

# Adaptive Bandit Convex Optimization with Heterogeneous Curvature

**Haipeng Luo\***

*University of Southern California*

HAIPENGL@USC.EDU

**Mengxiao Zhang\***

*University of Southern California*

ZHAN147@USC.EDU

**Peng Zhao\***

*National Key Laboratory for Novel Software Technology, Nanjing University*

ZHAOP@LAMDA.NJU.EDU.CN

**Editors:** Po-Ling Loh and Maxim Raginsky

## Abstract

We consider the problem of adversarial bandit convex optimization, that is, online learning over a sequence of arbitrary convex loss functions with only one function evaluation for each of them. While all previous works assume known and homogeneous curvature on these loss functions, we study a heterogeneous setting where each function has its own curvature that is only revealed after the learner makes a decision. We develop an efficient algorithm that is able to adapt to the curvature on the fly. Specifically, our algorithm not only recovers or *even improves* existing results for several homogeneous settings, but also leads to surprising results for some heterogeneous settings — for example, while Hazan and Levy (2014) showed that  $\tilde{O}(d^{3/2}\sqrt{T})$  regret is achievable for a sequence of  $T$  smooth and strongly convex  $d$ -dimensional functions, our algorithm reveals that the same is achievable even if  $T^{3/4}$  of them are not strongly convex, and sometimes even if a constant fraction of them are not strongly convex. Our approach is inspired by the framework of Bartlett et al. (2007) who studied a similar heterogeneous setting but with stronger gradient feedback. Extending their framework to the bandit feedback setting requires novel ideas such as lifting the feasible domain and using a logarithmically homogeneous self-concordant barrier regularizer.

## 1. Introduction

We consider the problem of adversarial bandit convex optimization, formulated as the following sequential learning process of  $T$  rounds. At the beginning, knowing the learner’s algorithm, an adversary decides an arbitrary sequence of  $T$  convex loss functions  $f_1, \dots, f_T : \mathcal{X} \mapsto \mathbb{R}$  over some convex domain  $\mathcal{X} \subset \mathbb{R}^d$ . Then, at each round  $t$ , the learner is required to select a point  $x_t \in \mathcal{X}$ , and afterwards observes and suffers her loss  $f_t(x_t)$ . The performance of the learner is measured by her regret, the difference between the her total suffered loss and that of the best fixed point in hindsight.

Without further assumption, Bubeck et al. (2021) achieve  $\tilde{O}(d^{10.5}\sqrt{T})$  regret with large computational complexity of  $\tilde{O}(\text{poly}(d)T)$  per round. Another work by Lattimore (2020) improves the regret to  $\tilde{O}(d^{2.5}\sqrt{T})$  but with even larger computational complexity that is exponential in  $d$  and  $T$  per round. On the other hand, the current best lower bound is  $\Omega(d\sqrt{T})$  (Dani et al., 2008), exhibiting a large gap in the  $d$  dependency. It has been shown that, however, curvature of the loss functions helps — for example, when the functions are all smooth and strongly convex, Hazan and Levy (2014) develop a simple and efficient Follow-the-Regularized-Leader (FTRL) type algorithm with  $\tilde{O}(d^{3/2}\sqrt{T})$  regret; even when the functions are only smooth, Saha and Tewari (2011) show that

---

\* Authors are listed in alphabetical order.

Table 1: A summary of our results for bandit convex optimization over  $T$  smooth  $d$ -dimensional functions, the  $t$ -th of which is  $\sigma_t$ -strongly convex.  $\mathcal{T} \subset [T]$  is a subset of rounds with no strong convexity. The dependency on parameters other than  $d$  and  $T$  can be found in the respective corollary (see also [Footnote 1](#)). Note that our results are all achieved by one single adaptive algorithm.

