{\epsilon}{C_2}}\right] \\ 2C_2\left(4\sqrt{\frac{\epsilon}{C_2}}\cdot \mathrm{sign}(x) - x\right) & \text{if } |x| \in \left(2\sqrt{\frac{\epsilon}{C_2}}, 4\sqrt{\frac{\epsilon}{C_2}} - \frac{C_1(6\epsilon)^\alpha}{2C_2}\right] \\ C_1(6\epsilon)^\alpha \cdot \mathrm{sign}(x) & \text{otherwise} \end{cases}$$

and $F(x) = \int_0^x f(y)\mathrm{d}y$.[5] We define

$$f_{\boldsymbol{r}}(\boldsymbol{x}) \triangleq F(||\boldsymbol{r}||_2 - \boldsymbol{x}\cdot\boldsymbol{e_r}) - F(||\boldsymbol{r}||_2) + C_2\left(||\boldsymbol{x}||_2^2 - (\boldsymbol{x}\cdot\boldsymbol{e_r})^2\right)$$

[5]The function f is well defined for sufficiently small ϵ.

for all $\boldsymbol{x} \in \mathbb{R}^d$, where $\boldsymbol{e_r} \triangleq \frac{\boldsymbol{r}}{||\boldsymbol{r}||_2}$. Then we construct the hard instances by letting

$$F_{\boldsymbol{r}}(\boldsymbol{x}) \triangleq \begin{cases} f_{\boldsymbol{r}}(\boldsymbol{x}) & \text{if } f_{\boldsymbol{r}}(\boldsymbol{x}) \leq 0 \\ \left(1 - \exp\left(-f_{\boldsymbol{r}}(\boldsymbol{x})\frac{C_2}{C_1^2}\epsilon^{-2\alpha}\right)\right)\cdot\frac{C_1^2}{C_2}\epsilon^{2\alpha} & \text{otherwise.} \end{cases}$$

The above hard instance function is illustrated in Figure 1. The construction idea is to ensure that the values of the hard instance functions is only close to f^* within a small region, but bounded away from the origin with any sufficiently large r. This property is enabled by the definition of $f_{\boldsymbol{r}}$. Then the function $F_{\boldsymbol{r}}$ is obtained by essentially "clipping" the values of $f_{\boldsymbol{r}}$ at approximately 0 while maintaining differentiability. So that measuring $F_{\boldsymbol{r}}$ at those clipped region only provides a limited amount of information.

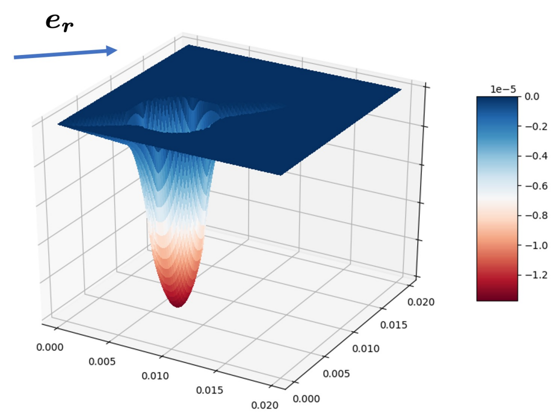

Figure 1: Illustration of a hard instance function, with $\alpha = \frac{2}{3}$, $d = 2$, $C_1 = C_2 = 1$, and $\epsilon = 10^{-6}$.

We restrict the length of $\boldsymbol{r}$ to be within a constant factor of $R_0 = C_1^{-\frac{d+3}{d+5}} C_2^{-\frac{1}{d+5}} \epsilon^{\frac{1}{2}-\frac{d+3}{d+5}(\alpha-\frac{1}{2})}$. Formally, let

$$\mathcal{F}_{\rm H} = \{F_{\boldsymbol{r}} \mid \boldsymbol{r} \in \mathbb{R}^d, ||\boldsymbol{r}||_2 \in [R_0, 2R_0]\}.$$

One can verify that $\mathcal{F}_{\rm H} \subseteq \mathcal{F}(\alpha, C_1, C_2)$ for any ϵ sufficiently small. Then we let p be defined with $\boldsymbol{r}$ having the uniform distribution within the hyperspherical shell $||\boldsymbol{r}||_2 \in [R_0, 2R_0]$. One can verify that for sufficiently small ϵ we have $P_{4\epsilon} > \frac{1}{2}$. Therefore it remains to derive an upper bound of the maximum local variance V.

We first show that the information that can be learned at any point $\boldsymbol{x}$ with $F_{\boldsymbol{r}}(\boldsymbol{x}) \geq 0$ is bounded by a negligible quantity. Formally,

$$\begin{aligned} V &\leq \sup_{\boldsymbol{x}\in\mathbb{R}^d} \mathbb{E}_{F_{\boldsymbol{r}}\sim p}[F_{\boldsymbol{r}}(\boldsymbol{x})^2] \\ &\leq \left(\frac{C_1^2}{C_2}\right)^2 \epsilon^{4\alpha} + \sup_{\boldsymbol{x}\in\mathbb{R}^d} \mathbb{E}_{F_{\boldsymbol{r}}\sim p}[\min\{F_{\boldsymbol{r}}(\boldsymbol{x}), 0\}^2]. \end{aligned}$$

The above bound implies that by replacing $F_{\boldsymbol{r}}$ with $\min\{F_{\boldsymbol{r}}, 0\}$, the resulting difference in V is at most $O\left(\left(\frac{C_1^2}{C_2}\right)^2 \epsilon^{4\alpha}\right)$. Therefore, we can focus on bounding the second term above.

