Adversarial bandit optimization for approximately linear functions

Anonymous authors
Paper under double-blind review

Abstract

We consider a variant of the standard Bandit linear optimization, 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.

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 [1]. 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.,Dani et al. (2007); Abernethy et al. (2008); Lee et al. (2020)). $\mathcal{O}(T^{3/4})$ regret bounds are shown by Flaxman et al. (2005). Lattimore (2020) shows an information-theoretic regret bound $\widetilde{O}(d^{2.5}\sqrt{T})$ for convex loss functions. For linear loss functions, Abernethy et al. (2008) propose the SCIRBLE algorithm and give an expected regret bound $\mathcal{O}(d\sqrt{T})$, achieving optimal dependence on T(Bubeck et al., 2012). Lee et al. (2020) propose the SCRIBLE with lifting and show a high-probability regret bound $\widetilde{O}(d^2\sqrt{T})$.

Recently, non-convex functions are also getting popular in this literature. For example, Agarwal et al. (2019) show a regret bound $\mathcal{O}(poly(d)T^{2/3})$ for smooth and bounded non-convex functions. Ghai et al. (2022) 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.

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., Abbasi-Yadkori et al. (2011); Amani et al. (2019)) cannot be applied to our problem. Also, standard Bandit linear optimization methods for estimating the gradient, such as self-concordant barrier regularizer(Hazan et al., 2016), cannot be effectively applied to our problem.

[1] 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 $x_1, \dots, x_{t-1}$.

1. When $\epsilon \neq 0$, we propose a modification of the SCRIBLE with lifting (Lee et al., 2020) and utilize the properties of the ν-normal barrier(Nesterov & Nemirovskii, 1994; Nemirovski, 2004) to prove its high probability regret bound $\widetilde{O}(\sqrt{T}+\epsilon T)$, and we also obtain its expected regret $\mathcal{O}(\sqrt{T\ln T}+\epsilon T)$.
2. When $\epsilon = 0$, this problem becomes Bandit linear optimization, a special case with no perturbation. Compared to Lee et al. (2020)'s results, 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 (2004), who achieved a regret bound of $\mathcal{O}(d^{3/5}T^{2/3})$ against an oblivious adversary. Later, McMahan & Blum (2004) established a regret bound of $\mathcal{O}(dT^{3/4})$ when facing an adaptive adversary. A foundational approach in bandit optimization problems involves gradient-based smoothing techniques. Hazan & Levy (2014) presented pioneering work in this area and achieved a regret bound of $\widetilde{O}(\sqrt{T})$.

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. (2018) considered both non-convex losses and non-stationary data and established a regret bound of $O(\sqrt{T+poly(T)}$. Yang et al. (2018) achieved a regret bound of $O(\sqrt{T\log T})$ for a non-convex loss functions. However, they both required the loss functions to have smoothness properties, and our loss functions are neither convex nor smooth.

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 $\|\theta\|\le G$.

4 Main Results

This section states our main contribution: the expected and high-probability regret of adversarial bandit optimization for approximately linear functions.

We simplify the SCRIBLE with lifting and increasing learning rates(Lee et al., 2020). We do not use the increasing learning rates part but retain the lifting. For a 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 $\mathcal{K}' = \{(x, 1) : x \in \mathcal{K}\}$.

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 $x'_{t+1} = \arg\min\limits_{x' \in \mathcal{K}'} \eta \sum_{\tau=1}^t g_\tau^\top x' + \mathcal{R}(x')$
7: **end for**

We use a different method to analyze regret. The first difference is that for ϵ-approximately linear functions, the traditional method that obtains an unbiased gradient estimate is impossible and the properties of self-concordant functions used by the SCRIBLE algorithm(Abernethy et al., 2008) do not help us bound the regret. That's why we employed the lifting so that we could leverage the properties of the normal barrier to bound the norm $\|h\|_{x'} \leq 2\nu$(see Lemma 8), where $x' \in int(\mathcal{K}', h \in con(\mathcal{K})$. The second difference is that we use a different regret decomposition approach to apply Lemma 7(Lee et al., 2020). This allows the high-probability regret bound we obtained to not only be effective in our problem setting but also achieve a better result for bandit linear optimization compared to previous outcomes. Besides, unlike SCRIBLE with lifting and increasing learning rates(Lee et al., 2020), which constrains the decision set from $\mathcal{K}'$ to $\mathcal{K}'_\delta$ to ensure that that x'_t is never too close to the boundary and in turn, that the eigenvalues of $\mathbf{A}_t$ are bounded. We are not concerned about x'_t being too close to the boundary as long as it doesn't reach it. Furthermore, we don't need to bound the eigenvalues of $\mathbf{A}_t$, allowing us to select a smaller value for δ, such as $\frac{1}{T^2}$, to obtain a better upper bound of regret. Finally, we prove the lower bound of regret in section 5.