Strong Convexity $\{\sigma_t\}_{t=1}^T$	Previous Works	Our Results ( <a href="#">Algorithm 1</a> )
$\sigma_t = 0, \forall t \in [T]$	$\tilde{\mathcal{O}}(dT^{2/3})$ ( <a href="#">Saha and Tewari, 2011</a> )	$\tilde{\mathcal{O}}(d^{2/3}T^{2/3})$ ( <a href="#">Corollary 8</a> )
$\sigma_t = \sigma > 0, \forall t \in [T]$	$\tilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ ( <a href="#">Hazan and Levy, 2014</a> )	$\tilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ ( <a href="#">Corollary 9</a> )
$\sigma_t = \sigma \mathbb{1}\{t \notin \mathcal{T}\},$ $ \mathcal{T}  = T^{3/4}$ or $\mathcal{T} = [T/2, T]$	N/A	$\tilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ ( <a href="#">Corollary 10</a> )
$\sigma_t = t^{-\alpha}, \forall t \in [T]$	N/A	$\begin{cases} \tilde{\mathcal{O}}(d^{3/2}T^{(1+\alpha)/2}), & \alpha \in [0, 1/3) \\ \tilde{\mathcal{O}}(d^{2/3}T^{2/3}), & \alpha \in [1/3, 1] \end{cases}$ ( <a href="#">Corollary 11</a> )

$\tilde{\mathcal{O}}(dT^{2/3})$  regret is achievable again via a simple and efficient FTRL variant, despite the suboptimal dependency in  $T$ .

However, all such existing results making use of curvature assume a homogeneous setting, that is, all loss functions share the same curvature parameters that are known ahead of time. Ignoring the ubiquitous heterogeneity in online data is either unrealistic or forcing one to use a conservative curvature parameter (e.g., the smallest strong convexity parameter among all functions), while intuitively, being able to exploit and adapt to the individual curvature of each loss function should result in much better performance.

Motivated by this fact and inspired by the work of [Bartlett et al. \(2007\)](#) who consider a heterogeneous setting in Online Convex Optimization with stronger gradient feedback, we study a similar setting where each  $f_t$  has its own strong convexity  $\sigma_t \geq 0$ , revealed only at the end of round  $t$  after the learner decides  $x_t$ . We provide examples in [Section 2](#) to illustrate why this is a realistic setup even in the bandit setting where traditionally only  $f_t(x_t)$  is revealed. In this setting, we develop efficient algorithms that automatically adapt to the heterogeneous curvature and enjoy strong adaptive regret bounds. These bounds not only recover or *even improve* the existing results in the homogeneous setting, but also reveal interesting new findings in some hybrid scenarios. More specifically, our results are as follows (for simplicity, only the dependency on  $d$  and  $T$  is shown; see respective sections for the complete bounds).

- We start with the case where all loss functions are smooth in [Section 3](#). Our algorithm achieves  $\tilde{\mathcal{O}}(d^{2/3}T^{2/3})$  regret if no functions are strongly convex (that is,  $\sigma_t = 0$  for all  $t$ ), *improving* the  $\tilde{\mathcal{O}}(dT^{2/3})$  bound of ([Saha and Tewari, 2011](#)), and  $\tilde{\mathcal{O}}(d^{3/2}\sqrt{T})$  regret if all functions happen to be  $\sigma$ -strongly convex (that is,  $\sigma_t = \sigma$  for all  $t$ ), matching that from ([Hazan and Levy, 2014](#)). In fact, our algorithm achieves the latter result *even if*  $\Theta(T^{3/4})$  of the functions have no strong convexity (and sometimes even if a constant fraction of the functions have no strong convexity). More generally, our bound interpolates between these two extremes. For example, if  $\sigma_t$  is decaying at the rate of  $1/t^\alpha$  for some  $\alpha \in [0, 1]$ , then the regret is  $\tilde{\mathcal{O}}(d^{3/2}T^{(1+\alpha)/2})$  when  $\alpha \leq 1/3$ , and  $\tilde{\mathcal{O}}(d^{2/3}T^{2/3})$  otherwise.<sup>1</sup> See [Table 1](#) for a summary. We note that the improvement over ([Saha and Tewari, 2011](#)) comes as a side product of the better regularization technique of our algorithm.

1. This is a simplified and loosen version of [Corollary 11](#), which explains the discontinuity in  $\alpha$ .

Table 2: A summary of our results for bandit convex optimization over  $T$  Lipschitz  $d$ -dimensional functions, the  $t$ -th of which is  $\sigma_t$ -strongly convex.  $\mathcal{T} \subset [T]$  is a subset of rounds with no strong convexity. The dependency on parameters other than  $d$  and  $T$  can be found in the respective corollary. Note that our results are all achieved by one single adaptive algorithm.

