An Optimal Algorithm for Bandit Convex Optimization with Strongly-Convex and Smooth Loss

Shinji Ito
NEC Corporation, The University of Tokyo

Abstract

We consider non-stochastic bandit convex optimization with strongly-convex and smooth loss functions. For this problem, Hazan and Levy have proposed an algorithm with a regret bound of $\tilde{O}(d^{3/2}\sqrt{T})$ given access to an $O(d)$-self-concordant barrier over the feasible region, where d and T stand for the dimensionality of the feasible region and the number of rounds, respectively. However, there are no known efficient ways for constructing self-concordant barriers for general convex sets, and a $\tilde{O}(\sqrt{d})$ gap has remained between the upper and lower bounds, as the known regret lower bound is $\Omega(d\sqrt{T})$. Our study resolves these two issues by introducing an algorithm that achieves an optimal regret bound of $\tilde{O}(d\sqrt{T})$ under a mild assumption, without self-concordant barriers. More precisely, the algorithm requires only a membership oracle for the feasible region, and it achieves an optimal regret bound of $\tilde{O}(d\sqrt{T})$ under the assumption that the optimal solution is an interior of the feasible region. Even without this assumption, our algorithm achieves $\tilde{O}(d^{3/2}\sqrt{T})$-regret.

1 Introduction

Bandit convex optimization (BCO) is a framework for online decision-making with limited feedback. In this framework, a player is given a convex *feasible region* $\mathcal{K}$ and the number T of *rounds*. In each round $t = 1, 2, \ldots, T$, the player chooses an *action* $a_t \in \mathcal{K}$, and the environment independently chooses convex *loss function* $f_t : \mathcal{K} \to [-1, 1]$. Not all information about the loss function is revealed to the player then, but only the *bandit feedback* is available, i.e., the player can observe $f_t(x_t)$ alone. The goal of the player is to minimize cumulative loss $\sum_{t=1}^{T} f_t(x_t)$, and performance is evaluated in terms of the *regret* $R_T(x^*)$ defined by

$$R_T(x^*) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(x^*) \tag{1}$$

for $x^* \in \mathcal{K}$.

This paper focuses on a *non-stochastic* or *adversarial* setting. In this setting, we do not assume any generative model for the loss functions f_t, and f_t can change arbitrarily. An alternative setting, a *stochastic* setting in which f_t independently follows an unknown probabilistic distribution, can be regarded as a special case of the non-stochastic setting. Indeed, algorithms for the non-stochastic setting work even for this stochastic setting, and regret upper bounds for the former setting apply even to the latter.

Work on non-stochastic BCO was initiated by Flaxman et al. [2005] and Kleinberg [2005], in which algorithms with regret bounds of $O(T^{3/4})$ were proposed. Note that there has been a lower bound of $\Omega(\sqrt{T})$, i.e., it is known that no algorithm can achieve a better regret bound than $O(\sqrt{T})$. A gap of $O(T^{1/4})$ between the upper and the lower bounds remained for a long time, until Bubeck et al. [2015], Bubeck and Eldan [2016] and Bubeck et al. [2017] proposed algorithms with $\tilde{O}(\sqrt{T})$-regret bounds, where $\tilde{O}(\cdot)$ notation ignores factors of poly-logarithmic terms. There has still been a large gap, however, w.r.t. the dimension d of the feasible region $\mathcal{K}$; the best known upper and lower bounds are of $\tilde{O}(d^{9.5}\sqrt{T})$ and $\Omega(d\sqrt{T})$, respectively. Bubeck et al. [2017] have conjectured that the optimal regret bound is of $\tilde{\Theta}(d^{3/2}\sqrt{T})$, but there have been no significant improvements since their study.

This paper focuses on an important special case of BCO in which the loss functions are strongly-convex and smooth. Work on such special cases is summarized in Table 1. Agarwal et al. [2010] showed that a modified version of the algorithm by Flaxman et al. [2005] can achieve a regret of $\tilde{O}(d\sqrt{T})$ for unconstrained problems, i.e., for problems with $\mathcal{K} = \mathbb{R}^d$. This result can be said to be minimax optimal because Shamir [2013] proved a lower bound of $\Omega(d\sqrt{T})$ that holds even for strongly-convex and smooth

Proceedings of the 23rdInternational Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108.

Table 1: Regret bound for bandit convex optimization with strongly-convex and smooth objective functions.

Reference	Regret bound	Notes
Flaxman et al. [2005],	$\tilde{O}(d^{2/3}T^{2/3})$	No additional assumptions.
Agarwal et al. [2010]	$\tilde{O}(d\sqrt{T})$	Assume that the optimization is unconstrained, i.e., $\mathcal{K} = \mathbb{R}^d$.
Hazan and Levy [2014]	$\tilde{O}(d\sqrt{\nu T})$	Require a ν-self-concordant barrier for $\mathcal{K}$. Parameter ν is at least d.
Corollary 1 [This work]	$\tilde{O}(d\sqrt{T})$	Assume that the optimal solution is an interior of $\mathcal{K}$.
Corollary 2 [This work]	$\tilde{O}(d^{3/2}\sqrt{T})$	No additional assumptions.
Shamir [2013]	$\Omega(d\sqrt{T})$	A lower bound that implies $O(d\sqrt{T})$-bounds are minimax optimal.

losses. On the other hand, for general constrained problems, the minimax optimal rate remains to be determined.

An important sign of progress in strongly-convex and smooth BCO has been shown by Hazan and Levy [2014]. They proposed an algorithm that can be applied to constrained problems and has a better regret bound. However, this algorithm does not directly apply to general problems because it requires a *ν-self-concordant barrier*[1] for the feasible region $\mathcal{K}$, where a self-concordant barrier is a convex function with certain properties and $\nu > 0$ is a parameter of it. For some special cases of convex sets, we have explicit forms of self-concordant barriers; for example, if $\mathcal{K}$ can be expressed by m linear inequalities, one has an m-self-concordant barrier for $\mathcal{K}$. For a general convex set, however, there are no known efficient ways for constructing a self-concordant barrier. Further, even if we were to have a ν-self-concordant barrier, the regret bound would be $O(d\sqrt{\nu T})$, which implies that there would still be a $\sqrt{\nu}$ gap between the upper and the lower bounds. Because ν is in general at least d for a compact convex set $\mathcal{K}$ (see, e.g., Nesterov and Nemirovskii [1994]), there is a gap of $\Omega(\sqrt{d})$ from the lower bound of $O(d\sqrt{T})$, and if we were to have only self-concordant barriers with a large ν, e.g., if $\mathcal{K}$ were expressed by $m(\gg d)$ linear inequalities, the gap would be even worse.