To introduce and prove our results, we introduce the following definitions and lemmas first.

Definition 2. *Let $\Psi \in \mathbb{R}^d$ 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:*

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

$$|\nabla \mathcal{R}(x)^\top h| \leq \sqrt{\nu} \|h\|_x, \tag{3}$$

$$\mathcal{R}(tx) = \mathcal{R}(x) - \nu \ln t, \forall x \in int(\Psi), t > 0. \tag{4}$$

Lemma 1 (Nemirovski (2004); Nesterov & Nemirovskii (1994))**.** *If $\mathcal{R}$ is a ν-normal barrier on Ψ, Then for any $x \in int(\Psi)$ and any $h \in \Psi$, we have*

$$\|x\|_x^2 = \nu, \tag{5}$$

$$\nabla^2 \mathcal{R}(x) x = -\nabla \mathcal{R}(x), \tag{6}$$

$$\|h\|_x \leq -\nabla\mathcal{R}(x)^\top h, \tag{7}$$

$$\nabla\mathcal{R}(x)^\top (h-x) \leq \nu. \tag{8}$$

Lemma 2 (Nemirovski (2004))**.** *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,*

$$\|h\|_y \geq \|h\|_x (1 - \|y-x\|_x) \tag{9}$$

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

Like in the SCRIBLE algorithm, the next minimizer x'_{t+1} is "close" to x'_t. However, there are two differences here: the first is that $\nabla\phi_{t-1}(x'_t) \neq 0$ is possible, where $\phi_t(x') = \eta \sum_{\tau=1}^t g_\tau^\top x' + \mathcal{R}(x')$. And the second is we need to satisfy $z \in \mathcal{K}'$ instead of $z \in \mathcal{K}$.

Lemma 3. *$x'_{t+1} \in W_{4d\eta}(x'_t)$, 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

$$|\alpha^*| = \frac{\eta |g_t^\top u|}{u^\top \nabla^2 \phi_t(\xi) u} = |\alpha^*| = \frac{\eta |g_t^\top u|}{u^\top \nabla^2 \mathcal{R}(\xi) u} \tag{10}$$

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

$$\nabla^2\mathcal{R}(\xi) \succeq (1 - \|\xi - x'_t\|_{x'_t})^2 \nabla^2\mathcal{R}(x'_t) \succ \frac{1}{4}\nabla^2\mathcal{R}(x'_t). \tag{11}$$

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 $\max_{x\in\mathcal{K}} |f_t(x)| \leq 1$,

$$g_t^\top u \leq \|g_t\|^*_{x'_t} \|u\|_{x'_t} \leq \|df_t(y_t)\mathbf{A}_t^{-1}\mu_t\|^*_{x'_t} \leq \sqrt{d^2 \mu_t^\top \mathbf{A}_t^{-\top} (\nabla^2\mathcal{R}(x'_t))^{-1} \mathbf{A}_t^{-1} \mu_t} \leq d, \tag{12}$$

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

$$x_{t+1} = \arg\min_{z \in \mathcal{K}'} \phi_t(z) \in W_{4d\eta}(x'_t). \tag{13}$$

□

Lemma 4 (Hazan et al. (2016))**.** *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 $\pi_x(z) = \inf\{t \geq 0 : x + t^{-1}(z-x) \in \Psi\}$.*

This next lemma is based on Lemma B.9.(Lee et al., 2020), 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 $i \in [d]$.