Strong Convexity $\{\sigma_t\}_{t=1}^T$	Previous Works	Our Results (Algorithm 2)
$\sigma_t = 0, \forall t \in [T]$	$\tilde{O}(d^{3/4}T^{3/4})$ (Luo, 2017)	$\tilde{O}(\sqrt{dT}^{3/4})$ (Corollary 13)
$\sigma_t = \sigma > 0, \forall t \in [T]$	$\tilde{O}(d^{4/3}T^{2/3})$ (Agarwal et al., 2010)	$\tilde{O}(d^{2/3}T^{2/3})$ (Corollary 14)
$\sigma_t = \sigma \mathbb{1}\{t \notin \mathcal{T}\},$ $ \mathcal{T}  = T^{8/9}$ or $\mathcal{T} = [T/2, T]$	N/A	$\tilde{O}(d^{2/3}T^{2/3})$ (Corollary 15)
$\sigma_t = t^{-\alpha}, \forall t \in [T]$	N/A	$\begin{cases} \tilde{O}(d^{2/3}T^{(2+\alpha)/3}), & \alpha \in [0, 1/4) \\ \tilde{O}(\sqrt{dT}^{3/4}), & \alpha \in [1/4, 1] \end{cases}$ (Corollary 16)

- We then consider another scenario where all loss functions are Lipschitz in Section 4. We develop another algorithm that achieves  $\tilde{O}(\sqrt{dT}^{3/4})$  regret when no functions are strongly convex, improving the  $\tilde{O}(d^{3/4}T^{3/4})$  bound of (Luo, 2017), and  $\tilde{O}(d^{2/3}T^{2/3})$  regret when all functions are  $\sigma$ -strongly convex, improving the  $\tilde{O}(d^{4/3}T^{2/3})$  bound of (Agarwal et al., 2010). These improvements again come as a side product of our better regularization. Similarly, the  $\tilde{O}(d^{2/3}T^{2/3})$  result holds even if  $\Theta(T^{8/9})$  of the functions (or sometimes a constant fraction of them) have no strong convexity. For similar intermediate bounds in the example when  $\sigma_t = 1/t^\alpha$ , see Table 2.

**Techniques.** Our algorithm is also a variant of FTRL, with two crucial new ingredients to handle heterogeneous curvature. First, we extend the idea of Bartlett et al. (2007) to adaptively add  $\ell_2$  regularization to the loss functions and adaptively tune the learning rate. Doing so in the bandit setting is highly nontrivial and requires our second technical ingredient, which is to lift the problem to the  $(d+1)$ -dimensional space and then apply a *logarithmically homogeneous self-concordant barrier* in the FTRL update. This technique is inspired by a recent work of Lee et al. (2020) on achieving high-probability regret bounds for adversarial linear bandits, but the extension from linear bandits to convex bandits is nontrivial. In fact, the purpose of using this technique is also different: they need to bound the variance of the learner’s loss, which is related to bounding  $x^\top \nabla^2 \psi(x) x$  for some regularizer  $\psi$ , while we need to bound the stability of the algorithm, which is related to bounding  $\nabla \psi(x) \nabla^{-2} \psi(x) \nabla \psi(x)$ , but it turns out that when  $\psi$  is a logarithmically homogeneous  $\nu$ -self-concordant barrier, then these two quantities are *exactly the same* and bounded by  $\nu$ .

**Related work.** Bandit convex optimization has been extensively studied under different loss function structures, including Lipschitz functions (Kleinberg, 2004; Flaxman et al., 2005), linear functions (Abernethy et al., 2008, 2012; Bubeck et al., 2012), smooth functions (Saha and Tewari, 2011), strongly convex functions (Agarwal et al., 2010), smooth and strongly convex functions (Hazan and Levy, 2014; Ito, 2020), quadratic functions (Suggala et al., 2021), pseudo-1-dimensional functions (Saha et al., 2021), and others. Without any structure (other than convexity), a series of progress has been made over recent years (Bubeck et al., 2015; Hazan and Li, 2016; Bubeck and Eldan, 2016; Bubeck et al., 2021), but as mentioned, even the best result (Bubeck et al., 2021) has a large dependency on  $d$  in the regret and is achieved by an impractical algorithm with large computational

complexity. Our comparisons in this work (such as those in [Table 1](#) and [Table 2](#)) thus mainly focus on more efficient and practical methods in the literature that share the same FTRL framework.<sup>2</sup>