Conditioned on any fixed $||\boldsymbol{r}||_2 = r \in [R_0, 2R_0]$, through a direct estimation, the expectation contained in the second term is given by

$$\mathbb{E}[\min\{F_{\boldsymbol{r}}(\boldsymbol{x}), 0\}^2 \mid ||\boldsymbol{r}||_2 = r] = \begin{cases} O\left(\left(\frac{C_1^2}{C_2}\right)^2 \epsilon^{4\alpha}\right) & \text{if } ||\boldsymbol{x}||_2 \in \left(0, \frac{C_1\epsilon^\alpha}{C_2}\right] \\ O\left(C_1^2\epsilon^{2\alpha}||\boldsymbol{x}||_2^2\left(\frac{\sqrt{C_1\epsilon^\alpha||\boldsymbol{x}||_2/C_2}}{||\boldsymbol{x}||_2}\right)^{d-1}\right) & \text{if } ||\boldsymbol{x}||_2 \in \left(\frac{C_1\epsilon^\alpha}{C_2}, r - 4\sqrt{\frac{\epsilon}{C_2}}\right] \\ O\left(\epsilon^2\left(\frac{\sqrt{\epsilon/C_2}}{r}\right)^{d-1}\right) & \text{if } ||\boldsymbol{x}||_2 \in \left(r - 4\sqrt{\frac{\epsilon}{C_2}}, r + 5\sqrt{\frac{\epsilon}{C_2}}\right] \\ O\left(C_1^2\epsilon^{2\alpha}r^2\left(\frac{\sqrt{C_1\epsilon^\alpha r/C_2}}{r}\right)^{d-1}\right) & \text{otherwise} \end{cases}$$

By taking the supremum over all possible values of $||\boldsymbol{x}||_2$ in each regime, we obtain the following upper bound for the same quantity.

$$\mathbb{E}[\min\{F_{\boldsymbol{r}}(\boldsymbol{x}), 0\}^2 \mid ||\boldsymbol{r}||_2 = r] = \begin{cases} O\left(\left(\frac{C_1^2}{C_2}\right)^2 \epsilon^{4\alpha}\right) & \text{if } ||\boldsymbol{x}||_2 \in \left(0, \frac{C_1\epsilon^\alpha}{C_2}\right] \\ O\left(\epsilon^2\left(\frac{\sqrt{\epsilon/C_2}}{R_0}\right)^{d-1}\right) & \text{if } ||\boldsymbol{x}||_2 \in \left(r - 4\sqrt{\frac{\epsilon}{C_2}}, r + 5\sqrt{\frac{\epsilon}{C_2}}\right] \\ O\left(C_1^2\epsilon^{2\alpha}R_0^2\left(\frac{\sqrt{C_1\epsilon^\alpha R_0/C_2}}{R_0}\right)^{d-1}\right) & \text{otherwise} \end{cases}$$

$$= \begin{cases} O\left(\left(\frac{C_1^2}{C_2}\right)^{\frac{d(d+3)}{2(d+5)}} \epsilon^{2+\frac{d(d+3)}{(d+5)}\left(\alpha-\frac{1}{2}\right)} \cdot \frac{R_0}{\sqrt{\epsilon/C_2}}\right) & \text{if } ||\boldsymbol{x}||_2 \in \left(r - 4\sqrt{\frac{\epsilon}{C_2}}, r + 5\sqrt{\frac{\epsilon}{C_2}}\right] \\ O\left(\left(\frac{C_1^2}{C_2}\right)^{\frac{d(d+3)}{2(d+5)}} \epsilon^{2+\frac{d(d+3)}{(d+5)}\left(\alpha-\frac{1}{2}\right)}\right) & \text{otherwise} \end{cases}$$

Then we take the expectation of the above upper bound over $||\boldsymbol{r}||_2$, to obtain an upper bound of V. Specifically, note that

$$\mathbb{E}_{F_{\boldsymbol{r}}\sim p}[\min\{F_{\boldsymbol{r}}(\boldsymbol{x}), 0\}^2] = \mathbb{E}_r[\mathbb{E}[\min\{F_{\boldsymbol{r}}(\boldsymbol{x}), 0\}^2 \mid ||\boldsymbol{r}||_2 = r]],$$

where r takes a distribution on $[R_0, 2R_0]$ with a density proportional to $(r/R_0)^{d-1}$. One can integrate the above bound by considering the two possible cases separately. The first scenario $||\boldsymbol{x}||_2 \in \left(r - 4\sqrt{\frac{\epsilon}{C_2}}, r + 5\sqrt{\frac{\epsilon}{C_2}}\right]$ occurs with a probability of at most $O\left(\frac{\sqrt{\epsilon/C_2}}{R_0}\right)$. Therefore, it contributes to the local variance V by at most $O\left(\left(\frac{C_1^2}{C_2}\right)^{\frac{d(d+3)}{2(d+5)}} \epsilon^{2+\frac{d(d+3)}{(d+5)}\left(\alpha-\frac{1}{2}\right)}\right)$. Note that this is identical to the contribution from the second scenario, which occurs with at most a probability of 1.[6] We can conclude that $V = O\left(\left(\frac{C_1^2}{C_2}\right)^{\frac{d(d+3)}{2(d+5)}} \epsilon^{2+\frac{d(d+3)}{(d+5)}\left(\alpha-\frac{1}{2}\right)}\right)$, and Proposition 7 leads to the needed statement.

Acknowledgements

JDL, QY, and QL acknowledge support of the ARO under MURI Award W911NF-11-1-0304, the Sloan Research Fellowship, NSF CCF 2002272, NSF IIS 2107304, NSF CIF 2212262, and NSF CAREER Award 2144994.

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. In *NIPS*, volume 11, pages 2312–2320.

Abernethy, J., Hazan, E., and Rakhlin, A. (2008). Competing in the dark: An efficient algorithm for bandit linear optimization. In *Conference on Learning Theory*, pages 263–273. COLT.

Agarwal, A., Dekel, O., and Xiao, L. (2010). Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Colt*, pages 28–40. Citeseer.

Agarwal, A., Dekel, O., and Xiao, L. (2011a). Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Proceedings of the 23rd Annual Conference on Learning*, page 2840. PMLR.

Agarwal, A., Foster, D. P., Hsu, D., Kakade, S. M., and Rakhlin, A. (2013). Stochastic convex optimization with bandit feedback. In *SIAM Journal on Optimization*, page 213–240.