Our contribution is to overcome the above issues by developing a novel algorithm with the following two strengths. (i) Under a mild assumption, our algorithm achieves $\tilde{O}(d\sqrt{T})$-regret, which is minimax optimal up to logarithmic factors. This represents the first tight bound for bandit convex optimization that applies even to constrained problems. More precisely, under the assumption that the optimal solution is an r-interior,[2] our algorithm enjoys a regret bound of $\tilde{O}(d\sqrt{T} + d^2/r^2)$, as given in Corollary 1. Also, even without the assumption of interiors, the algorithm has a regret bound of $\tilde{O}(d^{3/2}\sqrt{T})$, which is, at least, not worse than existing algorithms. (ii) Our algorithm does not require self-concordant barriers. Indeed, we only assume that we have access to a membership oracle for the feasible region. This means that our algorithm works well even if $\mathcal{K}$ is expressed by an exponentially large number of linear inequalities, or if we are not given explicit forms of $\mathcal{K}$.

A key ingredient in our algorithm is the *multiplicative weight update* (MWU) method [Arora et al., 2012], in which we update probabilistic distributions over $\mathcal{K}$ on the basis of estimators of objective functions. To estimate objective functions from bandit feedback, we use techniques of *smoothing* and *ellipsoidal sampling* [Flaxman et al., 2005]. Our analyses for regret bounds rely on theories for *log-concave distributions* [Lovász and Vempala, 2007], which is a class of continuous distributions that includes normal distributions, exponential distributions, and distributions in our algorithm. Further, this algorithm can be implemented so that it runs in polynomial time, thanks to efficient algorithms for sampling from log-concave distributions [Lovász and Vempala, 2007, Narayanan and Rakhlin, 2017].

2 Related Work

For *bandit linear optimization*, an important special case of BCO in which objectives are linear functions, there have been many signs of progress. Bubeck et al. [2012] and Cesa-Bianchi and Lugosi [2012] provided algorithms that achieve regret of $\tilde{O}(d\sqrt{T})$. These algorithms can be applied to *combinatorial bandits*, bandit linear optimization problems in which the feasible region $\mathcal{K}$ is a discrete finite set. The regret bounds of $\tilde{O}(d\sqrt{T})$ can be said to be non-improvable because Dani et al. [2008] showed a regret lower bound of $\Omega(d\sqrt{T})$. The computational complexity for combinatorial bandits depends on the feasible region $\mathcal{K}$, as mentioned in [Cesa-Bianchi and Lugosi, 2012]. Hazan and Karnin [2016] have proposed a computationally efficient algorithm that achieves $\tilde{O}(d\sqrt{T})$-regret if $\mathcal{K}$ is a convex set. For general $\mathcal{K}$ including discrete sets, Ito et al. [2019] have proposed an algorithm with a regret bound of $\tilde{O}(d^{3/2}\sqrt{T})$, which runs efficiently given an algorithm for linear optimization over $\mathcal{K}$.

Online convex optimization [Shalev-Shwartz, 2012, Hazan, 2016, Cesa-Bianchi and Lugosi, 2006] is a variant of BCO in which a player can get feedback on complete infor-

[1] Self-concordant barriers are special cases of convex functions that were introduced in order to develop interior-point methods for convex optimization. For details on self-concordant barriers, see, e.g., [Nesterov and Nemirovskii, 1994].

[2] The definition of r-interior is given in Section 3 and Figure 1.

mation about objective functions, rather than bandit feedback. For general convex objectives, it has been known that online gradient descent methods [Zinkevich, 2003] can achieve $O(\sqrt{dT})$ regret, and this bound is minimax optimal. For a special case called *exp-concave functions*, which involves a milder assumption than strong-convexity, Hazan et al. [2007] provided efficient algorithms with a regret bound of $O(d \log T)$, which is minimax optimal as well because there is a lower bound of $\Omega(d \log T)$ [Ordentlich and Cover, 1998].

It has been shown that one can achieve better regret for BCO if *multi-point feedback* is available, i.e., if the player can observe the values of objective functions on $k \geq 2$ different points in each round [Agarwal et al., 2010, Nesterov and Spokoiny, 2017, Duchi et al., 2015]. For general convex functions, it is known that one can achieve $O(d^2\sqrt{T})$ regret in a two-point feedback setting. Further, Agarwal et al. [2010] have shown that, under the assumption of strong-convexity and two-point feedback, one can achieve $O(d^2 \log T)$ regret. Agarwal et al. [2010] have performed a detailed analysis on the one-point gradient estimator with spherical perturbations [Flaxman et al., 2005] as well. On the basis of their analysis, we can see that spherical perturbations of radius $O(\min\{r, d/T^{1/4}\})$ provide a similar regret bound as Corollary 1 ($\log T$ instead of $\log d$). This approach, however, requires algorithms for projection onto $\mathcal{K}$ and does not work well for the case of small r, in contrast to our algorithm.

After the study by Hazan and Levy [2014], many works regarding BCO have followed. Bubeck and Eldan [2015] found the *entropic barrier*, which is a nearly d-self-concordant barrier for general convex sets. The algorithm by Hazan and Levy [2014] with the entropic barrier (with $\nu = d$) achieves an $\tilde{O}(d^{3/2}\sqrt{T})$-regret bounds, as Table 1 shows. In terms of computational complexity, however, the entropic barrier and corresponding optimization problems have not been proven to be efficiently computable. Hu et al. [2016] have considered a more general problem setting in which *biased noisy gradient* is available. Mohri and Yang [2016] provided a BCO algorithm that does not require a priori assumptions of strong convexity or smoothness. These two studies capture more general scenarios than ours, but their regret bounds achieved in our setting are not superior to those by Hazan and Levy [2014]. Kumagai [2017] considered a *dueling bandit* problem with strongly-convex and smooth costs, in which an algorithm based on self-concordant functions was proposed. Chen et al. [2019] focused on computationally efficient methods for BCO, and provided a projection-free algorithm that achieves sublinear regret for general convex losses.

Our proposed algorithm is based on the multiplicative weight update (MWU) method [Arora et al., 2012, Hoeven et al., 2018]. Algorithms similar to MWU can be found in the literature in the early 1950s in the context of game theory [Brown and Von Neumann, 1950, Brown, 1951, Robinson, 1951], and MWU has been independently rediscovered in other fields including computational geometry, and machine learning. Our approach is to use continuous MWU, multiplicative weight update over continuous domains, which has been applied to various online optimization problems, including Cover's universal portfolios [Cover, 1991], bandit linear optimization [Hazan and Karnin, 2016], and online improper learning [Hazan et al., 2018].

3 Problem Setting and Assumption

A player is given a convex *feasible region* $\mathcal{K} \subseteq \mathbb{R}^d$ and a number T of rounds of decision making, where d is a positive integer standing for the dimensionality of the feasible region. For each $t = 1, 2, \ldots, T$, the player chooses an *action* $a_t \in \mathcal{K}$, and an environment chooses a convex function $f_t : \mathcal{K} \to [-1, 1]$ at the same time. The player observes feedback of $f_t(a_t)$ before choosing the next action a_{t+1}. We assume that $\mathcal{K}$ has a positive volume, i.e., $\int_{x \in \mathcal{K}} 1 \mathrm{d}x > 0$. We assume that f_t is σ-strongly convex and β-smooth, i.e., that the following hold for all $x, y \in \mathcal{K}$:

$$f_t(y) \geq f_t(x) + \nabla f_t(x)^\top (y - x) + \frac{\sigma}{2}\|y - x\|_2^2, \qquad (2)$$

$$f_t(y) \leq f_t(x) + \nabla f_t(x)^\top (y - x) + \frac{\beta}{2}\|y - x\|_2^2, \qquad (3)$$

where $\nabla f_t(x) \in \mathbb{R}^d$ stands for the gradient of f_t at x.

