Adversarial bandit optimization for approximately linear functions

Zhuoyu Cheng[1][0009−0002−6631−4929], Kohei Hatano[1,2][0000−0002−1536−1269], and Eiji Takimoto[1][0000−0001−9542−2553]

[1] Department of Informatics, Kyushu University, Japan
[2] RIKEN AIP, Japan
cheng.zhuoyu.874@s.kyushu-u.ac.jp
{hatano,eiji}@inf.kyushu-u.ac.jp

Abstract. We consider a bandit optimization problem for non-convex and non-smooth functions, where in each trial the loss function is the sum of a linear function and a small but arbitrary perturbation chosen after observing the player's choice. We give both expected and high probability regret bounds for the problem. Our result also implies an improved high-probability regret bound for the bandit linear optimization, a special case with no perturbation. We also give a lower bound on the expected regret.

Keywords: Bandit linear optimization · Second keyword · Another keyword.

1 Introduction

Bandit optimization is a sequential game between a player and an adversary. The game is played over T rounds, where T is a positive natural number called the horizon. The game is specified by a pair $(\mathcal{K}, \mathcal{F})$, where $\mathcal{K} \subseteq \mathbb{R}^d$ is a bounded closed convex set and $\mathcal{F} \subseteq \{f : \mathcal{K} \to \mathbb{R}\}$ is a function class. In each round $t \in [T]$, the player first chooses an action $x_t \in \mathcal{K}$ and the adversary chooses a loss function $f_t \in \mathcal{F}$, and then the player receives the value $f_t(x_t)$ as the loss. Note that f_t itself is unknown to the player. In this paper, we assume the adversary is oblivious, i.e., the loss functions are specified before starting the game [3]. The goal of the player is to minimize the regret

$$\sum_{t=1}^{T} f_t(x_t) - \min_{x \in \mathcal{K}} \sum_{t=1}^{T} f_t(x) \tag{1}$$

in expectation (expected regret) or with high probability (high-probability regret).

For convex loss functions, the bandit optimization has been extensively studied (see, e.g.,[2,9,15]). $\mathcal{O}(d^{1/3}T^{3/4})$ regret bounds are shown by Flaxman et

---

[3] We do not consider the case where the adversary is adaptive, i.e., it can choose the t-th loss function f_t depending on the previous actions $x_1, \ldots, x_{t-1}$.

al.[10]. Lattimore shows an information-theoretic regret bound $\widetilde{O}(d^{2.5}\sqrt{T})$ for convex loss functions[14]. For linear loss functions, Abernethy et al. propose the SCIRBLE algorithm and give an expected regret bound $\mathcal{O}(d\sqrt{T \ln T})$[2], achieving optimal dependence on T[8]. Lee et al. propose the SCRIBLE with lifting and show a high-probability regret bound $\widetilde{O}(d^2\sqrt{T})$[15].

Recently, non-convex functions are also getting popular in this literature. For example, Agarwal et al. show a regret bound $\mathcal{O}(poly(d)T^{2/3})$ for smooth and bounded non-convex functions[4]. Ghai et al. propose algorithms with regret bounds $\mathcal{O}(poly(d)T^{2/3})$ under the assumption that non-convex functions are reparametlized as some convex functions[12].

In this paper, we investigate the bandit optimization problem for a class of non-convex non-smooth loss functions. The function class consists of non-smooth and non-convex functions that are "close" to linear functions, in the sense that functions in the class can be viewed as linear functions with adversarial non-convex perturbations whose amount is up to ϵ. Bandit optimization for linear loss functions with stochastic noise (e.g., [1,5]) cannot be applied to our problem. Also, standard Bandit linear optimization methods for estimating the gradient, such as self-concordant barrier regularizer[13], cannot be effectively applied to our problem. We propose a novel approach to analyzing high-probability regret, introducing a new method for decomposing regret. Additionally, we propose a novel method to bound $\|x_t - u\|$, where $u \in \mathcal{K}$, to account for the impact of perturbations.

1. When $\epsilon \neq 0$, we propose a modification of the SCRIBLE with lifting and increasing learning rates [15] and utilize the properties of the ν-normal barrier[17] to prove its high probability regret bound $\widetilde{O}(d\sqrt{T} + \epsilon dT)$, and we also obtain its expected regret $\mathcal{O}(d\sqrt{T \ln T} + \epsilon dT)$.
2. When $\epsilon = 0$, this problem becomes Bandit linear optimization, a special case with no perturbation. Compared to Lee et al.'s results[15], holding with probability $1 - \gamma$, $\mathcal{O}(\ln^2(dT)d^2 \ln T \sqrt{T \ln \frac{\ln(dT)}{\gamma}})$, we use a different regret decomposition approach to achieve a better high-probability regret bound $\mathcal{O}(d\sqrt{T \ln T} + \ln T \sqrt{T \ln(\frac{\ln T}{\gamma})} + \ln(\frac{\ln T}{\gamma}))$.
3. We prove a lower bound $\Omega(\epsilon T)$, implying that our bounds are tight w.r.t. the parameter ϵ.

2   Related Work

The bandit linear optimization was first proposed by Awerbuch & Kleinberg [6], who achieved a regret bound of $\mathcal{O}(d^{3/5}T^{2/3})$ against an oblivious adversary. Later, McMahan & Blum established a regret bound of $\mathcal{O}(dT^{3/4})$ when facing an adaptive adversary[16]. A foundational approach in bandit optimization problems involves gradient-based smoothing techniques. Abernethy et al. presented pioneering work in this area and achieved an expected regret bound of $\mathcal{O}(d\sqrt{T \ln T})$ when dealing with an oblivious adversary[3]. Bartlett et al.

proposed a high-probability regret bound of $\mathcal{O}(d^{2/3}\sqrt{T\ln dT})$ under a special condition[7]. Subsequently, Lee et al. presented a high-probability regret bound $\widetilde{O}(d^2\sqrt{T})$ for both oblivious and adaptive adversaries[15].

Unlike convex bandit problems, which have been extensively explored and analyzed, non-convex bandits introduce unique challenges due to the complexity of exploring and exploiting in a non-convex area. Gao et al. considered both non-convex losses and non-stationary data and established a regret bound of $O(\sqrt{T + poly(T)})$[11]. Yang et al. achieved a regret bound of $O(\sqrt{T\log T})$ for non-convex loss functions[19]. However, they both required the loss functions to have smoothness properties, and our loss functions are neither convex nor smooth.

2.1   Comparison to Lee et al.[15]

Our approach builds upon Lee et al.'s work. Below, we highlight the key differences between our method and Lee et al.'s in the context of the oblivious bandit setting:

1. Simplified Regret Analysis: While Lee et al.'s regret analysis introduces unnecessary complexity for the oblivious bandit setting, our approach simplifies the analysis, leading to more streamlined results.
2. Reduced Dependence on d: Lee et al.'s analysis results in a regret bound with greater dependence on d, whereas our method derives a bound with significantly reduced dependence on d (This distinction is demonstrated in the introduction and further illustrated in the case where $\epsilon = 0$).
3. Revised Generality of Problem Setting: Like the SCIRBLE algorithm, our approach is more general, treating bandit linear optimization as a special case within a broader problem framework.

3   Preliminaries

This section introduces some necessary notations and defines ϵ-approximately linear function. Then we give our problem setting.