Agarwal, A., Foster, D. P., Hsu, D. J., Kakade, S. M., and Rakhlin, A. (2011b). Stochastic convex optimization with bandit feedback. *Advances in Neural Information Processing Systems*, 24:1035–1043.

Agarwal, A., Kakade, S. M., Lee, J. D., and Mahajan, G. (2021). On the theory of policy gradient methods: Optimality, approximation, and distribution shift. *Journal of Machine Learning Research*, 22(98):1–76.

[6] In fact, the value of R_0 is exactly selected for them to be identical, in order to minimize their total contribution to V.

Arjevani, Y., Carmon, Y., Duchi, J. C., Foster, D. J., Srebro, N., and Woodworth, B. (2019). Lower bounds for non-convex stochastic optimization. *arXiv preprint arXiv:1912.02365*.

Audibert, J.-Y., Bubeck, S., et al. (2009). Minimax policies for adversarial and stochastic bandits. In *COLT*, volume 7, pages 1–122.

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

Balasubramanian, K. and Ghadimi, S. (2018). Zeroth-order (non)-convex stochastic optimization via conditional gradient and gradient updates. *Advances in Neural Information Processing Systems*, 31.

Bubeck, S., Lee, Y. T., and Eldan, R. (2017). Kernel-based methods for bandit convex optimization. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 72–85.

Bubeck, S. and Mikulincer, D. (2020). How to trap a gradient flow. In Abernethy, J. and Agarwal, S., editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 940–960. PMLR.

Chewi, S., Bubeck, S., and Salim, A. (2022). On the complexity of finding stationary points of smooth functions in one dimension. *arXiv preprint arXiv:2209.07513*.

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

Cutkosky, A. and Orabona, F. (2019). Momentum-based variance reduction in non-convex sgd. *Advances in neural information processing systems*, 32.

Dani, V., Hayes, T. P., and Kakade, S. M. (2008). Stochastic linear optimization under bandit feedback. *The 21st Annual Conference on Learning Theory*.

Duchi, J. C., Jordan, M. I., Wainwright, M. J., and Wibisono, A. (2015). Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, 61(5):2788–2806.

Fang, C., Li, C. J., Lin, Z., and Zhang, T. (2018). Spider: Near-optimal non-convex optimization via stochastic path-integrated differential estimator. *Advances in Neural Information Processing Systems*, 31.

Flaxman, A. D., Kalai, A. T., and McMahan, H. B. (2005). Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the 16th annualACM-SIAM symposium on Discrete Algorithms*, page 385–394.

Hazan, E. and Levy, K. (2014a). Bandit convex optimization: Towards tight bounds. *Advances in Neural Information Processing Systems*, 27.

Hazan, E. and Levy, K. Y. (2014b). Bandit convex optimization: Towards tight bounds. In *MIPS*.

Huang, B., Huang, K., Kakade, S. M., Lee, J. D., Lei, Q., Wang, R., and Yang, J. (2021). Optimal gradient-based algorithms for non-concave bandit optimization. *arXiv preprint arXiv:2107.04518*.

Ji, K., Wang, Z., Zhou, Y., and Liang, Y. (2019). Improved zeroth-order variance reduced algorithms and analysis for nonconvex optimization. In *International conference on machine learning*, pages 3100–3109. PMLR.

Karimi, H., Nutini, J., and Schmidt, M. (2016). Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 795–811. Springer.

Kizilcec, R. F., Reich, J., Yeomans, M., Dann, C., Brunskill, E., Lopez, G., Turkay, S., Williams, J. J., and Tingley, D. (2020). Scaling up behavioral science interventions in online education. *Proceedings of the National Academy of Sciences*, 117(26):14900–14905.

Kober, J., Bagnell, A., and Peters, J. (2014). Reinforcement learning in robotics: A survey. *Springer Tracts in Advanced Robotics*, 97:9–67.

Kurdyka, K. (1998). On gradients of functions definable in o-minimal structures. In *Annales de l'institut Fourier*, volume 48, pages 769–783.

Lan, G. (2019). First-order and stochastic optimization methods for machine learning. In *Springer Series in the Data Sciences*, pages 305–328. Springer.

Larson, J., Menickelly, M., and Wild, S. M. (2019). Derivative-free optimization methods. *Acta Numerica*, 28:287–404.

Lattimore, T. and György, A. (2021). Improved regret for zeroth-order stochastic convex bandits. In *Proceedings of Machine Learning Research*, page 134:1–27.

Lattimore, T. and Hao, B. (2021). Bandit phase retrieval. *arXiv preprint arXiv:2106.01660*.

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

Liu, L., Cheng, M., Hsieh, C.-J., and Tao, D. (2018a). Stochastic zeroth-order optimization via variance reduction method. *arXiv preprint arXiv:1805.11811*.

Liu, S., Kailkhura, B., Chen, P.-Y., Ting, P., Chang, S., and Amini, L. (2018b). Zeroth-order stochastic variance reduction for nonconvex optimization. *Advances in Neural Information Processing Systems*, 31.

Moulines, E. and Bach, F. (2011). Non-asymptotic analysis of stochastic approximation algorithms for machine learning. *Advances in neural information processing systems*, 24.

Nemirovskij, A. S. and Yudin, D. B. (1983). Problem complexity and method efficiency in optimization.

Nesterov, Y. and Spokoiny, V. (2017). Random gradient-free minimization of convex functions. In *Found Comput Math (2017)*, page 17:527–566.

Rusmevichientong, P. and Tsitsiklis, J. N. (2010). Linearly parameterized bandits. *Mathematics of Operations Research*, 35(2):395–411.

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

Wang, Y., Balakrishnan, S., and Singh, A. (2019a). Optimization of smooth functions with noisy observations: Local minimax rates. *IEEE Transactions on Information Theory*, 65(11):7350–7366.