Closest to our heterogeneous setting is the work on Online Convex Optimization by [Bartlett et al. \(2007\)](#), where at the end of each round,  $\sigma_t$  and  $\nabla f_t(x_t)$  are revealed (versus  $\sigma_t$  and  $f_t(x_t)$  in our setting). Due to the stronger feedback, their algorithm achieves  $\mathcal{O}(\sqrt{T})$  regret without any strong convexity,  $\mathcal{O}(\log T)$  regret if all functions are strongly convex, and generally something in between. Our results are in the same vein, and as mentioned, our algorithm is also heavily inspired by theirs.

Another potential approach to adapting to different environments is to have a meta algorithm learning over a set of base algorithms, each dedicated to a specific environment. Doing so in the bandit setting, however, is highly challenging ([Agarwal et al., 2017](#)) or even impossible sometimes ([Marinov and Zimmert, 2021](#)). For example, even if one only aims to adapt to two environments, one with only smooth functions and the other with smooth and strongly convex functions, the approach of [Agarwal et al. \(2017\)](#) is only able to achieve  $\tilde{\mathcal{O}}(T^{3/4})$  regret for the first environment if one insists to enjoy  $\tilde{\mathcal{O}}(\sqrt{T})$  regret in the second one.

## 2. Preliminaries and Problem Setup

We start by reviewing some basic definitions.

**Definition 1** *We say that a differentiable function  $f : \mathcal{X} \mapsto \mathbb{R}$  is  $\beta$ -smooth over the feasible set  $\mathcal{X}$  if for any  $x, y \in \mathcal{X}$ ,  $\|\nabla f(x) - \nabla f(y)\|_2 \leq \beta \|x - y\|_2$  holds.*

**Definition 2** *We say that a function  $f : \mathcal{X} \mapsto \mathbb{R}$  is  $L$ -Lipschitz over the feasible set  $\mathcal{X}$  if for any  $x, y \in \mathcal{X}$ ,  $|f(x) - f(y)| \leq L \|x - y\|_2$  holds.*

**Definition 3** *We say that a differentiable function  $f : \mathcal{X} \mapsto \mathbb{R}$  is  $\sigma$ -strongly convex over the feasible set  $\mathcal{X}$  if for any  $x, y \in \mathcal{X}$ ,  $f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\sigma}{2} \|x - y\|_2^2$  holds.*

**Problem setup.** Bandit Convex Optimization (BCO) can be modeled as a  $T$ -round games between a learner and an oblivious adversary. Before the game starts, the adversary (knowing the learner’s algorithm) secretly decides an arbitrary sequence of convex functions  $f_1, \dots, f_T : \mathcal{X} \mapsto \mathbb{R}$ , where  $\mathcal{X} \subset \mathbb{R}^d$  is a known compact convex domain. In [Section 3](#) we assume that all  $f_t$ ’s are  $\beta$ -smooth for some known parameter  $\beta \geq 0$ , while in [Section 4](#) we assume that they are all  $L$ -Lipschitz for some known parameter  $L \geq 0$ . In both cases, we denote by  $\sigma_t \geq 0$  the strong convexity parameter of  $f_t$ , initially *unknown* to the learner (note that  $\sigma_t$  could be zero).

At each round  $t \in [T] \triangleq \{1, \dots, T\}$  of the game, the learner chooses an action  $x_t \in \mathcal{X}$  based on all previous observations, and subsequently suffers loss  $f_t(x_t)$ . The adversary then reveals both  $f_t(x_t)$  and  $\sigma_t$  to the learner. Compared to previous works where  $\sigma_t$  is the same for all  $t$  and known to the learner ahead of time, our setting is clearly more suitable for applications with heterogeneous curvature. We provide such an example below, which also illustrates why it is reasonable for the learner to observe  $\sigma_t$  at the end of round  $t$ .

---

2. When the functions are smooth only, our comparison is based on ([Saha and Tewari, 2011](#)), instead of the seemingly better results of prior works ([Dekel et al., 2015](#); [Yang and Mohri, 2016](#)), because the latter ones are unfortunately wrong as pointed out in ([Hu et al., 2016](#)).