The performance of the player is evaluated in terms of the *regret* $R_T(x^*)$, which is defined as

$$R_T(x^*) = \sum_{t=1}^{T} f_t(a_t) - \sum_{t=1}^{T} f_t(x^*). \qquad (4)$$

In this paper, we suppose that a player arbitrarily chooses a convex *benchmark set* $\mathcal{K}' \subseteq \mathcal{K}$. We consider regret $R_T(x^*)$ for $x^* \in \mathcal{K}'$, i.e., we care about the value of $\sup_{x^* \in \mathcal{K}'} \mathbf{E}[R_T(x^*)]$, the expected gap between the cumulative losses for the algorithm's outputs and for the optimal single action x^* belonging to $\mathcal{K}'$. The value $\sup_{x^* \in \mathcal{K}'} \mathbf{E}[R_T(x^*)]$ is equal to the standard worst-case regret $\sup_{x^* \in \mathcal{K}} \mathbf{E}[R_T(x^*)]$, if the optimal single action $x^* \in \arg\min_{x \in \mathcal{K}} \sum_{t=1}^{T} f_t(x)$ belongs to $\mathcal{K}'$.

A point $x \in \mathbb{R}^d$ is called a γ-*interior* of $\mathcal{K}$ if $\|y - x\|_2 \leq \gamma$ implies $y \in \mathcal{K}$. For example, if $\mathcal{K}$ is expressed by m linear inequalities, i.e., $\mathcal{K}$ can be expressed as $\mathcal{K} = \{x \in \mathbb{R}^d \mid a_j^\top x \leq b_j \quad (j \in [m])\}$ with $a_j \in \mathbb{R}^d, b_j \in \mathbb{R}$ such that $\|a_j\|_2 = 1$, then the convex set $\mathcal{K}'$ defined by $\mathcal{K}' = \{x \in \mathbb{R}^d \mid a_j^\top x \leq b_j - r \quad (j \in [m])\}$ consists of r-interiors of $\mathcal{K}$. For general benchmark set $\mathcal{K}' \subseteq \mathcal{K}$, let $r \geq 0$ be a non-negative real value for which all members of $\mathcal{K}'$ are r-interiors. Figure 1 shows a geometric interpretation of how

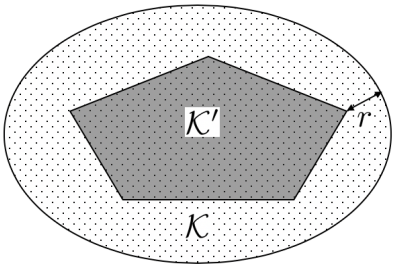

Figure 1: Geometric interpretation of $\mathcal{K}'$ that consists of r-interiors of $\mathcal{K}$.

r is determined for $\mathcal{K}$ and $\mathcal{K}' \subseteq \mathcal{K}$. For the special case in which $\mathcal{K}' = \mathcal{K}$, r is equal to zero.

We assume that we have access to a *membership oracle* for $\mathcal{K}'$. This means that, given $x \in \mathbb{R}^d$, we can determine if $x \in \mathcal{K}'$ or not, by calling on the membership oracle. If $\mathcal{K}'$ is expressed by m inequalities ($\mathcal{K}' = \{x \in \mathbb{R}^d \mid g_j(x) \leq 0 \quad (j \in [m])\}$), then we have access to the membership oracle for $\mathcal{K}'$ because one can check if $x \in \mathcal{K}$ by evaluating $g_i(x)$ for $i \in [m]$. Further, it is known that we will have a polynomial-time membership oracle for $\mathcal{K}'$ if we can solve linear optimization problems over $\mathcal{K}'$ in polynomial time [Schrijver, 1998].

4 Algorithm

4.1 Preliminary

Notation For a vector $x = (x_1, \ldots, x_d)^\top \in \mathbb{R}^d$, let $\|x\|_2$ denote the ℓ_2 norm of x, i.e., $\|x\|_2 = \sqrt{x^\top x} = \sqrt{\sum_{i=1}^d x_i^2}$. For a matrix $X \in \mathbb{R}^{d \times d}$, $\|X\|_2$ denotes the ℓ_2-operator norm, i.e., $\|X\|_2 = \max\{\|Xy\|_2 \mid y \in \mathbb{R}^d, \|y\|_2 = 1\}$. If X is a symmetric matrix, $\|X\|_2$ is equal to the maximum absolute value of eigenvalues of X. Given a positive semidefinite matrix $A \in \mathbb{R}^{d \times d}$ and a vector $x \in \mathbb{R}^d$, define $\|x\|_A$ by $\|x\|_A = \sqrt{x^\top A x} = \|A^{1/2}x\|_2$. Similarly, for a matrix $X \in \mathbb{R}^d$, let $\|X\|_A = \|A^{1/2} X A^{1/2}\|_2$. Given two symmetric matrices $A, B \in \mathbb{R}^{d \times d}$, denote $A \bullet B = \mathrm{tr}(AB)$.

Smoothed convex function Let v and u be random variables that follow uniform distributions over $\mathbb{B}^d = \{v \in \mathbb{R}^d \mid \|v\|_2 \leq 1\}$ and $\mathbb{S}^d = \{u \in \mathbb{R}^d \mid \|u\|_2 = 1\}$, respectively. For a convex function f over $\mathbb{R}^d$ and a positive-definite matrix $B \in \mathbb{R}^{d \times d}$, define the *smoothed function* $\hat{f}$ by

$$\hat{f}_B(x) = \mathbf{E}[f(x + Bv)]. \tag{5}$$

Then we have the following:

Lemma 1 (Hazan and Levy [2014])**.** *The gradient of $\hat{f}_B$ can be expressed as*

$$\nabla \hat{f}_B(x) = \mathbf{E}[d \cdot f(x + Bu) B^{-1} u]. \tag{6}$$

If f is β-smooth, it holds that

$$0 \leq \hat{f}_B(x) - f(x) \leq \frac{\beta}{2}\|B^\top B\|_2 = \frac{\beta}{2}\lambda_1(B^\top B). \tag{7}$$

If f is σ-strongly convex then so is $\hat{f}_B$.

Equation (6) can be shown using Stokes' theorem, and (7) follows from the definition of β-smoothness. In the bandit feedback setting, though unbiased estimators of the gradient of f_t are unavailable, those for the smoothed ones $\hat{f}_t$ can be constructed through (6). Differences between f_t and $\hat{f}_t$ will be bounded by means of (7).