Wang, Y., Du, S., Balakrishnan, S., and Singh, A. (2018). Stochastic zeroth-order optimization in high dimensions. In *International Conference on Artificial Intelligence and Statistics*, pages 1356–1365. PMLR.

Wang, Z., Ji, K., Zhou, Y., Liang, Y., and Tarokh, V. (2019b). Spiderboost and momentum: Faster variance reduction algorithms. *Advances in Neural Information Processing Systems*, 32.

Yu, C., Liu, J., and Nemati, S. (2020). Reinforcement learning in healthcare: A survey. In *arXiv preprint arXiv:1908.08796*.

Yu, Q., Wang, Y., Huang, B., lei, Q., and Lee, J. D. (2023). Sample complexities for quadratic bandits: Hessian dependent bounds and optimal algorithms. *online preprint*.

Zhang, L., Yang, T., Jin, R., and Zhou, Z.-H. (2015). Online bandit learning for a special class of non-convex losses. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Zhao, P. and Lai, L. (2021). Optimal stochastic nonconvex optimization with bandit feedback. *arXiv preprint arXiv:2103.16082*.

Zhou, D., Xu, P., and Gu, Q. (2018). Stochastic nested variance reduction for nonconvex optimization. *Advances in Neural Information Processing Systems*, 31.

A Gradient Based Method for Optimization under KL conditions

The main intuition of gradient-based method is as follows. Starting from any point $\boldsymbol{x} \in \mathbb{R}$, if one can measure an approximated gradient of f at the same point, then due to the smoothness condition, the function value for the updated point $\boldsymbol{x}' = \boldsymbol{x} - \frac{1}{2C_2}\nabla f(\boldsymbol{x})$ satisfies $f(\boldsymbol{x}') \leq f(\boldsymbol{x}) - \frac{1}{4C_2}||\nabla f(\boldsymbol{x})||_2^2$.

More specifically, when using zero-th order measurements, one can measure the partial gradient $f_{\boldsymbol{u}} = \boldsymbol{u} \cdot \nabla f$ for any unit vector $\boldsymbol{u}$ and achieve a decent of $\frac{1}{4C_2}||\boldsymbol{u} \cdot \nabla f(\boldsymbol{x})||_2^2$. In expectation, we have $\mathbb{E}[|f_{\boldsymbol{u}}|^2] = \frac{1}{d}||\nabla f(\boldsymbol{x})||_2^2$ for uniformly random $\boldsymbol{u}$. Hence, we can measure $\min\{f_{\boldsymbol{u}}, 0\}$ up to an error of $O(\frac{1}{\sqrt{d}}||\nabla f(\boldsymbol{x})||_2)$ with high probability by sampling at $\boldsymbol{x}$ and $\boldsymbol{x} + \frac{z}{2C_2\sqrt{d}}\boldsymbol{u}$ for any lower bound $z \in (0, ||\nabla f(\boldsymbol{x})||_2]$. It takes $O(\log \frac{1}{\delta})$ random samples of $\boldsymbol{u}$ to observe $f_{\boldsymbol{u}} \geq \frac{1}{\sqrt{d}}||\nabla f(\boldsymbol{x})||_2$ with probability $1 - \delta$, and the sample complexity for each fixed $\boldsymbol{u}$ is $\widetilde{O}(\frac{C_2^2 d^2}{z^4} \log \frac{1}{\delta})$.

Recall the KL condition. It takes $O\left(\frac{C_2}{C_1^2} d g^{1-2\alpha}\right)$ such decent steps to reduce $f(\boldsymbol{x}) - f^*$ by half when $f(\boldsymbol{x}) - f^* \leq g$, and it is achieved by choosing $z = \Theta(C_1 g^{\alpha})$. Hence, the overall sample complexity is $\widetilde{O}\left(\frac{C_2^3}{C_1^6} d^3 \epsilon^{1-6\alpha} \log^2 \frac{1}{\delta}\right)$ to achieve a simple regret ϵ with error probability δ. This implies a complexity upper bound of $O\left(\frac{C_2^3}{C_1^6} d^3 \epsilon^{1-6\alpha}\right)$ for achieving an expected simple regret of ϵ.

For clarity, we state a simplified procedure in Algorithm 4, which achieves the same $\widetilde{O}\left(\frac{C_2^3}{C_1^6} d^3 \epsilon^{1-6\alpha} \log^2 \frac{1}{\delta}\right)$ complexity guarantee.

Algorithm 4 Gradient-Based Method

```
Initialize x = 0, g = (4C_2/C_1^2)^(1/(2α−1)). Let T_1 = ⌈2log_{3/2}(g/ε)⌉.
for t_1 ← 1 to T_1 do
    Let z = C_1((2/3)g)^α, r = z/(8C_2√d), δ_t = δ/(2·2^(T_1−t)), T_2 = ⌈4g/(3C_2 r^2)⌉ · ⌈log_{10/9}(1/δ)⌉^+.
    for t_2 ← 1 to T_2 do
        Let u be uniformly random from the unit hypersphere
        Let D = Sample(x, C_2 r^2/4, δ_t/(4T_2)) − Sample(x + ru, C_2 r^2/4, δ_t/(4T_2))
        if D ≥ C_2 r^2/2 then
            x = x + ru
        end if
    end for
    g = (2/3)^α g
end for
    return x
```

B Monotonicity of Optimal Sample Complexity in d for Optimization under KL conditions

This can be proved by the fact that optimizing over the function class $\mathcal{F}(\alpha, C_1, C_2)$ for any smaller d is equivalent to optimizing over functions defined on any higher dimensional space, but each only depends on a certain fraction of input entries. The latter set belongs to $\mathcal{F}(\alpha, C_1, C_2)$ for larger d, so any algorithm stated for a higher dimensional space can be applied to lower dimensional spaces though a direct projection.