**Examples.** Consider a problem where  $f_t(x) = \sum_{i=1}^N g_{t,i}(c_{t,i}^\top x)$ . Here,  $N$  is the number of users in each round,  $c_{t,i}$  represents some context of the  $i$ -th user in round  $t$ , the learner's decision  $x$  is used to make a linear prediction  $c_{t,i}^\top x$  for this user, and her loss is evaluated via a convex function  $g_{t,i}$  which incorporates some ground truth for this user (e.g., labels in the case of classification where  $g_{t,i}$  could be the logistic loss, or responses in the case of regression where  $g_{t,i}$  could be the squared loss). Note that the (heterogeneous) strong convexity  $\sigma_t$  of  $f_t$  is at least  $\mu \lambda_{\min}(\sum_{i=1}^N c_{t,i} c_{t,i}^\top)$  where  $\lambda_{\min}(\cdot)$  denotes the minimum eigenvalue of a matrix, and  $\mu$  is a lower bound on the second derivative of  $g_{t,i}$ , often known ahead of time assuming some natural boundedness of  $x$  and  $c_{t,i}$ .

In this setup, there are several situations where our feedback model is reasonable. For example, it might be the case that the loss  $f_t(x_t)$  as well as the context  $c_{t,i}$  is visible to the learner, but the ground truth (and thus  $g_{t,i}$ ) is not. In this case, based on the earlier bound on the strong convexity, the learner can calculate  $\sigma_t$  herself. As another example, due to privacy consideration, the user's context  $c_{t,i}$  might not be revealed to the learner, but it is acceptable to reveal a single number  $\lambda_{\min}(\sum_{i=1}^N c_{t,i} c_{t,i}^\top)$  summarizing this batch of users. Clearly,  $\sigma_t$  can also be calculated by the learner in this case.

**Objective and simplifying assumptions.** The objective of the learner is to minimize her (expected) *regret*, defined as

$$\text{REG} = \mathbb{E} \left[ \sum_{t=1}^T f_t(x_t) \right] - \min_{x \in \mathcal{X}} \mathbb{E} \left[ \sum_{t=1}^T f_t(x) \right], \quad (1)$$

which is the difference between the expected loss suffered by the learner and that of the best fixed action (the expectation is with respect to the randomness of the learner). Without loss of generality, we assume  $\max_{x \in \mathcal{X}} |f_t(x)| \leq 1$  for all  $t \in [T]$ ,  $\mathcal{X}$  contains the origin, and  $\max_{x \in \mathcal{X}} \|x\|_2 = 1$ .<sup>3</sup>

**Notations.** We adopt the following notational convention throughout the paper. Generally, we use lowercase letters to denote vectors and capitalized letters to denote matrices. For a positive semi-definite matrix  $M \in \mathbb{R}^{d \times d}$  and a vector  $x \in \mathbb{R}^d$ ,  $\|x\|_2$  represents the standard Euclidean norm of  $x$  and  $\|x\|_M \triangleq \sqrt{x^\top M x}$  represents the Mahalanobis norm induced by  $M$ .  $\mathbb{S}^d$  and  $\mathbb{B}^d$  respectively denote the unit sphere and unit ball in  $\mathbb{R}^d$ ,  $\mathbf{1}$  and  $\mathbf{0}$  respectively denote the all-one and all-zero vectors with an appropriate dimension, and  $I$  denotes the identity matrix with an appropriate dimension. For a differentiable convex function  $\psi$ , define the corresponding Bregman divergence as  $D_\psi(u, v) = \psi(u) - \psi(v) - \langle \nabla \psi(v), u - v \rangle$ . For a vector  $x \in \mathbb{R}^d$ , we use  $x_{[i:j]} \in \mathbb{R}^{j-i+1}$  to denote the induced truncated vector consisting of the  $i$ -th to  $j$ -th coordinates of  $x$ . For a sequence of scalars  $a_1, \dots, a_t$ , we use  $a_{i:j} \triangleq \sum_{k=i}^j a_k$  to denote the cumulative summation, and  $\{a_s\}_{s=p}^q$  to denote the subsequence  $a_p, a_{p+1}, \dots, a_q$ . The  $\tilde{\mathcal{O}}(\cdot)$  notation omits the logarithmic dependence on the horizon  $T$ .<sup>4</sup>  $\mathbb{E}_t[\cdot]$  is a shorthand for the conditional expectation given the history before round  $t$ .