3.1   Notation

We abbreviate the 2-norm $\|\cdot\|_2$ as $\|\cdot\|$. For a twice differentiable convex function $\mathcal{R} : \mathbb{R}^d \to \mathbb{R}$ and any $x, h \in \mathbb{R}^d$, let $\|h\|_x = \|h\|_{\nabla^2\mathcal{R}(x)} = \sqrt{h^\top\nabla^2\mathcal{R}(x)h}$, and $\|h\|_x^* = \|h\|_{(\nabla^2\mathcal{R}(x))^{-1}} = \sqrt{h^\top(\nabla^2\mathcal{R}(x))^{-1}h}$, respectively.

For any $v \in \mathbb{R}^d$, let $v^\perp$ be the space orthogonal to v. Let $\mathbb{S}_1^d = \{x \mid \|x\| = 1\}$. The vector $e_i \in \mathbb{R}^d$ is a standard basis vector with a value of 1 in the i-th position and 0 in all other positions. I is an identity matrix with dimensionality implied by context.

3.2   Problem Setting

Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a bounded and closed convex set such that for any $x, y \in \mathcal{K}$, $\|x - y\| \le D$. Furthermore, we assume that $\mathcal{K}$ contains the unit ball centered at the zero vector. Otherwise, we can apply an affine transformation to translate the center point of the convex set to the origin. Let $\mathcal{K}' = \{(x, 1) : x \in \mathcal{K}\}$. For any $\delta \in (0, 1)$, let $\mathcal{K}_\delta = \{x | \frac{1}{1-\delta} x \in \mathcal{K}\}$ and $\mathcal{K}'_\delta = \{(x, 1) : x \in \mathcal{K}_\delta\}$, respectively.

Definition 1. *A function $f : \mathcal{K} \to \mathbb{R}$ is ϵ-approximately linear if there exists $\theta_f \in \mathbb{R}^d$ such that $\forall x \in \mathcal{K}$, $|f(x) - \theta_f^\top x| \le \epsilon$.*

For convenience, in the definition above, let $\sigma_f(x) = f(x) - \theta_f^\top x$, and we omit the subscript f of θ_f and σ_f if the context is clear. Note that $|\sigma(x)| \le \epsilon$ for any $x \in \mathcal{K}$.

In this paper, we consider the bandit optimization $(\mathcal{K}, \mathcal{F})$, where $\mathcal{F}$ is the set of ϵ-approximately linear functions $f(x) = \theta^\top x + \sigma(x)$ with $\|\theta\| \le G$. The bandit optimization for ϵ-approximately linear functions can be defined as the following statement. For every round $t = 1, .., T$, the player chooses an action $x_t \in \mathcal{K}$, and the adversary simultaneously chooses a linear function $f_t(x_t)(= \theta_t^\top x_t)$ at the same time. After observing the player's choice x_t, the adversary chooses a perturbation $\sigma_t(x_t)(|\sigma_t(x_t)| \le \epsilon)$. The value of ϵ-approximately linear function $f_t(x_t) = \theta_t^\top x_t + \sigma_t(x_t)$ is then revealed to the player. The goal of the player is to minimize the regret $\sum_{t=1}^T f_t(x_t) - \min_{x \in \mathcal{K}} \sum_{t=1}^T f_t(x)$.

4   Main Results

In this section, we first review SCRIBLE with lifting and increasing learning rates, followed by presenting the main contributions of this paper with detailed explanations.

4.1   SCRIBLE with lifting and increasing learning rates[15]

SCRIBLE with lifting and increasing learning rates introduces a dummy coordinate with a value of 1 appended to all actions, resulting in the lifted decision set $\mathcal{K}' = \{(x, 1) : x \in \mathcal{K}\}$. This transformation lifts the bandit linear optimization problem to $\mathbb{R}^{d+1}$. The conic hull of this set is $con(\mathcal{K}) = \{\mathbf{0}\} \cup \{(x, b) : \frac{x}{b} \in \mathcal{K}, x \in \mathbb{R}^d, b > 0\}$.

The algorithm performs SCRIBLE over the lifted decision set, using a ν-normal barrier $\mathcal{R}$ defined over the $con(\mathcal{K})$ (which always exists) as the regularizer to generate the sequence $x'_1, ..., x'_t$. It set $y'_t = x'_t + \mathbf{A}_t \mu_t$ where $\mathbf{A}_t = [\nabla^2 \mathcal{R}(x'_t)]^{-\frac{1}{2}}$ and μ_t is uniformly sampled at random from $\mathbb{S}_1^{d+1} \cap (\mathbf{A}_t e_{d+1})^\perp$. Since μ_t is orthogonal to $\mathbf{A}_t e_{d+1}$, the last coordinate of $\mathbf{A}_t \mu_t$ is zero, ensuring that $y'_t = (y_t, 1)$ remains within $\mathcal{K}'$. The actual point played is still y_t. After playing y_t and observing $f_t(= \theta_t^\top y_t)$, it constructs the loss estimator the same way as SCRIBLE: $g_t = df_t(y_t)\mathbf{A}_t^{-1} \mu_t$. The analysis by Lee et al. shows that the first d coordinates of g_t are indeed an unbiased estimator of θ_t[15].

SCRIBLE with lifting and increasing learning rates[15] utilizes the properties of the ν-normal barrier $\mathcal{R}$, which are not available in the κ-self concordant barrier $\mathcal{T}$, as well as increasing learning rates, to bound $\|h\|_{\nabla^2\mathcal{R}(x_t')} \leq -\|h\|_{\nabla^2\mathcal{R}(x_{t+1}')} + \nu\ln(\nu T + 1)$, where $h \in \mathcal{K}'$. Besides, Lee et al. divide the regret mainly into two parts: $\sum_{t=1}^{T}[y_t^\top\theta_t - x_t^\top g_t + u^\top(g_t - \theta_t)]$ and $\sum_{t=1}^{T}(y_t - u)^\top g_t$, where $u \in \mathcal{K}$. They provide a bound for $\sum_{t=1}^{T}(y_t - u)^\top g_t$ and a high-probability bound for $\sum_{t=1}^{T}[y_t^\top\theta_t - x_t^\top g_t + u^\top(g_t - \theta_t)]$, which helps derive high-probability regret bounds for both oblivious and adaptive adversaries. However, for our problem, the bound they obtain for $\|h\|_{\nabla^2\mathcal{R}(x_t')}$ is still too large and does not help in bounding $\|x_t' - h\|_{\nabla^2\mathcal{R}(x_t')}$. Again, since the largest eigenvalue of $\nabla^2\mathcal{R}(x_{t'})$[17] could potentially approach infinity, bounding $\|x_t' - h\|_{\nabla^2\mathcal{R}(x_t')}$ remains a challenge, and no previous work has addressed this.

4.2   SCRIBLE with lifting

For the decision set $\mathcal{K}$ with a ν-normal barrier on $con(\mathcal{K})$, where $con(\mathcal{K}) = \{\mathbf{0}\} \cup \{(x, b) : \frac{x}{b} \in \mathcal{K}, x \in \mathbb{R}^d, b > 0\}$, we apply Algorithm 1 to approximately linear functions. Recall $\mathcal{K}' = \{(x, 1) : x \in \mathcal{K}\}$.

We simplify the SCRIBLE with lifting and increasing learning rates[15]. We do not use the increasing learning rates part but retain the lifting. This preserves its advantages; for instance, the ν-normal barrier $\mathcal{R}$ always exists on $con(\mathcal{K})$. Additionally, the actual point played by the algorithm is y_t and $y_t' = (y_t, 1) = x_t' + \mathbf{A}_t\mu_t$ always remains within $\mathcal{K}'$. Furthermore, it adopts the same update method as FTRL algorithm[13]: $x_{t+1}' = \arg\min_{x'\in\mathcal{K}'} \eta\sum_{\tau=1}^{t} g_\tau^\top x' + \mathcal{R}(x')$.