C Proof of Proposition 3

Proof. We first prove the upper bound on $f(\mathbf{0}) - f^*$. By the stated assumptions, f is differentiable at $\mathbf{0}$. Hence, we can let $\boldsymbol{x}' \triangleq -\frac{1}{2C_2}\nabla f(\mathbf{0})$, and the smoothness condition implies that

$$f(\mathbf{0}) - f^* \geq f(\mathbf{0}) - f(\boldsymbol{x}') \geq \frac{||\nabla f(\mathbf{0})||_2^2}{4C_2}. \tag{2}$$

Recall the KL inequality, the RHS of above is lower bounded by $\frac{C_1^2|f(\mathbf{0})-f^*|^{2\alpha}}{4C_2}$. Therefore, we obtain a single-variate inequality on $f(\mathbf{0})-f^*$. For $\alpha\in[\frac{1}{2},1)$, the solution is given by

$$0\leq f(\mathbf{0})-f^*\leq\left(\frac{4C_2}{C_1^2}\right)^{\frac{1}{2\alpha-1}}.$$

Now we prove the rest of the statements. For any $g\in[0,f(\boldsymbol{x})-f^*]$, let $R(g)\triangleq\inf\{R\in[0,+\infty]\mid\exists\,\boldsymbol{r}\in\mathcal{X},||\boldsymbol{r}-\boldsymbol{x}||_2\leq R,f(\boldsymbol{r})-f^*\leq g\}$. By definition, $R(g)$ is non-increasing and $R(f(\boldsymbol{x})-f^*)=0$. It suffices to prove that $R(0)\leq L_{\boldsymbol{x}}$, and $R(0)<L$ holds under the additional conditions.

We first show that $R(g)$ is right-continuous under the natural topology of domain $[0,+\infty]$. By monotonicity, the right limit of $R(g)$ exists for any g, and we denote it by $R^+(g)$. It suffices to show that $R(g)\leq R^+(g)$. Consider the non-trivial case where $R^+(g)<+\infty$. There is a sequence of points $\boldsymbol{x}_1,\boldsymbol{x}_2...$ in $\mathcal{X}$ with $\lim_{k\to\infty}||\boldsymbol{x}_k-\boldsymbol{x}||_2=R^+(g)$, such that

$$\limsup_{k\to\infty}f(\boldsymbol{x}_k)\leq f^*+g. \tag{3}$$

By compactness, a subsequence of the above points converges, and we denote their limit by $\boldsymbol{x}_g^*$. Then by differentiability, the value of $f(\boldsymbol{x}_g^*)$ is identical to the limit of function values of the subsequence, which is no greater than f^*+g. Hence, we have proved a stronger statement showing that $R(g)$ can be defined as a minimum instead of infimum, and $\boldsymbol{x}_g^*$ serves as an instance for the upper bound $R(g)\leq R^+(g)$.

Now consider the left neighbourhood of any g. We prove that

$$\limsup_{g'\to g^-}\frac{R(g')-R(g)}{g-g'}\leq\frac{1}{C_1g^\alpha}. \tag{4}$$

Let $\boldsymbol{x}_g^*$ be defined as in the above steps.[7] If $f(\boldsymbol{x}_g^*)<f^*+g$, we have that $R(g')=R(g)$ for any $g'\in[f(\boldsymbol{x}_g^*)-f^*,g]$, and inequality (4) clearly holds. Hence, we can assume that $f(\boldsymbol{x}_g^*)=f^*+g$, and KL inequality implies that $||\nabla f(\boldsymbol{x}_g^*)||_2\geq C_1g^\alpha$. This bound implies that

$$\begin{aligned}&\limsup_{g'\to g^-}\frac{R(g')-R(g)}{g-g'}\\&\qquad\leq\limsup_{\phi\to0^+}\frac{||\boldsymbol{x}_g^*-\phi\nabla f(\boldsymbol{x}_g^*)||_2-R(g)}{f(\boldsymbol{x}_g^*)-f(\boldsymbol{x}_g^*-\phi\nabla f(\boldsymbol{x}_g^*))}\\&\qquad\leq\frac{1}{||\nabla f(\boldsymbol{x}_g^*)||_2}\leq\frac{1}{C_1g^\alpha}.\end{aligned}$$

Combine the above two facts, we have that $R(g)+\frac{g^{1-\alpha}}{C_1(1-\alpha)}$ is non-decreasing. Consequently,

$$\begin{aligned}R(g)&\leq R(f(\boldsymbol{x})-f^*)+\frac{(f(\boldsymbol{x})-f^*)^{1-\alpha}}{C_1(1-\alpha)}-\frac{g^{1-\alpha}}{C_1(1-\alpha)}\\&=L_{\boldsymbol{x}}-\frac{g^{1-\alpha}}{C_1(1-\alpha)}\end{aligned} \tag{5}$$

for any g. Taking $g=0$, we have $R(0)\leq L_{\boldsymbol{x}}$, which proves the second statement.

Now that if $L_{\boldsymbol{y}}>L_{\boldsymbol{x}}-L$ for any point $\boldsymbol{y}$ with $||\boldsymbol{y}-\boldsymbol{x}||_2=L$ for some $L\leq L_0$. We have $f(\boldsymbol{y})>f^*+(C_1(1-\alpha)(L_{\boldsymbol{x}}-L))^{\frac{1}{1-\alpha}}$ for any such $\boldsymbol{y}$. By the fact that any $R(g)$ is achieved as a minimum, we have $R(g)\neq L$ for all $g\leq g_L\triangleq(C_1(1-\alpha)(L_{\boldsymbol{x}}-L))^{\frac{1}{1-\alpha}}$. Further, we also have $R(g_L)\leq L$ due to inequality (5). Note that the two facts we proved earlier implies the continuity of $R(g)$. Hence, $R(0)<L$, and the third statement is proved. $\square$

[7] When $g=f(\boldsymbol{x})-f^*$, let $\boldsymbol{x}_g^*\triangleq\boldsymbol{x}$.

D Proof details on inequality (1)

Here we used the fact that the interval length L is reduced by a factor of $\frac{1}{2}$ within each global search step. Therefore, we always have $2L \geq L_{\text{th}}$ within the loop, and then inequality (1) implies that $f(x_{\text{c}} \pm L) \geq f(x_{\text{c}}) - 2C_0 \cdot L$.