Log-concave distribution A probability distribution over a convex set $\mathcal{K} \subseteq \mathbb{R}^d$ is called a *log-concave distribution* if its probability density function $p : \mathcal{K} \to \mathbb{R}$ can be expressed as $p(x) = \exp(-g(x))$ with a convex function $g : \mathcal{K} \to \mathbb{R}$, i.e., the logarithmic of $p(x)$ is a concave function. Our algorithm maintains log-concave distributions. Random samples from log-concave distributions can be efficiently generated under mild assumptions. Indeed, as shown in [Lovász and Vempala, 2007], there are computationally efficient MCMC algorithms for sampling from p that work given a membership oracle for $\mathcal{K}$ and an evaluation oracle for g. Accordingly, we can efficiently compute estimators of the mean $\mu(p)$ and the covariance matrix $\mathrm{Cov}(p)$ for p. The following lemma is useful for bounding the variance of log-concave distributions:

Lemma 2 (Prop. 10.1. in [Saumard and Wellner, 2014])**.** *Suppose that a log-concave distribution over $\mathcal{K}$ has a probability density function $p(x) = \exp(-g(x))$, where g is a σ-strongly convex function. Then, the covariance matrix Σ of p satisfies $\|\Sigma\|_2 \leq 1/\sigma$.*

The following lemma is used to prove a regret bound for our algorithm.

Lemma 3 (Lemma 5.7 in [Lovász and Vempala, 2007])**.** *Let X be a random point drawn from a log-concave distribution on $\mathbb{R}$. Assume that $\mathbf{E}[X^2] \leq 1$. Then for every $t > 1$, $\mathrm{Prob}[|X| > t] \leq \exp(-t + 1)$.*

We use the following lemma to guarantee that the outputs a_t of our algorithm are included in $\mathcal{K}$.

Lemma 4 (Lemma 5.5 (a) and Lemma 5.12 in [Lovász and Vempala, 2007])**.** *Let p be a log-concave distribution over $\mathcal{K}$. The ellipsoid $\{x \in \mathbb{R}^d \mid \|x - \mu(p)\|_{\mathrm{Cov}(p)^{-1}} \leq 1/\mathrm{e}\}$ will then be included in $\mathcal{K}$.*

4.2 Continuous Multiplicative Weight Update

In our algorithm, we maintain a function z_t over $\mathcal{K}'$ using the *multiplicative weight update* method [Arora et al.,

2012]. We initialize z_t by $z_1(x) = \sigma\|x\|_2^2/2$. In each round, let p_t be a probability distribution over $\mathcal{K}'$ with density proportional to $\exp(-\eta z_t(x))$, i.e., p_t is defined by

$$Z_t := \int_{x\in\mathcal{K}'} \exp(-\eta z_t(x))\mathrm{d}z, \quad p_t(x) = \frac{\exp(-\eta z_t(x))}{Z_t}. \tag{8}$$

Let μ_t and Σ_t denote the mean and the covariance matrix for p_t. We then compute estimators $\hat{\mu}_t$ and $\hat{\Sigma}_t$ for them such that

$$\|\hat{\mu}_t - \mu_t\|_{\Sigma_t^{-1}} \le 1/9, \quad \|\hat{\Sigma}_t - \Sigma_t\|_{\Sigma_t^{-1}} \le 1/9,$$
$$\mathbf{E}[\hat{\mu}_t|\mu_t] = \mu_t. \tag{9}$$

Specific methods for computing such $\hat{\mu}_t$ and $\hat{\Sigma}_t$ will be discussed in the next subsection. Let $B_t \in \mathbb{R}^{d\times d}$ be a positive-definite matrix for which $B_tB_t = \hat{\Sigma}_t$. Such a matrix can be computed, e.g., via eigenvalue decomposition. Consider smoothing f_t as (5) with $B = \alpha_t B_t$:

$$\hat{f}_t(x) := \mathbf{E}[f_t(x + \alpha_t B_t v)], \tag{10}$$

where α_t is the *exploration parameter* that we will adjust later, and v follows a uniform distribution over $\mathbb{B}^d$. An unbiased estimator of the gradient of $\hat{f}_t$ can then be constructed as follows. Choose u_t from a unit sphere $\mathbb{S}^d = \{u \in \mathbb{R}^d \mid \|u\|_2 = 1\}$, uniformly at random. Play an action of $a_t = \mu_t + \alpha_t B_t u_t$, and then observe $f_t(a_t)$. On the basis of this observation, define $\hat{g}_t \in \mathbb{R}^d$ by

$$\hat{g}_t = d \cdot f_t(a_t)(\alpha_t B_t)^{-1} u_t. \tag{11}$$

This is an unbiased estimator of the gradient $\nabla\hat{f}_t(\hat{\mu}_t)$, i.e., given $\hat{\mu}_t$ and B_t, the conditional expectation of $\hat{g}_t$ satisfies

$$\mathbf{E}[\hat{g}_t] = \mathbf{E}[d \cdot f_t(\hat{\mu}_t + \alpha_t B_t u_t)(\alpha_t B_t)^{-1} u_t] = \nabla\hat{f}_t(\hat{\mu}_t), \tag{12}$$

where the second inequality follows from (6). By means of this unbiased estimator $\hat{g}_t$, we update z_t as

$$z_{t+1}(x) = z_t(x) + \hat{g}_t^\top(x - \hat{\mu}_t) + \frac{\sigma}{2}\|x - \hat{\mu}_t\|_2^2. \tag{13}$$

4.3 Computation of $\hat{\mu}_t$ and $\hat{\Sigma}_t$

Estimators $\hat{\mu}_t$ and $\hat{\Sigma}_t$ satisfying (9) can be computed from samples $x_t^{(1)}, \dots, x_t^{(M)}$ generated by p_t as follows:

$$\hat{\mu}_t = \frac{1}{M}\sum_{j=1}^{M} x_t^{(j)}, \quad \hat{\Sigma}_t = \frac{1}{M}\sum_{j=1}^{M}(x_t^{(j)} - \hat{\mu}_t)(x_t^{(j)} - \hat{\mu}_t^{(j)}).$$

If we set M sufficiently large, (9) holds with high probability.

The remaining problem is how to get samples from p_t. A simple way for this is to use normal distributions; since p_t is defined by $z_1(x) = \sigma\|x\|_2^2/2$, (8) and (13), the distribution p_t is a multidimensional truncated normal distribution over $\mathcal{K}$ expressed as

$$p_t(x) \propto \exp(-\sigma\eta t\|x - \theta_t\|_2^2/2) \qquad (x \in \mathcal{K}'),$$
$$p_t(x) = 0 \qquad (x \in \mathbb{R}^d \setminus \mathcal{K}'),$$

where $\theta_t = \frac{1}{t}\sum_{j=1}^{t-1}(\hat{\mu}_j - \frac{1}{\sigma}\hat{g}_j)$. Hence, by sampling x from a normal distribution $\mathcal{N}(\theta_t, \frac{1}{\sigma\eta t}I)$ until $x \in \mathcal{K}'$, we can get x following p_t. Note that this procedure cannot, however, always terminate in polynomial time, though it will be practical enough in many cases. Even if the simple procedure does not work well, we can apply an alternative polynomial-time sampling method based on MCMC [Lovász and Vempala, 2007], since p_t is a log-concave distribution. For more efficient ways of computing $\hat{\mu}$ and $\hat{\Sigma}$ and sampling from p_t, see, e.g., [Belloni et al., 2015, Narayanan and Rakhlin, 2017].