Although it constructs the same loss estimator, $g_t = df_t(y_t)\mathbf{A}_t^{-1}\mu_t$ as the original algorithm, g_t is no longer an unbiased estimator of θ_t

---

Algorithm 1 SCRIBLE with lifting

Input: T, parameters $\eta \in \mathbb{R}, \delta \in (0, 1)$, ν-normal barrier $\mathcal{R}$ on $con(\mathcal{K})$
 1: Initialize: $x_1' = \arg\min_{x'\in\mathcal{K}'} \mathcal{R}(x')$
 2: **for** $t = 1, .., T$ **do**
 3:     let $\mathbf{A}_t = [\nabla^2\mathcal{R}(x_t')]^{-\frac{1}{2}}$
 4:     Draw μ_t from $\mathbb{S}_1^{d+1} \cap (\mathbf{A}_t e_{d+1})^\perp$ uniformly, set $y_t' = (y_t, 1) = x_t' + \mathbf{A}_t\mu_t$.
 5:     Play y_t, observe and incur loss $f_t(y_t)$. Let $g_t = df_t(y_t)\mathbf{A}_t^{-1}\mu_t$.
 6:     Update $x_{t+1}' = \arg\min_{x'\in\mathcal{K}'} \eta\sum_{\tau=1}^{t} g_\tau^\top x' + \mathcal{R}(x')$
 7: **end for**

We present our main results: expected and high-probability regret bounds for the problem.

Theorem 1. *The algorithm with parameters* $\eta = \frac{\sqrt{2\nu \log T}}{2d\sqrt{T}}, \delta = \frac{1}{T^2}$ *guarantees the following expected regret bound*

$$\mathbb{E}[\sum_{t=1}^{T} f_t(y_t) - \min_{x \in \mathcal{K}} \sum_{t=1}^{T} f_t(x)] \le 4d\sqrt{2\nu T \log T} + \frac{GD}{T} + 2T\epsilon + dT\epsilon(2\nu + \sqrt{\nu}). \quad (2)$$

Theorem 2. *The algorithm with parameters* $\eta = \frac{\sqrt{2\nu \ln T}}{2d\sqrt{T}}, \delta = \frac{1}{T^2}$ *ensures that with probability at least* $1 - \gamma$

$$\sum_{t=1}^{T} f_t(y_t) - \min_{x \in \mathcal{K}} \sum_{t=1}^{T} f_t(x) \le 4d\sqrt{2\nu T \ln T} + \frac{GD}{T} + Td\epsilon(2\nu + \sqrt{\nu}) + C(1+\epsilon)\sqrt{8T \ln \frac{C}{\gamma}} + 2GD \ln \frac{C}{\gamma} + 2T\epsilon$$

$$(3)$$

where $C = \lceil \ln GD \rceil \lceil \ln((GD)^2 T) \rceil$.

We primarily divide the regret into parts $\sum_{t=1}^{T}(x_t - u)^\top \theta_t$ and $\sum_{t=1}^{T}(y_t - x_t)^\top \theta_t$ rather than following the approach of Lee et al.[15], where $u \in \mathcal{K}$. This means that, for an oblivious adversary, calculating the regret does not require considering the variance of the estimator g_t and θ_t, but only the variance between y_t and x_t. This difference is a key factor that enables us to achieve a better high-probability regret bound when $\epsilon = 0$.

Firstly, we bound $\sum_{t=1}^{T}(x_t - u)^\top \theta_t$. As mentioned earlier, for ϵ-approximately linear functions, bounding $\sum_{t=1}^{T}(x_t - u)^\top \theta_t$ requires considering the term $(d\sigma_t(y_t)\mathbf{A}_t^{-1}\mu_t)^\top (x_t - u)$. The main difficulty in bounding $(d\sigma_t(y_t)\mathbf{A}_t^{-1}\mu_t)^\top (x_t - u)$ is that the norm $\|x_t - u\|$ is hard to bound. Thus, we make the entire analysis hold in $\mathbb{R}^{d+1}$ and transform the problem from bounding the norm $\|x_t - u\|$ to how to bound the norm $\|x_t' - h\|$, where $h \in \mathcal{K}'$. We present a straightforward yet necessary Lemma 4, which helps to bound $\|h\|_{\nabla^2 \mathcal{R}(x_t')} \le 2\nu$. In addition, the properties of the ν normal barrier tell us $\|x_t'\|_{\nabla^2 \mathcal{R}(x_t')} = \sqrt{\nu}$. With these two conditions, we can immediately deduce the bound $\|x_t - h\|_{\nabla^2 \mathcal{R}(x_t')}$ as $2\nu + \sqrt{\nu}$. This also implies that increasing learning rates are not required in our case, as they are solely aimed at controlling $\|h\|_{\nabla^2 \mathcal{R}(x_t')} \le -\|h\|_{\nabla^2 \mathcal{R}(x_{t+1}')} + \nu \ln(\nu T + 1)$ in Lee et al.'s paper[15].

Secondly, by obtaining the expected bound and high-probability bound for $\sum_{t=1}^{T}(y_t - x_t)^\top \theta_t$, we can derive the expected regret bound and high-probability regret bound, respectively. For high-probability bound of $\sum_{t=1}^{T}(y_t - x_t)^\top \theta_t$, unlike SCRIBLE with lifting and increasing learning rates, which constrains the decision set from $\mathcal{K}'$ to $\mathcal{K}_\delta'$ to ensure that x_t' is never too close to the boundary (thus ensuring that the eigenvalues of $\mathbf{A}_t$ are bounded, especially for bounding $\|h\|_{\nabla^2 \mathcal{R}(x_t')}$). Our approach does not require x_t' to stay away from the boundary. Furthermore, we do not need to bound the eigenvalues of $\mathbf{A}_t$, which gives us greater flexibility in choosing the value of δ (such as $\frac{1}{T^2}$), leading to a better upper bound for the regret.

Finally, we prove the lower bound of regret in section 6.

5 Proof

This section introduces the preliminary of ν-normal barrier, presents several essential lemmas, and gives the proof of main theorems.

5.1 ν-normal barrier

We introduce the ν-normal barrier, providing its definitions and highlighting several key properties that will be frequently used in the subsequent analysis.

Definition 2. *Let $\Psi \subseteq \mathbb{R}^d$ be a closed and proper convex cone and let $\nu \geq 1$. A function $\mathcal{R} : int(\Psi) \to \mathbb{R}$: is called a ν-logarithmically homogeneous self-concordant barrier (or simply ν-normal barrier) on Ψ if*

1. *$\mathcal{R}$ is three times continuously differentiable and convex and approaches infinity along any sequence of points approaching the boundary of Ψ.*
2. *For every $h \in \mathbb{R}^d$ and $x \in int(\Psi)$ the following holds:*

$$\sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{d} \frac{\partial^3 \mathcal{R}(x)}{\partial x_i \partial x_j \partial x_k} h_i h_j h_k \leq 2\|h\|_x^3, \tag{4}$$

$$\|\nabla \mathcal{R}(x)^\top h\| \leq \sqrt{\nu} \|h\|_x, \tag{5}$$

$$\mathcal{R}(tx) = \mathcal{R}(x) - \nu \ln t, \forall x \in int(\Psi), t > 0. \tag{6}$$

Lemma 1 ([17,18]). *If $\mathcal{R}$ is a ν-normal barrier on Ψ, Then for any $x \in int(\Psi)$ and any $h \in \Psi$, we have*

$$\|x\|_x^2 = \nu, \tag{7}$$

$$\nabla^2 \mathcal{R}(x)x = -\nabla \mathcal{R}(x), \tag{8}$$

$$\|h\|_x \leq -\nabla \mathcal{R}(x)^\top h, \tag{9}$$

$$\nabla \mathcal{R}(x)^\top (h - x) \leq \nu. \tag{10}$$

Lemma 2 ([17]). *If $\mathcal{R}$ is a ν-normal barrier on Ψ, then the Dikin ellipsoid centered at $x \in int(\Psi)$, defined as $\{y : \|y - x\|_x \leq 1\}$, is always within Ψ. Moreover,*

$$\|h\|_y \geq \|h\|_x (1 - \|y - x\|_x) \tag{11}$$

holds for any $h \in \mathbb{R}^d$ and any y with $\|y - x\|_x \leq 1$.