For the induction proof, we first check the initial condition. When $L = L_0$,

$$\begin{aligned} f(x_{\text{c}} \pm L) &\geq f^* \\ &\geq f(x_{\text{c}}) - g_t \\ &= f(x_{\text{c}}) - C_0 L_0 \geq f(x_{\text{c}}) - (C_0 + \Delta_C)\, L. \end{aligned}$$

Hence, inequality (1) holds at the beginning of the global search, and it suffices to consider the case where L_0 is large enough to enter the global search loop.

Then assume inequality (1) holds at the beginning of any given search step, we prove that it holds at the end of the same step. There are three possible cases. When x_{c} is not updated, the interval is updated to $[x_{\text{c}} - \frac{L}{2}, x_{\text{c}} + \frac{L}{2}]$. From Proposition 2 and the argmin condition, we have the following bounds with high probability,

$$\begin{aligned} f\left(x_{\text{c}} \pm \frac{L}{2}\right) &\geq f(x_{\text{c}}) - (C_0 + \Delta_C) \cdot \frac{L}{2} - 2\epsilon_L \\ &> f(x_{\text{c}}) - \left(C_0 + 2^{\frac{2}{3}} \Delta_C\right) \cdot \frac{L}{2}, \end{aligned}$$

which proves the induction statement. Among the other two cases, without loss of generality, we assume the updated interval is given by $[x_{\text{c}}, x_{\text{c}} + L]$ in terms of the initial parameters. From Proposition 2 and the induction assumption, we have

$$\begin{aligned} f\left(x_{\text{c}} + \frac{L}{2}\right) &\leq f(x_{\text{c}}) - (C_0 + \Delta_C) \cdot \frac{L}{2} + 2\epsilon_L \\ &\leq f(x_{\text{c}} + L) + (C_0 + \Delta_C) \cdot \frac{L}{2} + 2\epsilon_L \\ &< f(x_{\text{c}} + L) + (C_0 + 2^{\frac{2}{3}} \Delta_C) \cdot \frac{1}{2} L. \end{aligned}$$

We also have

$$\begin{aligned} f\left(x_{\text{c}} + \frac{L}{2}\right) &\leq f\left(x_{\text{c}}\right) \\ &< f\left(x_{\text{c}}\right) + (C_0 + 2^{\frac{2}{3}} \Delta_C) \cdot \frac{1}{2} L. \end{aligned}$$

Therefore, the induction statement for the updated interval also holds in these case and the proof for global search phase is concluded.

E Proof of Proposition 4

Proof. We first generalize Proposition 3 by showing that for any $\boldsymbol{x} \in \mathcal{X}$ and any set $\mathcal{B}$ that contains $\boldsymbol{x}$, if $L_{\boldsymbol{y}} > L_{\boldsymbol{x}} - ||\boldsymbol{y} - \boldsymbol{x}||_2$ for any $\boldsymbol{y} \in \partial\mathcal{B}$, then a global minima of f exists within $\mathcal{B}$. The proof can be obtained by including an additional constraint $\boldsymbol{r} \in \mathcal{B}$ to the definition of $R(g)$, i.e., let $R(g) \triangleq \inf\{R \in [0, +\infty] \mid \exists\, \boldsymbol{r} \in \mathcal{B}, ||\boldsymbol{r} - \boldsymbol{x}||_2 \leq R, f(\boldsymbol{r}) - f^* \leq g\}$. By the given assumptions on $L_{\boldsymbol{x}}$, any $R(g)$ can not be achieved at the boundary of $\mathcal{B}$. Therefore, the proof of right continuity still holds and $R(g)$ can be achieved as a minimum with some point $\boldsymbol{r}$ in the interior of $\mathcal{B}$. Then following the same arguments, we have $R(0) \leq L_{\boldsymbol{x}} < +\infty$. Hence, f^* can be achieved by a point in the interior of $\mathcal{B}$.

Going back to the Proposition, when $f(\boldsymbol{x}) - f^* > \left(\frac{C}{C_1(1-\alpha)}\right)^{\frac{1}{\alpha}}$, we have

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) - C_1(1-\alpha) \cdot (f(\boldsymbol{x}) - f^*)^{\alpha} \cdot ||\boldsymbol{x} - \boldsymbol{y}||_2$$

for all $\boldsymbol{y} \in \partial\mathcal{B}$. By the convexity of $L_{\boldsymbol{x}}$ with respect to $f(\boldsymbol{x}) - f^*$, we have $L_{\boldsymbol{y}} > L_{\boldsymbol{x}} - ||\boldsymbol{y} - \boldsymbol{x}||_2$. Hence, the achievability of f^* in $\mathcal{B}$ is proved.

□

F Proof of Proposition 5

Proof. Let $\widehat{f}(\boldsymbol{y})$ denote the value returned for each call of the Sample function. By union bound and Proposition 2, we have $\widehat{f}(\boldsymbol{y}) \in [f(\boldsymbol{y}) - \epsilon_{\rm r}, f(\boldsymbol{y}) + \epsilon_{\rm r}]$ for all sampled $\boldsymbol{y}$ w.p. $1 - \delta_{\rm r}$. We will focus on this event in the rest of the proof.

Depending on the value of $\boldsymbol{y}^*$, we consider two possible cases. First, if $\boldsymbol{y}^* = \boldsymbol{x}$, we let $F(\boldsymbol{y}) \triangleq (C_0 + 2\Delta)||\boldsymbol{y} - \boldsymbol{y^*}||_2 + f(\boldsymbol{y})$ and $\boldsymbol{y}_F \triangleq \text{argmin}_{\boldsymbol{y} \in \mathcal{S}_p \cup \mathcal{S}_q} F(\boldsymbol{y})$. To prove that $\boldsymbol{y}^*$ is a dominant point on the new set, it suffices to show that $F(\boldsymbol{y}_F) > F(\boldsymbol{x})$. We prove this statement by contradiction. Assume that $F(\boldsymbol{y}_F) \leq F(\boldsymbol{x})$. By the dominant condition of $\boldsymbol{x}$ on $\mathcal{B}$, the point $\boldsymbol{y}_F$ can not be on the boundary of $\boldsymbol{B}$. Hence, we have $\nabla F(\boldsymbol{y}_F)$ is orthogonal to $\mathcal{S}_p$ and $\mathcal{S}_q$. Therefore, the value of $F(\boldsymbol{y}_F)$ can be bounded by the sampled results using the smoothness condition.