Algorithm 1 Continuous multiplicative weight update method for bandit convex optimization

Require: Time horizon $T \in \mathbb{N}$, learning rate $\eta > 0$, membership oracle $\mathcal{M}$ for $\mathcal{K}'$, exploration parameters $\{\alpha_t\}_{t=1}^T \subseteq \mathbb{R}_{>0}$, strong-convexity parameter $\sigma > 0$.
1: Set $z_1 : \mathcal{K}' \to \mathbb{R}$ by $z_1(x) = \sigma\|x\|_2^2/2$.
2: **for** $t = 1, 2, \dots, T$ **do**
3: Compute $\hat{\mu}_t$ and $\hat{\Sigma}_t$ for which (9) holds.
4: Compute a positive-definite matrix $B_t \in \mathbb{R}^{d\times d}$ such that $B_tB_t = \hat{\Sigma}_t$.
5: Pick u_t from $\mathbb{S}^d$ uniformly at random.
6: Play $a_t = \hat{\mu}_t + \alpha_t B_t u_t$ and observe $f_t(a_t)$.
7: Set $\hat{g}_t$ by (11) and update z_t by (13).
8: **end for**

4.4 Choice of Exploration Parameter α_t

It is necessary to choose α_t so that $a_t = \hat{\mu}_t + \alpha_t B_t u_t$ is a feasible solution, i.e., $a_t \in \mathcal{K}$. The following proposition provides a sufficient condition for this.

Proposition 1. *If α_t is bounded as $0 < \alpha_t \le 1/9 + r\sqrt{t\eta\sigma/2}$ then $a_t = \hat{\mu}_t + \alpha_t B_t u_t$ is in $\mathcal{K}$.*

Proof. Let α_{t1} and α_{t2} be positive numbers such that $\alpha_{t1} \le 1/9$, $\alpha_{t2} \le r\sqrt{t\eta\sigma/2}$ and $\alpha_t = \alpha_{t1} + \alpha_{t2}$. Since a_t can be expressed as $a_t = \hat{\mu}_t + \alpha_{t1}B_tu_t + \alpha_{t2}B_tu_t$ and since all points of $\mathcal{K}'$ are r-interior of $\mathcal{K}$, it suffices to show that (i) $\hat{\mu}_t + \alpha_{t1}B_tu_t \in \mathcal{K}'$ and (ii) $\|\alpha_{t2}B_tu_t\|_2 \le r$.

From Lemma 4, $\|\hat{\mu}_t + \alpha_{t1}B_tu_t - \mu_t\|_{\Sigma_t^{-1}} \le 1/\mathrm{e}$ implies $\hat{\mu}_t + \alpha_{t1}B_tu_t \in \mathcal{K}'$. From the triangle inequality, we have

$$\|\hat{\mu}_t + \alpha_{t1}B_tu_t - \mu_t\|_{\Sigma_t^{-1}} \le \|\hat{\mu}_t - \mu_t\|_{\Sigma_t^{-1}} + \alpha_{t1}\|B_tu_t\|_{\Sigma_t^{-1}}$$
$$\le \frac{1}{9} + \frac{1}{9}\|\Sigma_t^{-1/2}B_tu_t\|_2 \le \frac{1}{9}(1 + \|\Sigma_t^{-1/2}B_t\|_2). \tag{14}$$

From (9), we have $\|\Sigma_t^{-1/2} B_t\|_2 \leq 2$. Combining this and (14), we have $\|\hat{\mu}_t + \alpha_{t1} B_t u_t - \mu_t\|_{\Sigma_t^{-1}} \leq 1/3 \leq 1/\mathrm{e}$, which implies that (i) holds.

Since $\eta z_t(x)$ is a $(t\eta\sigma)$-strongly convex function, from Lemma 2, the covariance matrix $\Sigma_t = \mathrm{Cov}(p_t)$ is bounded as $\|\Sigma_t\|_2 \leq 1/(t\eta\sigma)$. From this and (9), we have $\|\hat{\Sigma}_t\|_2 \leq 2/(t\eta\sigma)$. Accordingly, we have

$$\|B_t\|_2 \leq \sqrt{\|\hat{\Sigma}_t\|_2} \leq \sqrt{2/(t\eta\sigma)}. \tag{15}$$

From this, $\|u_t\| = 1$, and $\alpha_t \leq r\sqrt{t\eta\sigma/2}$, we have (ii). □

Proposition 1 implies that, under the assumption of $0 < \alpha_t \leq 1/9 + r\sqrt{t\eta\sigma/2}$, for arbitrary $x \in \mathcal{K}'$, the value of $\hat{f}_t(x)$ can be defined by (10). In addition, we have $\hat{f}_t(x) \in [-1, 1]$ for $x \in \mathcal{K}'$ since $\hat{f}_t(x)$ is defined to be a convex combination of values of $f_t(y)$ for $y \in \mathcal{K}$. Hereafter, we suppose that $0 < \alpha_t \leq 1/9 + r\sqrt{t\eta\sigma/2}$ holds.

5 Regret Analysis

This section shows regret upper bounds for Algorithm 1.

5.1 Main Results

We analyze the expected regret for the case in which α_t is defined by

$$\alpha_t = \min\left\{\frac{1}{9} + r\sqrt{\frac{t\eta\sigma}{2}}, \sqrt{d}\right\}. \tag{16}$$

We also assume that the learning rate η is bounded as

$$\eta \leq \frac{\alpha_t}{2d\log(50T)}. \tag{17}$$

We then have the following regret bound:

Theorem 1. *Suppose that α_t is chosen as* (16) *and that η satisfies* (17)*. Then, for the output of Algorithm 1 and for arbitrary $x^* \in \mathcal{K}'$, the regret is bounded as*

$$\mathbf{E}[R_T(x^*)] = O\left(\frac{d\beta\log T}{\sigma\eta} + \eta d T + \frac{d^2\log d}{r^2\sigma}\right)$$

and

$$\mathbf{E}[R_T(x^*)] = O\left(\frac{d\beta\log T}{\sigma\eta} + \eta d^2 T\right),$$

where the expectation is taken w.r.t. the randomness of the algorithm.

From this theorem, by setting $\eta = \Theta(T^{-1/2})$ (ignoring factors in d, β, σ, r), we obtain regret bound of $\tilde{O}(\sqrt{T})$.[3] More precisely, if $r > 0$, we have the following regret bound:

[3]Note that, if the number T of rounds is sufficiently large and if the parameter η is of order $O(T^{-1/2})$, then the condition (17) is automatically satisfied.