Lemma 3 ([13]). *Let $\mathcal{R}$ is a ν-normal barrier over Ψ, then for all $x, z \in int(\Psi) : \mathcal{R}(z) - \mathcal{R}(x) \leq \nu \log \frac{1}{1 - \pi_x(z)}$, where $\pi_x(z) = \inf\{t \geq 0 : x + t^{-1}(z - x) \in \Psi\}$.*

ν-normal barrier plays a crucial role in addressing one of the key challenges in this problem: bounding $\|x_t' - h\|_{x_t'}$, where $h \in \mathcal{K}'$. Equation 7 give $\|x_t'\|_{x_t'} = \sqrt{\nu}$. Building upon Lemma 1, we introduce an effective Lemma 4 that aids in bounding $\|h\|_{x_t'}$.

Lemma 4. *If $\mathcal{R}$ be a ν-normal barrier for $\Psi \subseteq R^d$, then for any $x \in int(\Psi)$ and any $h \in \Psi$, we have*

$$\|h\|_x \le 2\nu. \tag{12}$$

Proof. From Lemma 1, we have

$$\|h\|_x \le -\nabla\mathcal{R}(x)^\top h \le \|\nabla\mathcal{R}(x)^\top h|. \tag{13}$$

Then,

$$\|\nabla\mathcal{R}(x)^\top h| = \|\nabla\mathcal{R}(x)^\top (h - x + x)| \le \|\nabla\mathcal{R}(x)^\top (h - x)| + \|\nabla\mathcal{R}(x)^\top x|. \tag{14}$$

By Lemma 1, $\|\nabla\mathcal{R}(x)^\top (h - x)| + \|\nabla\mathcal{R}(x)^\top x| \le \nu + \|x^\top \nabla^2 \mathcal{R}(x)x| = 2\nu$.

With the help of Equation 7 and Lemma 4, it is easy to apply the triangle inequality to derive $\|x_t' - h\|_{x_t'} \le \|x_t'\|_{x_t'} + \|h\|_{x_t'} \le \sqrt{\nu} + 2\nu$, where $h \in \mathcal{K}'$..

5.2   Useful lemmas

In addition to the properties of the normal barrier and its related lemmas, we also need to introduce some additional necessary lemmas.

Like Lemma 6 in the SCRIBLE algorithm[2], the next minimizer x_{t+1}' is "close" to x_t'. However, there are two differences here: the first is that $\nabla\phi_{t-1}(x_t') \ne 0$ is possible, where $\phi_t(x') = \eta \sum_{\tau=1}^{t} g_\tau^\top x' + \mathcal{R}(x')$. And the second is that for $z = x_t' + \alpha u$, where u is a vector such that $\|u\|_{x_t'} = 1$ and $\alpha \in (-\frac{1}{2}, \frac{1}{2})$, we need to satisfy $z \in \mathcal{K}'$ instead of $z \in \mathcal{K}$.

Lemma 5. $x_{t+1}' \in W_{4d\eta}(x_t')$, *where* $W_r(x') = \{y \in \mathcal{K}' : \|y - x'\|_{x'} < r\}$.

Proof. Recall that $x_{t+1}' = \arg\min_{x' \in \mathcal{K}'} \phi_t(x')$, where $\phi_t(x') = \eta \sum_{\tau=1}^{t} g_\tau^\top x' + \mathcal{R}(x')$. Let $h_t(x) = \phi_t((x, 1)) = \phi_t(x')$, then $\min h_t(x) = \min \phi_t(x')$. Noticing that h_t is a convex function on $\mathbb{R}^d$ and still holds the barrier property(approaches infinity along any sequence of points approaching the boundary of $\mathcal{K}$). By properties of convex functions, we can get $\nabla h_{t-1}(x_t) = 0$ and for the first d coordinates $\nabla\phi_{t-1}(x_t') = 0$.

Consider any point in $z \in W_{\frac{1}{2}}(x_t')$. It can be written as $z = x_t' + \alpha u$ for some vector u such that $\|u\|_{x_t'} = 1$ and $\alpha \in (-\frac{1}{2}, \frac{1}{2})$. Noticing the $d + 1$ coordinate of u is 0. Expanding,

$$
\begin{aligned}
\phi_t(z) &= \phi_t(x_t' + \alpha u) \\
&= \phi_t(x_t') + \alpha \nabla\phi_t(x_t')^\top u + \alpha^2 \frac{1}{2} u^\top \nabla^2 \phi_t(\xi)u \\
&= \phi_t(x_t') + \alpha(\nabla\phi_{t-1}(x_t') + \eta g_t)^\top u + \alpha^2 \frac{1}{2} u^\top \nabla^2 \phi_t(\xi)u \\
&= \phi_t(x_t') + \alpha\eta g_t^\top u + \alpha^2 \frac{1}{2} u^\top \nabla^2 \phi_t(\xi)u,
\end{aligned}
$$

for some ξ on the path between x'_t and $x'_t + \alpha u$ and the last equality holds because $\nabla \phi_{t-1}(x'_t)^\top u = 0$. Setting the derivative with respect to α to zero, we obtain

$$\|\alpha^*\| = \frac{\eta \|g_t^\top u\|}{u^\top \nabla^2 \phi_t(\xi) u} = \frac{\eta \|g_t^\top u\|}{u^\top \nabla^2 \mathcal{R}(\xi) u} \tag{15}$$

The fact that ξ is on the line x'_t to $x'_t + \alpha u$ implies that $\|\xi - x'_t\|_{x'_t} \le \|\alpha u\|_{x'_t} \le \frac{1}{2}$. Hence, by Lemma 2

$$\nabla^2 \mathcal{R}(\xi) \succeq (1 - \|\xi - x'_t\|_{x'_t})^2 \nabla^2 \mathcal{R}(x'_t) \succ \frac{1}{4} \nabla^2 \mathcal{R}(x'_t). \tag{16}$$

Thus $u^\top \nabla^2 \mathcal{R}(\xi) u > \frac{1}{4} \|u\|_{x'_t} = \frac{1}{4}$, and $\alpha^* < 4\eta \|g_t^\top u\|$. Using assumption $\max_{x \in \mathcal{K}} \|f_t(x)\| \le 1$,

$$g_t^\top u \le \|g_t\|_{x'_t}^* \|u\|_{x'_t} \le \|df_t(y_t) \mathbf{A}_t^{-1} \mu_t\|_{x'_t}^* \le \sqrt{d^2 \mu_t^\top \mathbf{A}_t^{-\top} (\nabla^2 \mathcal{R}(x'_t))^{-1} \mathbf{A}_t^{-1} \mu_t} \le d, \tag{17}$$

we conclude that $\|g_t^\top u\| \le d$, and $\|\alpha^*\| < 4d\eta < \frac{1}{2}$ by our choice of η and T. We conclude that the local optimum $\arg\min z \in W_{\frac{1}{2}(x'_t)} \phi_t(z)$ is strictly inside $W_{4d\eta}(x'_t)$, and since ϕ_t is convex, the global optimum is

$$x_{t+1} = \arg\min_{z \in \mathcal{K}'} \phi_t(z) \in W_{4d\eta}(x'_t). \tag{18}$$

Lemma 5 implies $\|x'_{t+1} - x'_t\|_{x'_t} \le 4d\eta$. This result will help us bound $g_t^\top(x'_t - h)$, where $h \in \mathcal{K}'$(see Lemma 7).