Lemma 5. *Let $l = d(\theta_t, 0)(x'_t + \mathbf{A}_t\mu_t)\mathbf{A}_t^{-1}\mu_t$, for algorithm 1, we have $\mathbb{E}_t[l_{t,i}] = \theta_{t,i}$, for $i \in [d]$.*

Proof. Let $v = \frac{\mathbf{A}_t e_{d+1}}{\|\mathbf{A}_t e_{d+1}\|_2}$.First note that $\mathbb{E}_t[\mu_t\mu_t^\top] = \frac{1}{d}(I - vv^\top)$

$$\begin{aligned}
\mathbb{E}_t[d(\theta_t,0)(x'_t + \mathbf{A}_t\mu_t)\mathbf{A}_t^{-1}\mu_t] &= d\mathbb{E}_t[(\theta_t,0)x'_t\mathbf{A}_t^{-1}\mu_t] + d\mathbb{E}_t[(\theta_t,0)\mathbf{A}_t\mu_t\mathbf{A}_t^{-1}\mu_t] \\
&= d(\theta_t,0)x'_t\mathbf{A}_t^{-1}\mathbb{E}_t[\mu_t] + d\mathbb{E}_t[(\theta_t,0)\mathbf{A}_t\mu_t\mathbf{A}_t^{-1}\mu_t] \\
&= d\mathbb{E}_t[(\theta_t,0)\mathbf{A}_t\mu_t\mathbf{A}_t^{-1}\mu_t] \\
&= d\mathbf{A}_t^{-1}\mathbb{E}_t[\mu_t\mu_t^\top]\mathbf{A}_t(\theta_t,0)^\top \\
&= \mathbf{A}_t^{-1}(I - vv^\top)\mathbf{A}_t(\theta_t,0)^\top \\
&= (\theta_t,0)^\top - \mathbf{A}_t^{-1}vv^\top\mathbf{A}_t(\theta_t,0)^\top \\
&= (\theta_t,0)^\top - \frac{e_{d+1}e_{d+1}^\top\mathbf{A}_t^2}{\|\mathbf{A}_t e_{d+1}\|_2^2}(\theta_t,0)^\top.
\end{aligned}$$

Noticing that $\mathbb{E}_t[\mu_t] = 0$ by symmetry and the first d coordinates of $e_{d+1}e_{d+1}^\top\mathbf{A}_t^2(\theta_t,0)^\top$ are all zeros concludes the proof. ☐

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(Hazan et al., 2016), we can apply (Lemma 5.3. in Hazan et al. (2016)) to Algorithm 1 as follow.

Lemma 6. *For algorithm 1 and for every $u \in \mathcal{K}'$, $\sum_{t=1}^{T} g_t^\top x'_t - \sum_{t=1}^{T} g_t^\top u \le \sum_{t=1}^{T}[g_t^\top x'_t - g_t^\top x'_{t+1}] + \frac{1}{\eta}[\mathcal{R}(u) - \mathcal{R}(x'_1)]$.*

Proof. Define the functions $h_0(x') = \frac{1}{\eta}\mathcal{R}(x')$, $h_t(x') = g_t^\top x'$. We first prove for every $u \in \mathcal{K}'$,

$$\sum_{t=0}^{T} h_t(u) \ge \sum_{t=0}^{T} h_t(x'_{t+1}). \tag{14}$$

By induction on T:

Induction base: by definition, we have that $x'_1 = \arg\min_{x'\in\mathcal{K}'} \mathcal{R}(x')$, and thus $h_0(u) \ge h_0(x'_1)$ for all u.

Induction step: assume that for T, we have

$$\sum_{t=0}^{T} h_t(u) \ge \sum_{t=0}^{T} h_t(x'_{t+1}), \tag{15}$$

and let us prove the statement for $T+1$. Since $x'_{T+2} = \arg\min_{x'\in\mathcal{K}'} \sum_{t=0}^{T+1} h_t(x')$, we have:

$$\begin{aligned}
\sum_{t=0}^{T+1} h_t(u) &\ge \sum_{t=0}^{T+1} h_t(x'_{T+2}) \\
&\ge \sum_{t=0}^{T} h_t(x'_{T+2}) + h_{T+1}(x'_{T+2}) \\
&\ge \sum_{t=0}^{T} h_t(x'_{t+1}) + h_{T+1}(x'_{T+2}) \\
&= \sum_{t=0}^{T+1} h_t(x'_{t+1}),
\end{aligned}$$

Where in the third line we used the induction hypothesis for $u = x'_{T+2}$.