Corollary 1. *If we set $\eta = \sqrt{(\beta\log T)/(\sigma T)}$, and if* (16)*,* (17)*, and $r > 0$ hold, for all $x^* \in \mathcal{K}'$, we have*

$$\mathbf{E}[R_T(x^*)] = O\left(d\sqrt{\frac{\beta T\log T}{\sigma}} + \frac{d^2\log d}{r^2\sigma}\right).$$

In addition, even if $r = 0$, e.g., even when $\mathcal{K}' = \mathcal{K}$, we have the following regret bound:

Corollary 2. *If we set $\eta = \sqrt{(\beta\log T)/(d\sigma T)}$, and if* (16) *and* (17) *hold, for all $x^* \in \mathcal{K}'$, we have*

$$\mathbf{E}[R_T(x^*)] = O\left(d^{3/2}\sqrt{\frac{\beta T\log T}{\sigma}}\right).$$

5.2 Proof of Theorem 1

To prove Theorem 1, we start by bounding the regret $R_T(x^*)$ by means of the smoothed objectives $\hat{f}_t$ defined by (10), on the basis of Lemmas 1 and 2.

Lemma 5. *For arbitrary $x^* \in \mathcal{K}'$, the regret for Algorithm 1 is bounded as*

$$\mathbf{E}[R_T(x^*)] \leq \sum_{t=1}^{T}\left(\mathbf{E}[\hat{f}_t(\hat{\mu}_t) - \hat{f}_t(x^*)] + \frac{\beta(\alpha_t^2+1)}{\eta\sigma t}\right).$$

Proof. Since f_t is a β-smooth function, we have

$$\begin{aligned}
&\mathbf{E}[f_t(a_t)] = \mathbf{E}[f_t(\hat{\mu}_t + \alpha_t B_t u_t)] \\
&\leq \mathbf{E}\left[f_t(\hat{\mu}_t) + \alpha_t \nabla f_t(\hat{\mu}_t)^\top B_t u_t + \frac{\beta}{2}\|\alpha_t B_t u_t\|_2^2\right] \\
&\leq \mathbf{E}\left[f_t(\hat{\mu}_t)\right] + \frac{\beta\alpha_t^2\|B_t\|_2^2}{2} \leq \mathbf{E}\left[\hat{f}_t(\hat{\mu}_t)\right] + \frac{\beta\alpha_t^2}{t\eta\sigma},
\end{aligned}$$

where the first inequality comes from (3), the second inequality follows from that $\mathbf{E}[u_t] = 0$ and that $\|u_t\|_2 = 1$, and the last inequality comes from (7) and (15). Similarly, from (7) and (15), we have

$$\begin{aligned}
-\mathbf{E}[f_t(x^*)] &\leq -\mathbf{E}[\hat{f}_t(x^*) + \beta\|B_t\|_2^2/2] \\
&\leq -\mathbf{E}[\hat{f}_t(x^*) + \beta/(t\eta\sigma)].
\end{aligned}$$

By combining the above two inequalities and taking the sum for $t \in [T]$, we obtain the bound for $\mathbf{E}[R_T(x^*)]$. □

We can provide a bound for the value $\sum_{t=1}^{T}(\hat{f}_t(x^*) - \hat{f}_t(\hat{\mu}_t))$ by combining an analysis for continuous multiplicative weight update methods [Arora et al., 2012, Hazan and Karnin, 2016], (12), Lemma 2, and assumption of (9).

Lemma 6. *Suppose that η satisfies* (17)*. For arbitrary $x^* \in \mathcal{K}'$ and arbitrary $\gamma > 0$, we have*

$$\begin{aligned}
&\sum_{t=1}^{T}\mathbf{E}[\hat{f}_t(\hat{\mu}_t) - \hat{f}_t(x^*)] \\
&\leq \frac{4d}{\eta}\log\frac{1}{\gamma} + 4\gamma T + \sum_{t=1}^{T}\left(\frac{4\eta d^2}{\alpha_t^2} + \frac{2\beta}{\eta\sigma t}\right) + 8.
\end{aligned}$$

A proof of this lemma is given in Section 8 in the supplementary material. We here give a rough sketch of the proof.

Proof sketch of Lemma 6. From a standard analysis of MWU [Arora et al., 2012], we have

$$\sum_{t=1}^{T} \mathbf{E}[\hat{f}_t(\hat{\mu}_t) - \hat{f}_t(x^*)] = O\left(\frac{d}{\eta}\log\frac{1}{\gamma} + \gamma T + \frac{1}{\eta}\sum_{t=1}^{T}\mathbf{E}\left[\log\int_{x\in\mathcal{K}'} p_t(x)\exp(-\eta\hat{g}_t^\top(x-\hat{\mu}_t))\mathrm{d}x\right]\right).$$

We provide a bound for the integral above, via separating the domain $\mathcal{K}'$ of integration into two sets $\{x \in \mathcal{K}' \mid -\eta\hat{g}_t^\top(x-\hat{\mu}_t) \le 1\}$ and $\{x \in \mathcal{K}' \mid -\eta\hat{g}_t^\top(x-\hat{\mu}_t) > 1\}$. The integral over the latter set can be bounded using Lemma 3 and the fact that the log-concavity of distributions is preserved under any linear transformation (see, e.g., [Saumard and Wellner, 2014]). The integral over the former set can be bounded by means of the inequality $\exp(y) \le 1 + y + y^2$ that holds for $y \le 1$, and (9), (5). Further, by applying $\log(1+w) \le w$, we have

$$\mathbf{E}\left[\log\int_{x\in\mathcal{K}'} p_t(x)\exp(-\eta\hat{g}_t^\top(x-\hat{\mu}_t))\mathrm{d}x\right] \lesssim \mathbf{E}\left[\int_{x\in\mathcal{K}'} p_t(x)(-\eta\hat{g}_t^\top(x-\hat{\mu}_t) + (\eta\hat{g}_t^\top(x-\hat{\mu}_t))^2)\mathrm{d}x\right],$$

ignoring $O(1/T)$ terms. Each term of the above can be bounded as follows:

$$\begin{aligned}
&\mathbf{E}\left[\int_{x\in\mathcal{K}'} p_t(x)(-\eta\hat{g}_t^\top(x-\hat{\mu}_t))\right] \\
&= \eta\,\mathbf{E}[\hat{g}_t^\top(\hat{\mu}_t-\mu_t)] = \eta\,\mathbf{E}[\nabla\hat{f}_t(\hat{\mu}_t)^\top(\hat{\mu}_t-\mu_t)] \\
&\le \eta\,\mathbf{E}[\nabla\hat{f}_t(\mu_t)^\top(\hat{\mu}_t-\mu_t)] + \eta\beta\|\hat{\mu}_t-\mu_t\|_2^2 \\
&= \eta\beta\|\hat{\mu}_t-\mu_t\|_2^2 \le \eta\beta\|\hat{\mu}_t-\mu_t\|_{\Sigma_t^{-1}}^2\|\Sigma_t\|_2 \le \frac{\beta}{\sigma t},
\end{aligned}$$

where the second equality comes from (12), the first inequality follows from (3) for β-smooth convex functions, the third equality follows from $\mathbf{E}[\hat{\mu}_t|\mu_t] = \mu_t$ in (9), and the last inequality follows from (9) and Lemma 2 and the $(\eta\sigma t)$-strong convexity of z_t. Similarly, we have