This next lemma is based on Lemma B.9.[15], but due to the differences in the loss functions, what we obtain is an unbiased estimate regarding $g_{t,i}$ rather than $\theta_{t,i}$, for $i \in [d]$. Lee et al. state that $\mathbb{E}_t[l_{t,i}] = \theta_{t,i}$, for $i \in [d]$[15]. Since $l_t = d(\theta_t, 0)(x'_t + \mathbf{A}_t \mu_t) \mathbf{A}_t^{-1} \mu_t$ is identical to ours, we directly apply it to our analyze.

Lemma 6. *Let $l_t = d(\theta_t, 0)(x'_t + \mathbf{A}_t \mu_t) \mathbf{A}_t^{-1} \mu_t$. For Algorithm 1, we have $\mathbb{E}_t[l_{t,i}] = \theta_{t,i}$, for $i \in [d]$.*

The regret bound of FTRL algorithm[13] states that for every $u \in \mathcal{K}$, $\sum_{t=1}^T \nabla_t^\top x_t - \sum_{t=1}^T \nabla_t^\top u \le \sum_{t=1}^T [\nabla_t^\top x_t - \nabla_t^\top x_{t+1}] + \frac{1}{\eta}[\mathcal{R}(u) - \mathcal{R}(x_1)]$, where ∇_t represents the gradient of the loss function f_t. In our adaptation, we replaced ∇_t with g_t and $\mathcal{K}$ with $\mathcal{K}'$. This modification does not fundamentally alter the original result. Since the update way $x'_{t+1} = \arg\min_{x' \in \mathcal{K}'} \eta \sum_{\tau=1}^t g_\tau^\top x' + \mathcal{R}(x')$ satisfied the condition of FTRL algorithm[13], we can apply Lemma 5.3. in [13] to Algorithm 1 as follow.

Lemma 7. *For Algorithm 1 and for every $h \in \mathcal{K}'$, $\sum_{t=1}^T g_t^\top x'_t - \sum_{t=1}^T g_t^\top h \le \sum_{t=1}^T [g_t^\top x'_t - g_t^\top x'_{t+1}] + \frac{1}{\eta}[\mathcal{R}(h) - \mathcal{R}(x'_1)]$.*

The following lemma represents a key proof of this paper. Specifically, it provides a bound for $\sum_{t=1}^T \theta_t^\top x_t - \sum_{t=1}^T \theta_t^\top x^*$. Due to Lemma 8, we only need to consider $\sum_{t=1}^T \theta_t^\top y_t - \sum_{t=1}^T \theta_t^\top x_t$ when calculating the regret bound. This result plays a crucial role in deriving both expected and high-probability regret bounds.

Lemma 8. *For Algorithm 1, let $f_t(x_t) = \theta_t^\top x_t + \sigma_t(x_t)$ and $x^* = \arg\min_{x \in \mathcal{K}} \sum_{t=1}^T f_t(x)$ and we have*

$$\sum_{t=1}^T \theta_t^\top x_t - \sum_{t=1}^T \theta_t^\top x^* \leq 2\eta d^2 T + \frac{\nu \log(\frac{1}{\delta})}{\eta} + T d\epsilon (2\nu + \sqrt{\nu}) + \delta DGT. \quad (19)$$

Proof. Recall for any $\delta \in (0,1)$, $\mathcal{K}_\delta = \{x | \frac{1}{1-\delta} x \in \mathcal{K}\}$ and $\mathcal{K}_\delta' = \{(x,1) : x \in \mathcal{K}_\delta\}$. Let $x_\delta^* = \prod_{\mathcal{K}_\delta} x^*$, by properties of projections, then

$$\|x^* - x_\delta^*\| = \min_{a \in \mathcal{K}_\delta} \|x^* - a\|. \quad (20)$$

Since $(1 - \delta)x^* \in \mathcal{K}_\delta$, then

$$\min_{a \in \mathcal{K}} \|x^* - a\| \leq \|x^* - (1 - \delta)x^*\| \leq \delta D. \quad (21)$$

So,

$$\|x_\delta^* - x^*\| \leq \delta D. \quad (22)$$

By Cauchy–Schwarz inequality and the fact that $\|\theta\| \leq G$ and $\|x_\delta^* - x^*\| \leq \delta D$,

$$\sum_{t=1}^T \theta_t^\top x_\delta^* - \sum_{t=1}^T \theta_t^\top x^* \leq \delta DGT. \quad (23)$$

So,

$$\sum_{t=1}^T \theta_t^\top x_t - \sum_{t=1}^T \theta_t^\top x^* = \sum_{t=1}^T \theta_t^\top x_t - \sum_{t=1}^T \theta_t^\top x_\delta^* + \sum_{t=1}^T \theta_t^\top x_\delta^* - \sum_{t=1}^T \theta_t^\top x^*$$

$$\leq \sum_{t=1}^T \theta_t^\top x_t - \sum_{t=1}^T \theta_t^\top x_\delta^* + \delta DGT.$$

Let $\theta_t' = (\theta_t, z)$, where z is the $(d+1)$th coordinate of $d\mathbb{E}_t[(\theta_t, 0)^\top (x_t' + \mathbf{A}_t \mu_t)\mathbf{A}_t^{-1}\mu_t]$. From Lemma 6, we know $d\mathbb{E}_t[(\theta_t, 0)^\top (x_t' + \mathbf{A}_t \mu_t)\mathbf{A}_t^{-1}\mu_t] = \theta_t'$. Since $g_t = df(y_t)\mathbf{A}_t^{-1}\mu_t = d\theta_t'^\top (x_t' + \mathbf{A}_t \mu_t)\mathbf{A}_t^{-1}\mu_t + d\sigma_t(y_t)\mathbf{A}_t^{-1}\mu_t$, and let $M_t = \mathbb{E}_t[d\sigma_t(y_t)\mathbf{A}_t^{-1}\mu_t]$, then $\theta_t' = \mathbb{E}_t[g_t] - M_t$ and that we have

$$\sum_{t=1}^T \theta_t^\top x_t - \sum_{t=1}^T \theta_t^\top x_\delta^* = \sum_{t=1}^T (\theta_t'^\top x_t' - z) - \sum_{t=1}^T (\theta_t'^\top x_\delta^{*'} - z)$$

$$= \sum_{t=1}^T (\mathbb{E}_t[g_t] - M_t)^\top x_t' - \sum_{t=1}^T (\mathbb{E}_t[g_t] - M_t)^\top x_\delta^{*'}$$

$$= \sum_{t=1}^T \mathbb{E}_t[g_t]^\top (x_t' - x_\delta^{*'}) + \sum_{t=1}^T M_t^\top (x_\delta^{*'} - x_t').$$