Without loss of generality, assume $\boldsymbol{y}_F \in \mathcal{S}_p$. We rely on the following fact to show a smoothness condition of F.

Proposition 8. *For any real numbers $a < x < b$, both the outputs p, q from Balanced Partition(a, b, x) belong to (a, b) and the distance from any of them to any other 4 points belongs to $[(b-a)/5, 4(b-a)/5]$.*

The above proposition provides a lower bound of $||\boldsymbol{y} - \boldsymbol{x}||_2$ for any $\boldsymbol{y} \in \mathcal{S}_p$. Hence, by taking the derivatives, one can show that

$$\begin{aligned} F(\boldsymbol{y}) - F(\boldsymbol{y}_F) &\leq \left(\frac{5(C_0 + 2\Delta)}{2(b_k - a_k)} + C_2\right) ||\boldsymbol{y}_F - \boldsymbol{y}||_2^2 \\ &\leq 2C_2||\boldsymbol{y}_F - \boldsymbol{y}||_2^2. \end{aligned}$$

We remind the reader the definition of ϵ-net.

Definition 2 (ϵ-net)**.** *Let $\mathcal{Y}$ be a subset of $\mathcal{S}$ and $\epsilon > 0$ be a general parameter. We say $\mathcal{Y}$ is an* ϵ-net *of $\mathcal{S}$, if the packing radius of $\mathcal{Y}$ is no smaller than $\epsilon/2$, and the covering radius is no greater than ϵ.*

By the covering property, there is a sampled point $\widehat{\boldsymbol{y}}_F$ on $\mathcal{S}_p$ with $||\boldsymbol{y}_F - \boldsymbol{y}||_2 \leq \sqrt{\epsilon_{\rm r}/C_2}$. Consequently,

$$F(\widehat{\boldsymbol{y}}_F) - F(\boldsymbol{y}_F) < 2\epsilon_{\rm r}.$$

Recall that $f(\widehat{\boldsymbol{y}}_F)$ is sampled with an error of at most $\epsilon_{\rm r}$. We have

$$\widehat{f}(\widehat{\boldsymbol{y}}_F) + (C_0 + 2\Delta)||\boldsymbol{x} - \widehat{\boldsymbol{y}}_F||_2 - F(\boldsymbol{y}_F) < 3\epsilon_{\rm r}.$$

Now we use the fact that $\boldsymbol{y}^* = \boldsymbol{x}$ and $f(\boldsymbol{x})$ is also sampled with an accuracy of $\epsilon_{\rm r}$.

$$\widehat{f}(\widehat{\boldsymbol{y}}_F) + (C_0 + \Delta)||\boldsymbol{x} - \widehat{\boldsymbol{y}}_F||_2 \geq \widehat{f}(\widehat{\boldsymbol{x}}) > F(\boldsymbol{x}) - \epsilon_{\rm r}.$$

Combine the above two inequalities, we have $F(\boldsymbol{y}_F) - F(\boldsymbol{x}) > \Delta||\boldsymbol{x} - \widehat{\boldsymbol{y}}_F||_2 - 4\epsilon_{\rm r} \geq 0$ from Proposition 8, which contradicts the assumption on $\boldsymbol{y}_F$.

For the other case, we have $\boldsymbol{y}^* \in \mathcal{S}_p \cup \mathcal{S}_q$. Then by the sampling guarantee and the definition of $\boldsymbol{y}^*$, we have

$$\begin{aligned} f(\boldsymbol{x}) - f(\boldsymbol{y}^*) &\geq \widehat{f}(\boldsymbol{x}) - \widehat{f}(\boldsymbol{y}^*) - 2\epsilon_{\rm r} \\ &\geq (C_0 + \Delta)||\boldsymbol{x} - \boldsymbol{y}^*||_2 - 2\epsilon_{\rm r}. \end{aligned}$$

Recall that $\Delta||\boldsymbol{x} - \boldsymbol{y}^*|| \geq 2\epsilon_{\rm r}$ can be implied from Proposition 8. The above inequality shows that

$$f(\boldsymbol{x}) - f(\boldsymbol{y}^*) \geq C_0||\boldsymbol{x} - \boldsymbol{y}^*||_2.$$

Using triangle inequality, $\boldsymbol{y}^*$ is also a dominant point of $\mathcal{B}$ with $C = C_0$. Obviously, the same holds for a larger parameter $C = C_0 + 2\Delta$.

Now without loss of generality, we can assume that $\boldsymbol{y}^* \in \mathcal{S}_p$. It remains to show that $F(\boldsymbol{y}) > F(\boldsymbol{y}^*)$ for any $\boldsymbol{y} \in \mathcal{S}_q$. Similar to the first case, we assume the opposite and define $\boldsymbol{y}_q \triangleq \text{argmin}_{\boldsymbol{y} \in \mathcal{S}_q} F(\boldsymbol{y})$. Recall that $\boldsymbol{y}^*$ dominates $\boldsymbol{B}$. The point $\boldsymbol{y}_q$ is not on the boundary of $\boldsymbol{B}$. Following the same arguments used in the first case, one can show the existence of $\widehat{\boldsymbol{y}}_F \in \mathcal{S}_{\text{net},q}$ for the following inequalities to hold.