Then we conclude that,

$$\begin{aligned}\sum_{t=1}^{T} h_t(x'_t) - \sum_{t=1}^{T} h_t(u) &\leq \sum_{t=1}^{T} h_t(x'_t) - \sum_{t=1}^{T} h_t(x'_{t+1}) + [h_0(u) - h_0(x'_1)] \\ &= \sum_{t=1}^{T} [g_t^\top x'_t - g_t^\top x'_{t+1}] + \frac{1}{\eta}[\mathcal{R}(u) - \mathcal{R}(x'_1)]\end{aligned}$$

□

Lemma 7 (Theorem 2.2. in Lee et al. (2020))**.** *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*

$$\sum_{t=1}^{T} X_t \leq C(\sqrt{8V ln(C/\gamma)} + 2B^* ln(C/\gamma)), \tag{16}$$

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

Lemma 8. *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*

$$\|h\|_x \leq 2\nu. \tag{17}$$

Proof. From Lemma 1, we have

$$\|h\|_x \leq -\nabla\mathcal{R}(x)^\top h \leq |\nabla\mathcal{R}(x)^\top h|. \tag{18}$$

Then,

$$|\nabla\mathcal{R}(x)^\top h| = |\nabla\mathcal{R}(x)^\top(h - x + x)| \leq |\nabla\mathcal{R}(x)^\top(h - x)| + |\nabla\mathcal{R}(x)^\top x|. \tag{19}$$

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

Lemma 9. *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*

$$\sum_{t=1}^{T} \theta_t^\top x_t - \sum_{t=1}^{T} \theta_t^\top x^* \leq 2\eta d^2 T + \frac{\nu\log(\frac{1}{\delta})}{\eta} + Td\epsilon(2\nu + \sqrt{\nu}) + \delta DGT. \tag{20}$$

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

$$\|x^* - x^*_\delta\| = \min_{a\in\mathcal{K}_\delta} \|x^* - a\|. \tag{21}$$

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

$$\min_{a\in\mathcal{K}} \|x^* - a\| \leq \|x^* - (1 - \delta)x^*\| \leq \delta D. \tag{22}$$

So,

$$\|x^*_\delta - x^*\| \leq \delta D. \tag{23}$$

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

$$\sum_{t=1}^{T} \theta_t^\top x^*_\delta - \sum_{t=1}^{T} \theta_t^\top x^* \leq \delta DGT. \tag{24}$$

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 5, 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,

$$\begin{aligned}
\|M_t\|_{x_t'}^* &= \sqrt{M_t^\top\nabla^2(\mathcal{R}(x_t'))^{-1}M_t} \\
&= \sqrt{\mathbb{E}_t[d\sigma_t(y_t)\mathbf{A}_t^{-1}\mu_t]^\top\nabla^2(\mathcal{R}(x_t'))^{-1}\mathbb{E}_t[d\sigma_t(y_t)\mathbf{A}_t^{-1}\mu_t]} \\
&= \sqrt{d^2\mathbb{E}_t[\sigma_t(y_t)\mu_t]^\top\mathbf{A}_t^{-1}\mathbf{A}_t^2\mathbf{A}_t^{-1}\mathbb{E}_t[\sigma_t(y_t)\mu_t]} \\
&= \sqrt{d^2\mathbb{E}_t[\sigma_t(y_t)\mu_t]^\top\mathbb{E}_t[\sigma_t(y_t)\mu_t]} \\
&\leq \sqrt{d^2\mathbb{E}_t[\sigma_t^2(y_t)\mu_t^\top\mu_t]} \\
&\leq \sqrt{d^2\epsilon^2} \\
&= d\epsilon.
\end{aligned}$$

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 8, 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 6,

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

The proof of Lemma 3 implies $\|x'_t - x'_{t+1}\|_{x'_t} \le 4d\eta$ is true by choice of η and $\|g_t\|^*_{x'_t} \le d$. Therefore,

$$\|g_t\|^*_{x'_t}\|x'_t - x'_{t+1}\|_{x'_t} \le 4\eta d^2, \tag{25}$$

$$\mathbb{E}_t\{\sum_{t=1}^{T}[\|g_t\|^*_{x'_t}\|x'_t - x'_{t+1}\|_{x'_t}]\} \le 4\eta d^2 T. \tag{26}$$