We bound $\sum_{t=1}^T M_t^\top (x_\delta^{*'} - x_t')$ firstly. By Cauchy–Schwarz inequality,

$$\sum_{t=1}^T M_t^\top (x_\delta^{*'} - x_t') \leq \sum_{t=1}^T \|M_t\|_{x_t'}^* \|x_\delta^{*'} - x_t'\|_{x_t'}.$$

By Jensen's inequality,

$$\|M_t\|_{x'_t}^* = \sqrt{M_t^\top \nabla^2(\mathcal{R}(x'_t))^{-1} M_t} \tag{24}$$

$$= \sqrt{\mathbb{E}_t[d\sigma_t(y_t)\mathbf{A}_t^{-1}\mu_t]^\top \nabla^2(\mathcal{R}(x'_t))^{-1} \mathbb{E}_t[d\sigma_t(y_t)\mathbf{A}_t^{-1}\mu_t]} \tag{25}$$

$$= \sqrt{d^2 \mathbb{E}_t[\sigma_t(y_t)\mu_t]^\top \mathbf{A}_t^{-1} \mathbf{A}_t^2 \mathbf{A}_t^{-1} \mathbb{E}_t[\sigma_t(y_t)\mu_t]} \tag{26}$$

$$= \sqrt{d^2 \mathbb{E}_t[\sigma_t(y_t)\mu_t]^\top \mathbb{E}_t[\sigma_t(y_t)\mu_t]} \tag{27}$$

$$\leq \sqrt{d^2 \mathbb{E}_t[\sigma_t^2(y_t)\mu_t^\top \mu_t]} \tag{28}$$

$$\leq \sqrt{d^2 \epsilon^2} \tag{29}$$

$$= d\epsilon. \tag{30}$$

Then we bound $\|x_\delta^{*'} - x'_t\|_{x'_t}$. From the triangle inequality,

$$\|x_\delta^{*'} - x'_t\|_{x'_t} \leq \|x_\delta^{*'}\|_{x'_t} + \|x'_t\|_{x'_t}.$$

By Lemma 1 and Lemma 4, we obtain $\|x_\delta^{*'}\|_{x'_t} \leq 2\nu$, $\|x'_t\|_{x'_t} = \sqrt{\nu}$.

So $\|x_\delta^{*'} - x'_t\|_{x'_t} \leq 2\nu + \sqrt{\nu}$ and $\sum_{t=1}^T M_t^\top (x_\delta^{*'} - x'_t) \leq Td\epsilon(2\nu + \sqrt{\nu})$. Then bound $\sum_{t=1}^T \mathbb{E}_t[g_t]^T (x'_t - x_\delta^{*'})$

By Lemma 7,

$$\sum_{t=1}^T \mathbb{E}_t[g_t]^T (x'_t - x_\delta^{*'}) = \mathbb{E}_t\{\sum_{t=1}^T g_t^\top (x'_t - x_\delta^{*'})\}$$

$$\leq \mathbb{E}_t\{\sum_{t=1}^T [g_t^\top x'_t - g_t^\top x'_{t+1}] + \frac{1}{\eta}[\mathcal{R}(x_\delta^{*'}) - \mathcal{R}(x'_1)]\}$$

$$\leq \mathbb{E}_t\{\sum_{t=1}^T [\|g_t\|_{x'_t}^* \|x'_t - x'_{t+1}\|_{x'_t}]\} + \frac{1}{\eta}(\mathcal{R}(x_\delta^{*'}) - \mathcal{R}(x'_1)).$$

Lemma 5 implies that $\|x'_t - x'_{t+1}\|_{x'_t} \leq 4d\eta$ is true by choice of η. Additionally, from Eq. (17), we deduce that $\|g_t\|_{x'_t}^* \leq d$. Therefore,

$$\|g_t\|_{x'_t}^* \|x'_t - x'_{t+1}\|_{x'_t} \leq 4\eta d^2, \tag{31}$$

$$\mathbb{E}_t\{\sum_{t=1}^T [\|g_t\|_{x'_t}^* \|x'_t - x'_{t+1}\|_{x'_t}]\} \leq 4\eta d^2 T. \tag{32}$$

With Lemma 3,

$$\frac{1}{\eta}(\mathcal{R}(x_\delta^{*'}) - \mathcal{R}(x'_1)) \leq \frac{\nu \log(\frac{1}{\delta})}{\eta}. \tag{33}$$

Combine everything, we get

$$\sum_{t=1}^T \theta_t^\top x_t - \sum_{t=1}^T \theta_t^\top x_\delta^* \leq 4\eta d^2 T + \frac{\nu \log(\frac{1}{\delta})}{\eta} + Td\epsilon(2\nu + \sqrt{\nu}). \tag{34}$$

Now we are ready to prove Theorem 1.

5.3 Proof of Theorem 1

Proof. Recall ϵ-approximately linear function can be write as: $f(x) = \theta^\top x + \sigma(x)$. Thus, the regret of SCRIBLE with lifting algorithm

$$\mathbb{E}[\sum_{t=1}^{T} f_t(y_t) - \sum_{t=1}^{T} f_t(x^*)] = \mathbb{E}[\sum_{t=1}^{T}[\theta_t^\top y_t + \sigma_t(y_t)] - \sum_{t=1}^{T}[\theta_t^\top x^* + \sigma_t(x^*)]]$$

$$= \mathbb{E}[\sum_{t=1}^{T} \theta_t^\top y_t - \sum_{t=1}^{T} \theta_t^\top x^*] + \mathbb{E}[\sum_{t=1}^{T} \sigma_t(y_t) - \sum_{t=1}^{T} \sigma_t(x^*)].$$

Firstly, we bound the front of the above equation,

$$\mathbb{E}[\sum_{t=1}^{T} \theta_t^\top y_t - \sum_{t=1}^{T} \theta_t^\top x^*] = \sum_{t=1}^{T} \mathbb{E}[\theta_t^\top y_t] - \sum_{t=1}^{T} \mathbb{E}[\theta_t^\top x_t] + \sum_{t=1}^{T} \mathbb{E}[\theta_t^\top x_t] - \sum_{t=1}^{T} \mathbb{E}[\theta_t^\top x^*].$$

From the Law of total expectation, we know

$$\sum_{t=1}^{T} \mathbb{E}[\theta_t^\top y_t] - \sum_{t=1}^{T} \mathbb{E}[\theta_t^\top x_t] = \sum_{t=1}^{T} \mathbb{E}[\theta_t^\top (y_t - x_t)]$$

$$= \sum_{t=1}^{T} \mathbb{E}[\mathbb{E}_t[\theta_t^\top (y_t - x_t)]]$$

$$= \sum_{t=1}^{T} \mathbb{E}[\mathbb{E}_t[\theta_t^\top (\mathbf{A}_t \mu_t)]]$$

$$= \sum_{t=1}^{T} \mathbb{E}[\theta_t^\top \mathbb{E}_t[(\mathbf{A}_t \mu_t)]]$$

$$= \sum_{t=1}^{T} \mathbb{E}[\theta_t^\top \mathbf{0}]$$

$$= \mathbf{0}.$$

Thus,

$$\mathbb{E}[\sum_{t=1}^{T} \theta_t^\top y_t - \sum_{t=1}^{T} \theta_t^\top x^*] = \mathbb{E}[\sum_{t=1}^{T} \theta_t^\top x_t - \sum_{t=1}^{T} \theta_t^\top x^*]. \tag{35}$$