$$\widehat{f}(\widehat{\boldsymbol{y}}_F) + (C_0 + 2\Delta)||\boldsymbol{y}^* - \widehat{\boldsymbol{y}}_F||_2 - F(\boldsymbol{y}_F) < 3\epsilon_{\text{r}}$$
$$\widehat{f}(\widehat{\boldsymbol{y}}_F) + (C_0 + \Delta)||\boldsymbol{x} - \widehat{\boldsymbol{y}}_F||_2$$
$$> F(\boldsymbol{y}^*) + (C_0 + \Delta)||\boldsymbol{x} - \widehat{\boldsymbol{y}^*}||_2 - \epsilon_{\text{r}}.$$

Combine the above inequalities and use the triangle inequality, we can conclude that

$$F(\boldsymbol{y}_F) - F(\boldsymbol{y}^*) > \Delta||\boldsymbol{y}^* - \widehat{\boldsymbol{y}}_F||_2 - 4\epsilon_{\text{r}} \geq 0.$$

Finally, we prove the sample complexity. We use the following well know upper bound for the size of ϵ-nets.

Proposition 9. *For any fixed d and any set $\mathcal{B}$ that belongs to $\mathbb{R}^d$, an ϵ-net of $\mathcal{B}$ exists. If*

$$V \triangleq \text{Vol}\{\boldsymbol{x} \in \mathbb{R}^d \mid \exists\, \boldsymbol{y} \in \mathcal{B}, ||\boldsymbol{x} - \boldsymbol{y}||_2 \leq \epsilon/2\}$$

is finite, then any ϵ-net is finite, with a size of $O(V/\epsilon^d)$.

The algorithm runs $O(|\mathcal{S}|)$ numbers of Sample calls. Given the stated assumptions and the above proposition, we have

$$|\mathcal{S}| = O\left(\frac{\prod_{j\neq k}(b_j - a_j + \sqrt{\epsilon_{\text{r}}/C_2})}{\sqrt{\epsilon_{\text{r}}/C_2}^{d-1}}\right)$$
$$= O\left(\frac{(b_k - a_k)^{\frac{d-1}{2}}}{(\Delta/C_2)^{\frac{d-1}{2}}}\right).$$

Hence, the overall sample complexity is $O\left(\frac{|\mathcal{S}|}{\epsilon_{\text{r}}^2}\log(\frac{|\mathcal{S}|}{\delta})\right) = O\left(\frac{C_2^{\frac{d-1}{2}}(b_k-a_k)^{\frac{d-5}{2}}}{\Delta^{\frac{d+3}{2}}}\log(\frac{C_2(b_k-a_k)}{\Delta\delta})\right)$. □

G Proof of Proposition 7

Consider any algorithm that achieves a minimax regret of ϵ with T samples. Let $\widehat{\boldsymbol{x}}_f$ denote the returned estimator given any fixed function f and let $\widehat{\boldsymbol{x}}_p$ denote the estimator when the observation model at each point $\boldsymbol{x}$ is $\boldsymbol{y} = \mathbb{E}_{f\sim p}[f(\boldsymbol{x})] + w$. We have

$$\begin{aligned}\epsilon \geq &\mathbb{E}_{f\sim p}[\mathbb{E}_{\widehat{\boldsymbol{x}}_f}[f(\widehat{\boldsymbol{x}}_f) - \inf f]]\\ \geq &\mathbb{E}_{f\sim p}[\mathbb{P}_{\widehat{\boldsymbol{x}}_f}[f(\widehat{\boldsymbol{x}}_f) - \inf f \geq 2\epsilon/c]] \cdot 2\epsilon/c\\ \geq &\mathbb{E}_{f\sim p}[\mathbb{P}_{\widehat{\boldsymbol{x}}_p}[f(\widehat{\boldsymbol{x}}_p) - \inf f \geq 2\epsilon/c]\\ &- \text{TV}(\widehat{\boldsymbol{x}}_f, \widehat{\boldsymbol{x}}_p)] \cdot 2\epsilon/c,\end{aligned} \tag{6}$$

where TV$(\cdot)$ denotes the total variation distance between the distributions of the given variables. Let $\mathcal{O}_f$, $\mathcal{O}_p$ denote the sets of all T action-observation pairs under each model. Because $\widehat{\boldsymbol{x}}_f$ and $\widehat{\boldsymbol{x}}_p$ are determined from the same algorithm, we have

$$\begin{aligned}\mathbb{E}_{f\sim p}[\text{TV}(\widehat{\boldsymbol{x}}_f, \widehat{\boldsymbol{x}}_p)] &\leq \mathbb{E}_{f\sim p}[\text{TV}(\mathcal{O}_f, \mathcal{O}_p)]\\ &\leq \sqrt{\frac{1}{2}\mathbb{E}_{f\sim p}[\text{KL}(\mathcal{O}_p||\mathcal{O}_f)]},\end{aligned}$$

where KL$(\cdot)$ denotes the KL distance. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, ...\boldsymbol{x}_T$ denote the T sample points in $\mathcal{O}_p$. By the additive Gaussian noise assumption,

$$\text{KL}(\mathcal{O}_p||\mathcal{O}_f) = \frac{1}{2}\sum_{t=1}^{T}(f(\boldsymbol{x}_t) - \mathbb{E}_{f\sim p}[f(\boldsymbol{x}_t)])^2.$$

Therefore, we have

$$\begin{aligned}\mathbb{E}_{f\sim p}[\mathrm{TV}(\widehat{\boldsymbol{x}}_f, \widehat{\boldsymbol{x}}_p)] &\leq \frac{1}{2}\sqrt{\sum_{t=1}^{T} \mathrm{Var}_{f\sim p}[f(\boldsymbol{x}_t)]} \\ &\leq \frac{1}{2}\sqrt{VT}.\end{aligned}$$

Consequently, the condition $P_{2\epsilon/c} \geq c$ and inequality (6) implies that

$$\epsilon \geq \left(c - \frac{1}{2}\sqrt{VT}\right) \cdot \frac{2\epsilon}{c},$$

which requires $T \geq c^2/V$.