With Lemma 4,

$$\frac{1}{\eta}(\mathcal{R}(x^{*'}_\delta) - \mathcal{R}(x'_1)) \le \frac{\nu\log(\frac{1}{\delta})}{\eta}. \tag{27}$$

Combine everything, we get

$$\sum_{t=1}^{T}\theta_t^\top x_t - \sum_{t=1}^{T}\theta_t^\top x^*_\delta \le 4\eta d^2 T + \frac{\nu\log(\frac{1}{\delta})}{\eta} + Td\epsilon(2\nu + \sqrt{\nu}). \tag{28}$$

□

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*

$$\mathbb{E}[\sum_{t=1}^{T} f_t(y_t) - \min_{x\in\mathcal{K}}\sum_{t=1}^{T} f_t(x)] \le 4d\sqrt{2\nu T\log T} + \frac{GD}{T} + 2T\epsilon + dT\epsilon(2\nu + \sqrt{\nu}). \tag{29}$$

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

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

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

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}[\theta_t^\top y_t]-\sum_{t=1}^{T}\mathbb{E}[\theta_t^\top x_t] &= \sum_{t=1}^{T}\mathbb{E}[\theta_t^\top(y_t-x_t)]\\ &= \sum_{t=1}^{T}\mathbb{E}[\mathbb{E}_t[\theta_t^\top(y_t-x_t)]]\\ &= \sum_{t=1}^{T}\mathbb{E}[\mathbb{E}_t[\theta_t^\top(\mathbf{A}_t\mu_t)]]\\ &= \sum_{t=1}^{T}\mathbb{E}[\theta_t^\top\mathbb{E}_t[(\mathbf{A}_t\mu_t)]]\\ &= \sum_{t=1}^{T}\mathbb{E}[\theta_t^\top\mathbf{0}]\\ &= \mathbf{0}.\end{aligned}$$

Thus,

$$\mathbb{E}[\sum_{t=1}^{T}\theta_t^\top y_t-\sum_{t=1}^{T}\theta_t^\top x^*]=\mathbb{E}[\sum_{t=1}^{T}\theta_t^\top x_t-\sum_{t=1}^{T}\theta_t^\top x^*]. \tag{30}$$

From Lemma 9, we have

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

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

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

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

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 $1-\gamma$*

$$\sum_{t=1}^{T}f_t(y_t)-\min_{x\in\mathcal{K}}\sum_{t=1}^{T}f_t(x)\leq 4d\sqrt{2\nu T\ln T}+\frac{GD}{T}+Td\epsilon(2\nu+\sqrt{\nu})+C(1+\epsilon)\sqrt{8T\ln\frac{C}{\gamma}}+2GD\ln\frac{C}{\gamma}+2T\epsilon \tag{31}$$

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

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,

$$\begin{aligned}\sum_{t=1}^{T} f_t(y_t) - \sum_{t=1}^{T} f_t(x^*) &= \sum_{t=1}^{T}[\theta_t^\top y_t + \sigma_t(y_t)] - \sum_{t=1}^{T}[\theta_t^\top x^* + \sigma_t(x^*)] \\ &\leq \sum_{t=1}^{T}\theta_t^\top y_t - \sum_{t=1}^{T}\theta_t^\top x^* + \sum_{t=1}^{T}\sigma_t(y_t) - \sum_{t=1}^{T}\sigma_t(x^*) \\ &\leq \sum_{t=1}^{T}\theta_t^\top y_t - \sum_{t=1}^{T}\theta_t^\top x^* + 2T\epsilon \\ &= \sum_{t=1}^{T}\theta_t^\top y_t - \sum_{t=1}^{T}\theta_t^\top x_t + \sum_{t=1}^{T}\theta_t^\top x_t - \sum_{t=1}^{T}\theta_t^\top x^* + 2T\epsilon.\end{aligned}$$