From Lemma 8, we have

$$\mathbb{E}[\sum_{t=1}^{T} \theta_t^\top x_t - \sum_{t=1}^{T} \theta_t^\top x^*] \le \mathbb{E}[4\eta d^2 T + \frac{\nu \log(\frac{1}{\delta})}{\eta} + Td\epsilon(2\nu + \sqrt{\nu}) + \delta DGT]$$

$$\le 4\eta d^2 T + \frac{\nu \log(\frac{1}{\delta})}{\eta} + Td\epsilon(2\nu + \sqrt{\nu}) + \delta DGT.$$

Since σ_t is chosen after knowing the player's action, it can cause as large a perturbation as possible. We using $\|\sigma_t(x)\| \leq \epsilon$ to bound $\sum_{t=1}^{T} \mathbb{E}[\sigma_t(y_t) - \sum_{t=1}^{T} \sigma_t(x^*)] \leq 2T\epsilon$ and combination of everything, we get

$$
Regret = \mathbb{E}[\sum_{t=1}^{T} f_t(y_t) - \sum_{t=1}^{T} f_t(x^*)]
$$
$$
\leq 4d\sqrt{2\nu T \log T} + \frac{GD}{T} + Td\epsilon(2\nu + \sqrt{\nu}) + 2T\epsilon,
$$

where $\eta = \frac{\sqrt{2\nu \log T}}{2d\sqrt{T}}, \delta = \frac{1}{T^2}$.

5.4   Proof of Theorem 2

To establish the high-probability regret bound, we first introduce the necessary Lemma 9.

Lemma 9 (Theorem 2.2. in [15]). *Let $X_1, ..., X_T$ be a martingale difference sequence with respect to a filtration $F_1 \subseteq ... \subseteq F_T$ such that $\mathbb{E}[X_t \mid F_t] = 0$. Suppose $B_t \in [1, b]$ for a fixed constant b is F_t-measurable and such that $X_t \leq B_t$ holds almost surely. Then with probability at least $1 - \gamma$ we have*

$$
\sum_{t=1}^{T} X_t \leq C(\sqrt{8V ln(C/\gamma)} + 2B^* ln(C/\gamma)), \tag{36}
$$

where $V = \max\{1; \sum_{t=1}^{T} \mathbb{E}[X_t^2 \mid F_t]\}, B^* = \max_{t\in[T]} B_t,$ and $C = \lceil logb \rceil \lceil log(b^2 T) \rceil$.

The analysis in [15] employs Lemma 9 to derive a high-probability bound for $\sum_{t=1}^{T}[y_t^\top \theta_t - x_t^\top g_t + u^\top(g_t - \theta_t)]$. In contrast, our approach defines $X_t = \theta_t^\top y_t - \theta_t^\top x_t$ and derives the high-probability bound for $\sum_{t=1}^{T}(\theta_t^\top y_t - \theta_t^\top x_t)$. This distinction in the application of Lemma 9 enables us to derive a tighter high-probability upper bound for bandit linear optimization.

With the support of Lemma 8 and Lemma 9, we are ready to prove Theorem 2.

Proof. Let $X_t = \theta_t^\top y_t - \theta_t^\top x_t$, then $\mathbb{E}_t[X_t] = \mathbb{E}_t[\theta_t^\top y_t - \theta_t^\top x_t] = 0$, $X_t = \theta_t^\top y_t - \theta_t^\top x_t \leq \|\theta_t\|\|y_t - x_t\| \leq GD$ and

$$
\begin{aligned}
\mathbb{E}_t[X_t^2] &= \mathbb{E}_t[(\theta_t^\top y_t - \theta_t^\top x_t)^2] \\
&= \mathbb{E}_t[(\theta_t^\top y_t)^2 + (\theta_t^\top x_t)^2 - 2\theta_t^\top y_t \theta_t^\top x_t] \\
&= \mathbb{E}_t[(\theta_t^\top y_t)^2] + \mathbb{E}_t[(\theta_t^\top x_t)^2] - \mathbb{E}_t[2\theta_t^\top y_t \theta_t^\top x_t] \\
&= \mathbb{E}_t[(\theta_t^\top y_t)^2] - \theta_t^\top x_t \theta_t^\top x_t \\
&\leq (1 + \epsilon)^2.
\end{aligned}
$$

Then,

$$\sum_{t=1}^{T} f_t(y_t) - \sum_{t=1}^{T} f_t(x^*) = \sum_{t=1}^{T}[\theta_t^\top y_t + \sigma_t(y_t)] - \sum_{t=1}^{T}[\theta_t^\top x^* + \sigma_t(x^*)]$$

$$\leq \sum_{t=1}^{T} \theta_t^\top y_t - \sum_{t=1}^{T} \theta_t^\top x^* + \sum_{t=1}^{T} \sigma_t(y_t) - \sum_{t=1}^{T} \sigma_t(x^*)$$

$$\leq \sum_{t=1}^{T} \theta_t^\top y_t - \sum_{t=1}^{T} \theta_t^\top x^* + 2T\epsilon$$

$$= \sum_{t=1}^{T} \theta_t^\top y_t - \sum_{t=1}^{T} \theta_t^\top x_t + \sum_{t=1}^{T} \theta_t^\top x_t - \sum_{t=1}^{T} \theta_t^\top x^* + 2T\epsilon.$$

From Lemma 8, we know

$$\sum_{t=1}^{T} \theta_t^\top x_t - \sum_{t=1}^{T} \theta_t^\top x^* \leq 4d\sqrt{2\nu T \ln T} + \frac{GD}{T} + Td\epsilon(2\nu + \sqrt{\nu}), \qquad (37)$$

where $\eta = \frac{\sqrt{2\nu \ln T}}{2d\sqrt{T}}, \delta = \frac{1}{T^2}$. Then by Lemma 9,

$$\mathbb{P}(\sum_{t=1}^{T}(\theta_t^\top y_t - \theta_t^\top x_t) \leq C(\sqrt{8V \ln(C/\gamma)} + 2B^* \ln(C/\gamma))) \geq 1 - \gamma, \qquad (38)$$

where $V = (1+\epsilon)^2 T, B^* = b = GD$, and $C = \lceil \ln GD \rceil \lceil \ln((GD)^2 T) \rceil$. Combine everything to conclude the proof.

5.5   Application to black-box optimization

From online to offline transformation, the result of this paper can also apply to black-box optimization for ϵ-approximately linear function. This problem is important in that previous theoretical analyses for black box optimization can only deal with linear/convex/smooth objectives in the adversarial environments (via bandit convex optimization). So, it is quite meaningful to clarify the possibility of the black box optimization problems without such restrictions. In fact, our objective is not linear, nor smooth, even with a simple assumption.

Let $\hat{x}$ be the output of algorithm 1, then from Theorem 1. we can easily prove and ensure $f(\hat{x}) - \min_{x \in \mathcal{K}} f(x) \leq \frac{4d\sqrt{2\nu \ln T}}{\sqrt{T}} + \frac{GD}{T^2} + d\epsilon(2\nu + \sqrt{\nu}) + 2\epsilon$. Additionally, we provide a lower bound 2ϵ for this problem (see Lemma 10). We can see that the difference between the lower bound and the upper bound is only $d\epsilon(2\nu + \sqrt{\nu})$ as T approaches infinity. This suggests the potential existence of "easier" settings between the adversarial environment and the standard stochastic environment, where better algorithms might be found. It also motivates us to explore these settings further.

6   Lower bound

In this section, we show a lower bound of the regret. To do so, we consider a black-box optimization problem for the set $\mathcal{F}$ of ϵ-approximately linear functions $f : \mathcal{K} \to \mathbb{R}$. In the problem, we are given access to the oracle O_f for some $f \in \mathcal{F}$, which returns the value $f(x)$ given an input $x \in \mathcal{K}$. The goal is to find a point $\hat{x} \in \mathcal{K}$ such that $f(\hat{x}) - \min_{x \in \mathcal{K}} f(x)$ is small enough. Then, the following statement holds.