$$\begin{aligned}
&\mathbf{E}\left[\int_{x\in\mathcal{K}'} p_t(x)(\eta\hat{g}_t^\top(x-\hat{\mu}_t))^2\right] \\
&= \eta^2\,\mathbf{E}\left[\hat{g}_t^\top\left(\Sigma_t + (\mu_t-\hat{\mu}_t)(\mu_t-\hat{\mu}_t)^\top\right)\hat{g}_t\right] \\
&\le \frac{\eta^2 d^2}{\alpha_t^2}(\Sigma_t + (\mu_t-\hat{\mu}_t)(\mu_t-\hat{\mu}_t)^\top) \bullet (\mathbf{E}[B_t^{-1}u_t u_t^\top B_t^{-1}]) \\
&\le \frac{\eta^2 d}{\alpha_t^2}(\Sigma_t + (\mu_t-\hat{\mu}_t)(\mu_t-\hat{\mu}_t)^\top) \bullet \hat{\Sigma}_t^{-1} \le \frac{2\eta^2 d^2}{\alpha_t^2},
\end{aligned}$$

where the first inequality follows from (11) and the assumption that $|f_t(a_t)| \le 1$, and the last inequality follows from (9). Combining the above four displayed inequalities, we obtain Lemma 6. For the details on the proof, see Section 8 in the supplementary material. □

We shall complete the proof of Theorem 1 by means of Lemmas 5 and 6.

Proof of Theorem 1. By combining the above and Lemmas 5 and 6, and by setting $\gamma = \frac{1}{T}$, we obtain

$$\mathbf{E}[R_T(x^*)] \le 12 + \frac{4d}{\eta}\log T + \sum_{t=1}^{T}\left(\frac{4\eta d^2}{\alpha_t^2} + \frac{\beta(\alpha_t^2+3)}{\eta\sigma t}\right).$$

If we set α_t by (16), since $\alpha_t^2 \le d$ holds, we have

$$\sum_{t=1}^{T}\frac{\beta(\alpha_t^2+3)}{\eta\sigma t} \le \frac{4d\beta}{\eta}\sum_{t=1}^{T}\frac{1}{t} \le \frac{4d\beta\log(\mathrm{e}T)}{\eta}.$$

Combining the above two displayed inequalities and the fact that $1 \le \beta/\sigma$, we have

$$\mathbf{E}[R_T(x^*)] \le 12 + \frac{8d\beta\log(\mathrm{e}T)}{\eta\sigma} + 4\eta d^2\sum_{t=1}^{T}\frac{1}{\alpha_t^2}. \quad (18)$$

Let us consider bounding $\sum_{t=1}^{T} 1/\alpha_t^2$. From (16), we have

$$\sum_{t=1}^{T}\frac{1}{\alpha_t^2} = \sum_{t=1}^{T}\max\left\{\frac{2}{(\sqrt{2}/9 + r\sqrt{t\eta\sigma})^2}, \frac{1}{d}\right\} \le 81T, \quad (19)$$

where the inequality holds for arbitrary $r \ge 0$ (even if $r = 0$) since we have $1/\alpha_t^2 \le 81$. Assuming $r > 0$, we obtain a tiger bound for it; Denote $T' := \lfloor\frac{2d}{r^2\eta\sigma}\rfloor$. Since $2/(\sqrt{2}/9 + r\sqrt{t\eta\sigma})^2 \le 1/d$ holds for $t > T'$, we have

$$\begin{aligned}
\sum_{t=1}^{T}\frac{1}{\alpha_t^2} &\le \sum_{t=1}^{T'}\frac{2}{(\sqrt{2}/9 + r\sqrt{t\eta\sigma})^2} + \frac{\max\{T-T',0\}}{d} \\
&\le \frac{2}{r^2\eta\sigma}\sum_{t=1}^{T'}\frac{1}{t + 2/(81r^2\eta\sigma)} + \frac{T}{d} \\
&\le \frac{2}{r^2\eta\sigma}\log(1 + 81r^2\eta\sigma T'/2) + \frac{T}{d} \\
&\le \frac{2}{r^2\eta\sigma}\log(1+81d) + \frac{T}{d}, \quad (20)
\end{aligned}$$

where the third inequality follows from the fact that $\sum_{t=1}^{T'} 1/(t+y) \le \log(1 + T'/y)$ holds for any $y > 0$, and the last inequality follows from $T' \le \frac{2d}{r^2\eta\sigma}$. Combining (18), (19) and (20), we obtain

$$\begin{aligned}
\mathbf{E}[R_T(x^*)] &\le 12 + \frac{8d\beta\log(\mathrm{e}T)}{\eta\sigma} + 4\eta d^2\min\left\{81T, \frac{2}{r^2\eta\sigma}\log(1+81d) + \frac{T}{d}\right\} \\
&= O\left(\frac{d\beta\log T}{\eta\sigma} + \eta d T + \min\left\{\eta d^2 T, \frac{d^2\log d}{r^2\sigma}\right\}\right).
\end{aligned}$$

□

6 Discussion

We discuss the possibility of removing the assumption of $r > 0$, i.e., the assumption that the optimal solutions (or the benchmark set) are interiors, in Corollary 1.

From Lemmas 5 and 6, if we can set $\alpha_t = \tilde{\Theta}(\sqrt{d})$, we have

$$\mathbf{E}[R_T(x^*)] = \tilde{O}\left(\frac{d\beta}{\sigma\eta} + \eta d T\right) = \tilde{O}\left(d\sqrt{\frac{\beta T}{\sigma}}\right) \quad (21)$$

by setting $\eta = \tilde{\Theta}(\sqrt{\frac{\beta}{\sigma T}})$. Setting $\alpha_t = \tilde{\Theta}(\sqrt{d})$, however, may cause an infeasible action $a_t = \hat{\mu}_t + \alpha_t B_t u_t \notin \mathcal{K}$ in Algorithm 1 without the assumption on r, and consequently, may make it impossible to bound the regret. The possibility of $a_t \notin \mathcal{K}$ seems to be, on the other hand, quite small if we set $\alpha_t = c\sqrt{d}$ with a small constant $c > 0$, even when $\mathcal{K}' = \mathcal{K}$. Under the assumption that the possibility of $a_t \notin \mathcal{K}$ is sufficiently small, our analysis in Section 5 works similarly and leads to a regret bound of $\tilde{O}(d\sqrt{T})$. A sufficient condition for this assumption of small possibilities can be formulated as follows:

Conjecture 1. *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a d-dimensional convex set. Let p be a multidimensional truncated normal distribution over $\mathcal{K}$, i.e., $p(x) \propto \exp(-\|x\|_2^2/2)$ for $x \in \mathcal{K}$ and $p(x) = 0$ for $x \in \mathbb{R}^d \setminus \mathcal{K}$. Let $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ be the mean and the covariance matrix of p, respectively. Then, for a random variable u following a uniform distribution $U(\mathbb{S}^d)$ over the unit sphere $\mathbb{S}^d$, the probability of that $\mu + \alpha\Sigma^{1/2}u$ is not in $\mathcal{K}$ is bounded as*

$$\begin{aligned}&\Pr_{u\sim U(\mathbb{S}^d)}\left[\mu + \alpha\Sigma^{1/2}u \notin \mathcal{K}\right] \\ &\quad\le \exp\left((1+\log d)^{O(1)} - \frac{\sqrt{d}}{\alpha\cdot(1+\log d)^{O(1)}}\right)\end{aligned}$$

for all $\alpha > 0$.