From Lemma 9, we know

$$\sum_{t=1}^{T}\theta_t^\top x_t - \sum_{t=1}^{T}\theta_t^\top x^* \leq 4d\sqrt{2\nu T\ln T} + \frac{GD}{T} + Td\epsilon(2\nu + \sqrt{\nu}), \tag{32}$$

where $\eta = \frac{\sqrt{2\nu\ln T}}{2d\sqrt{T}}, \delta = \frac{1}{T^2}$. Then by Lemma 7,

$$\mathbb{P}(\sum_{t=1}^{T}(\theta_t^\top y_t - \theta_t^\top x_t) \leq C(\sqrt{8V\ln(C/\gamma)} + 2B^*\ln(C/\gamma))) \geq 1-\gamma, \tag{33}$$

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

4.1 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. Let $\hat{x}$ be the output of algorithm 1, then from Theorem 1. we can easily prove and ensure $f(\hat{x}) - \min_{x\in\mathcal{K}} f(x) \leq \frac{4d\sqrt{2\nu\ln T}}{\sqrt{T}} + \frac{GD}{T^2} + d\epsilon(2\nu + \sqrt{\nu}) + 2\epsilon$.

5 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*

$$f(\hat{x}) - \min_{x\in\mathcal{K}} f(x) \geq 2\epsilon. \tag{34}$$

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,

$$x_t = \mathcal{A}((x_1, y_1), ..., (x_{t-1}, y_{t-1})). \tag{35}$$

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

$$f(x) = \begin{cases} \epsilon, & x \neq z, \\ -\epsilon, & x = z. \end{cases} \tag{36}$$

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

$$f(\hat{x}) - \min_{x \in \mathcal{K}} f(x) \geq 2\epsilon. \tag{37}$$

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 (36), 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

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Jacob D Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *COLT*, pp. 263–274, 2008.

Naman Agarwal, Alon Gonen, and Elad Hazan. Learning in non-convex games with an optimization oracle. In *Conference on Learning Theory*, pp. 18–29. PMLR, 2019.

Sanae Amani, Mahnoosh Alizadeh, and Christos Thrampoulidis. Linear stochastic bandits under safety constraints. *Advances in Neural Information Processing Systems*, 32, 2019.

Baruch Awerbuch and Robert D Kleinberg. 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.

Sébastien Bubeck, Nicolo Cesa-Bianchi, and Sham M Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Conference on Learning Theory*, pp. 41–1. JMLR Workshop and Conference Proceedings, 2012.

Varsha Dani, Sham M Kakade, and Thomas Hayes. The price of bandit information for online optimization. *Advances in Neural Information Processing Systems*, 20, 2007.

Abraham D Flaxman, Adam Tauman Kalai, and H Brendan Mcmahan. 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. doi: 10.5555/1070432.

Xiand Gao, Xiaobo Li, and Shuzhong Zhang. Online learning with non-convex losses and non-stationary regret. In *International Conference on Artificial Intelligence and Statistics*, pp. 235–243. PMLR, 2018.

Udaya Ghai, Zhou Lu, and Elad Hazan. Non-convex online learning via algorithmic equivalence. *Advances in Neural Information Processing Systems*, 35:22161–22172, 2022.

Elad Hazan and Kfir Levy. Bandit convex optimization: Towards tight bounds. *Advances in Neural Information Processing Systems*, 27, 2014.

Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2 (3-4):157–325, 2016.

Tor Lattimore. Improved regret for zeroth-order adversarial bandit convex optimisation. *Mathematical Statistics and Learning*, 2:311–334, 10 2020. ISSN 2520-2316. doi: 10.4171/msl/17.

Chung-Wei Lee, Haipeng Luo, Chen-Yu Wei, and Mengxiao Zhang. Bias no more: high-probability data-dependent regret bounds for adversarial bandits and mdps. *Advances in neural information processing systems*, 33:15522–15533, 2020.

H Brendan McMahan and Avrim Blum. 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.

Arkadi Nemirovski. Interior point polynomial time methods in convex programming. *Lecture notes*, 42(16): 3215–3224, 2004.

Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

Lin Yang, Lei Deng, Mohammad H Hajiesmaili, Cheng Tan, and Wing Shing Wong. An optimal algorithm for online non-convex learning. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2(2):1–25, 2018.