Lemma 10. *For any algorithm $\mathcal{A}$ for the black-box optimization problem for $\mathcal{F}$, there exists an ϵ-approximately linear function $f \in \mathcal{F}$ such that the output $\hat{x}$ of $\mathcal{A}$ satisfies*

$$f(\hat{x}) - \min_{x \in \mathcal{K}} f(x) \geq 2\epsilon. \tag{39}$$

Proof. Firstly, suppose that the algorithm $\mathcal{A}$ is deterministic. At iteration $t = 1, ..., T$, for any feedback $y_1, ..., y_{t-1} \in \mathbb{R}$, $\mathcal{A}$ should choose the next query point x_t based on the data observed so far. That is,

$$x_t = \mathcal{A}((x_1, y_1), ..., (x_{t-1}, y_{t-1})). \tag{40}$$

Assume that the final output $\hat{x}$ is returned after T queries to the oracle O_f. In particular, we fix the T feedbacks $y_1 = y_2 = \cdots = y_T = \epsilon$. Let $z \in \mathcal{K}$ be such that $z \notin \{x_1, ..., x_T, \hat{x}\}$. Then we define a function $f : \mathcal{K} \to \mathbb{R}$ is as

$$f(x) = \begin{cases} \epsilon, & x \neq z, \\ -\epsilon, & x = z. \end{cases} \tag{41}$$

The function f is indeed an ϵ-approximately linear function, as $f(x) = 0^\top x + \sigma(x)$, where $\sigma(x) = \epsilon$ for $x \neq z$ and $\sigma(x) = -\epsilon$ for $x = z$. Further, we have

$$f(\hat{x}) - \min_{x \in \mathcal{K}} f(x) \geq 2\epsilon. \tag{42}$$

Secondly, if algorithm $\mathcal{A}$ is randomized. It means each x_t is chosen randomly. We assume the same feedbacks $y_1 = y_2 = \cdots = y_T = \epsilon$. Let $X = \{x_1, ..., x_T, \hat{x}\}$. Then, there exists a point $z \in \mathcal{K}$ such that $P_X(z \in X) = 0$, since $\mathbb{E}_{z'}[P_X(z' \in X|z')] = P_{z',X}(z' \in X) = \mathbb{E}_X[P_{z'}(z' \in X|X)] = 0$, where the expectation on z' is defined w.r.t. the uniform distribution over $\mathcal{K}$. For the objective function f defined in (41), we have $f(\hat{x}) - \min_{x \in \mathcal{K}} f(x) \geq 2\epsilon$ while f is ϵ-approximately linear.

Theorem 3. *For any horizon $T \geq 1$ and any player, there exists an adversary such that the regret is at least $2\epsilon T$.*

Proof. We prove the statement by contradiction. Suppose that there exists a player whose regret is less than $2\epsilon T$. Then we can construct an algorithm for the blackbox optimization problem from it by feeding the online algorithm

with T feedbacks of the blackbox optimization problem and by setting $\hat{x} = \min_{t \in [T]} f(x_t)$. Then,

$$f(\hat{x}) - \min_{x \in \mathcal{K}} f(x) \leq \frac{\sum_{t=1}^{T} f(x_t) - \sum_{t=1}^{T} \min_{x \in \mathcal{K}} f(x)}{T} < 2\epsilon,$$

which contradicts Lemma 10.

This lower bound indicates that $\Omega(\epsilon T)$ regret is inevitable for the bandit optimization problem for ϵ-approximately linear functions. We conjecture that the lower bound can be tightened to $\Omega(d\epsilon T)$, but we leave it as an open problem.

References

1. Abbasi-Yadkori, Y., Pál, D., Szepesvári, C.: Improved algorithms for linear stochastic bandits. Advances in neural information processing systems **24** (2011)
2. Abernethy, J.D., Hazan, E., Rakhlin, A.: Competing in the dark: An efficient algorithm for bandit linear optimization. In: COLT. pp. 263–274 (2008)
3. Abernethy, J.D., Hazan, E., Rakhlin, A.: Interior-point methods for full-information and bandit online learning. IEEE Transactions on Information Theory **58**(7), 4164–4175 (2012)
4. Agarwal, N., Gonen, A., Hazan, E.: Learning in non-convex games with an optimization oracle. In: Conference on Learning Theory. pp. 18–29. PMLR (2019)
5. Amani, S., Alizadeh, M., Thrampoulidis, C.: Linear stochastic bandits under safety constraints. Advances in Neural Information Processing Systems **32** (2019)
6. Awerbuch, B., Kleinberg, R.D.: Adaptive routing with end-to-end feedback: Distributed learning and geometric approaches. In: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing. pp. 45–53 (2004)
7. Bartlett, P., Dani, V., Hayes, T., Kakade, S., Rakhlin, A., Tewari, A.: High-probability regret bounds for bandit online linear optimization. In: Proceedings of the 21st Annual Conference on Learning Theory-COLT 2008. pp. 335–342. Omnipress (2008)
8. Bubeck, S., Cesa-Bianchi, N., Kakade, S.M.: Towards minimax policies for online linear optimization with bandit feedback. In: Conference on Learning Theory. pp. 41–1. JMLR Workshop and Conference Proceedings (2012)
9. Dani, V., Kakade, S.M., Hayes, T.: The price of bandit information for online optimization. Advances in Neural Information Processing Systems **20** (2007)
10. Flaxman, A.D., Kalai, A.T., Mcmahan, H.B.: Online convex optimization in the bandit setting: gradient descent without a gradient. In: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms. pp. 385 – 394 (2005). https://doi.org/10.5555/1070432
11. Gao, X., Li, X., Zhang, S.: Online learning with non-convex losses and non-stationary regret. In: International Conference on Artificial Intelligence and Statistics. pp. 235–243. PMLR (2018)
12. Ghai, U., Lu, Z., Hazan, E.: Non-convex online learning via algorithmic equivalence. Advances in Neural Information Processing Systems **35**, 22161–22172 (2022)
13. Hazan, E., et al.: Introduction to online convex optimization. Foundations and Trends® in Optimization **2**(3-4), 157–325 (2016)

14. Lattimore, T.: Improved regret for zeroth-order adversarial bandit convex optimisation. Mathematical Statistics and Learning **2**, 311–334 (10 2020). `https://doi.org/10.4171/msl/17`
15. Lee, C.W., Luo, H., Wei, C.Y., Zhang, M.: Bias no more: high-probability data-dependent regret bounds for adversarial bandits and mdps. Advances in neural information processing systems **33**, 15522–15533 (2020)
16. McMahan, H.B., Blum, A.: Online geometric optimization in the bandit setting against an adaptive adversary. In: Learning Theory: 17th Annual Conference on Learning Theory, COLT 2004, Banff, Canada, July 1-4, 2004. Proceedings 17. pp. 109–123. Springer (2004)
17. Nemirovski, A.: Interior point polynomial time methods in convex programming. Lecture notes **42**(16), 3215–3224 (2004)
18. Nesterov, Y., Nemirovskii, A.: Interior-point polynomial algorithms in convex programming. SIAM (1994)
19. Yang, L., Deng, L., Hajiesmaili, M.H., Tan, C., Wong, W.S.: An optimal algorithm for online non-convex learning. Proceedings of the ACM on Measurement and Analysis of Computing Systems **2**(2), 1–25 (2018)