If this conjecture holds, we have a regret bound of $\mathbf{E}[R_T(x^*)] = \tilde{O}(d\sqrt{T})$ for arbitrary $x^* \in \mathcal{K}$, without the assumption of interior optimal solutions, i.e., the bound holds even for the case of $\mathcal{K}' = \mathcal{K}$.

Since a truncated normal distribution is a log-concave distribution, in Conjecture 1, the probability of $\mu + \alpha\Sigma^{1/2}u \notin \mathcal{K}$ is equal to zero for $\alpha < 1/\mathrm{e}$, from Lemma 4. This fact is used to prove Proposition 1. The question is if we can obtain a bound of the probability for $\alpha = \tilde{\Omega}(\sqrt{d})$.

7 Conclusion

This paper considered bandit convex optimization problems with strongly-convex and smooth objectives. We provided an algorithm with tight regret bounds, w.r.t. the number T of rounds as well as the dimension d of the feasible region, under milder assumptions than existing works. More precisely, we gave a regret bound of $\tilde{O}(d\sqrt{T}+d^2/r^2)$ under the assumption that the optimal solutions (or the benchmark set) are r-interiors, and without this assumption, our algorithm achieves a regret of $\tilde{O}(d^{3/2}\sqrt{T})$. Our algorithm, further, works given only a membership oracle for the feasible region, without self-concordant barriers.

The assumption of interior optimal solutions, however, might be abundant, and the tight regret bounds might apply to more general problem settings. To prove that our algorithm achieves $\tilde{O}(d\sqrt{T})$-regret without the assumption, we introduced an approach based on the inequality for probability regarding log-concave distributions (more precisely, multi-dimensional truncated normal distributions) in Conjecture 1. We leave it as a feature work to prove this conjecture.

Another future direction is to pursue tighter regret bounds for general BCO without assumptions of strong convexity and smoothness of objective functions. For this more general problem, there is a larger gap w.r.t. d between the upper bound of $\tilde{O}(d^{9.5}\sqrt{T})$ and the lower bound of $\Omega(d\sqrt{T})$. If a lower bound of $\Omega(d^{3/2}\sqrt{T})$ would hold as conjectured in [Bubeck et al., 2017], it would, together with our results, imply a $\tilde{\Omega}(\sqrt{d})$ gap between minimax regrets for BCO with strongly-convex and smooth losses and that for general BCO.

Acknowledgment

We would like to thank the anonymous reviewers for their insightful comments and suggestion, which help improve the manuscript. This work was supported by JST, ACT-I, Grant Number JPMJPR18U, Japan.

References

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

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

A. Belloni, T. Liang, H. Narayanan, and A. Rakhlin. Escaping the local minima via simulated annealing: Optimization of approximately convex functions. In *Conference on Learning Theory*, pages 240–265, 2015.

G. W. Brown. Iterative solution of games by fictitious play. *Activity Analysis of Production and Allocation*, 13 (1):374–376, 1951.

G. W. Brown and J. Von Neumann. Solutions of games by differential equations. *Ann. Math. Studies*, 24:73–79, 1950.

S. Bubeck and R. Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. In *Conference on Learning Theory*, pages 279–279, 2015.

S. Bubeck and R. Eldan. Multi-scale exploration of convex functions and bandit convex optimization. In *Conference on Learning Theory*, pages 583–589, 2016.

S. Bubeck, N. Cesa-Bianchi, and S. Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Conference on Learning Theory*, pages 41.1–41.14, 2012.

S. Bubeck, O. Dekel, T. Koren, and Y. Peres. Bandit convex optimization: $\sqrt{T}$ regret in one dimension. In *Conference on Learning Theory*, pages 266–278, 2015.

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

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge university press, 2006.

N. Cesa-Bianchi and G. Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

L. Chen, M. Zhang, and A. Karbasi. Projection-free bandit convex optimization. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2047–2056, 2019.

T. M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

V. Dani, S. M. Kakade, and T. P. Hayes. The price of bandit information for online optimization. In *Advances in Neural Information Processing Systems*, pages 345–352, 2008.

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

A. D. Flaxman, A. T. Kalai, and H. B. McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete algorithms*, pages 385–394. Society for Industrial and Applied Mathematics, 2005.

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

E. Hazan and Z. Karnin. Volumetric spanners: an efficient exploration basis for learning. *The Journal of Machine Learning Research*, 17(1):4062–4095, 2016.

E. Hazan and K. Levy. Bandit convex optimization: Towards tight bounds. In *Advances in Neural Information Processing Systems*, pages 784–792, 2014.

E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

E. Hazan, W. Hu, Y. Li, and Z. Li. Online improper learning with an approximation oracle. In *Advances in Neural Information Processing Systems*, pages 5652–5660, 2018.

D. Hoeven, T. Erven, and W. Kotłowski. The many faces of exponential weights in online learning. In *Conference on Learning Theory*, pages 2067–2092, 2018.

X. Hu, L. Prashanth, A. György, and C. Szepesvári. (bandit) convex optimization with biased noisy gradient oracles. In *Artificial Intelligence and Statistics*, pages 819–828, 2016.

S. Ito, D. Hatano, H. Sumita, K. Takemura, T. Fukunaga, N. Kakimura, and K.-I. Kawarabayashi. Oracle-efficient algorithms for online linear optimization with bandit feedback. In *Advances in Neural Information Processing Systems*, pages 10590–10599, 2019.

R. D. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems*, pages 697–704, 2005.

W. Kumagai. Regret analysis for continuous dueling bandit. In *Advances in Neural Information Processing Systems*, pages 1489–1498, 2017.

L. Lovász and S. Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms*, 30(3):307–358, 2007.

M. Mohri and S. Yang. Adaptive algorithms and data-dependent guarantees for bandit convex optimization. In *The 32nd Conference on Uncertainty in Artificial Intelligence 2016*, pages 815–824, 2016.

H. Narayanan and A. Rakhlin. Efficient sampling from time-varying log-concave distributions. *The Journal of Machine Learning Research*, 18(1):4017–4045, 2017.

Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, 1994.

Y. Nesterov and V. Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017.

E. Ordentlich and T. M. Cover. The cost of achieving the best portfolio in hindsight. *Mathematics of Operations Research*, 23(4):960–982, 1998.

J. Robinson. An iterative method of solving a game. *Annals of Mathematics*, pages 296–301, 1951.

A. Saumard and J. A. Wellner. Log-concavity and strong log-concavity: a review. *Statistics Surveys*, 8:45, 2014.

A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1998.

S. Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

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

M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 928–936, 2003.