3. This is without loss of generality because for any problem with  $|f_t(x)| \leq B$  and  $\max_{x, x' \in \mathcal{X}} \|x - x'\|_2 = D$ , we can solve it via solving a modified problem with convex domain  $\mathcal{U} = \{u = \frac{1}{D}(x - x_0) : x \in \mathcal{X}\}$  for any fixed  $x_0 \in \mathcal{X}$  and loss functions  $g_t(u) = \frac{1}{B} f_t(Du + x_0)$ , which then satisfies our simplifying assumptions  $\max_{u \in \mathcal{U}} |g_t(u)| \leq 1$ ,  $\mathbf{0} \in \mathcal{U}$ , and  $\max_{u \in \mathcal{U}} \|u\|_2 = 1$ .

4. In the texts, for simplicity  $\tilde{\mathcal{O}}(\cdot)$  might also hide dependency on other parameters such as the dimension  $d$ . However, in all formal theorem/lemma statements, this will not be the case.

### 3. Smooth Bandit Convex Optimization with Heterogeneous Strong Convexity

Throughout this section, we assume that all loss functions  $f_1, \dots, f_T$  are  $\beta$ -smooth (see [Definition 1](#)). To present our adaptive algorithm in this case, we start by reviewing two important existing algorithms upon which ours is built.

**Review of Adaptive Online Gradient Descent (AOGD).** As mentioned, the AOGD algorithm of [Bartlett et al. \(2007\)](#) is designed for a similar heterogeneous setting but with the stronger gradient feedback. The first key idea of AOGD is that, instead of learning over the original loss functions  $\{f_t\}_{t=1}^T$ , one should learn over their  $\ell_2$ -regularized version  $\{\tilde{f}_t\}_{t=1}^T$  where  $\tilde{f}_t(x) = f_t(x) + \frac{\lambda_t}{2} \|x\|_2^2$  for some coefficient  $\lambda_t$ . Intuitively,  $\lambda_t$  is large when previous loss functions exhibit not enough strong convexity to stabilize the algorithm, and small otherwise. How to exactly tune  $\lambda_t$  based on the observed  $\{\sigma_s\}_{s=1}^t$  is their second key idea — they show that  $\lambda_t$  should balance two terms and satisfy

$$\frac{3}{2}\lambda_t = \frac{1}{\sigma_{1:t} + \lambda_{1:t}}, \quad (2)$$

which results in a quadratic equation of  $\lambda_t$  and can be solved in closed form. The final component of AOGD is simply to run gradient descent on  $\{\tilde{f}_t\}_{t=1}^T$  with some adaptively decreasing learning rates.

**Review of BCO with smoothness and strong convexity.** [Hazan and Levy \(2014\)](#) consider the BCO problem with  $\beta$ -smooth and  $\sigma$ -strongly convex loss functions. Their FTRL-based algorithm maintains an auxiliary sequence  $y_1, \dots, y_T \in \mathcal{X}$  via

$$y_{t+1} = \operatorname{argmin}_{x \in \mathcal{X}} \left\{ \sum_{s=1}^t \left( \langle g_s, x \rangle + \frac{\sigma}{2} \|x - y_s\|_2^2 \right) + \frac{1}{\eta} \psi(x) \right\}, \quad (3)$$

where  $g_s$  is some estimator of  $\nabla f_s(y_s)$ ,  $\eta > 0$  is some fixed learning rate, and the regularizer  $\psi$  is a  $\nu$ -self-concordant barrier whose usage in BCO is pioneered by [Abernethy et al. \(2008\)](#) (see [Appendix C](#) for definition). The rational behind the squared distance terms in this update is that, due to strong convexity, we have  $f_s(y_s) - f_s(x) \leq \tilde{g}_s(y_s) - \tilde{g}_s(x)$  for any  $x$  and  $\tilde{g}_s(x) = \nabla f_s(y_s)^\top x + \frac{\sigma}{2} \|x - y_s\|_2^2$ , meaning that it suffices to consider  $\tilde{g}_1, \dots, \tilde{g}_T$  as the loss functions. Having  $y_t$ , the algorithm makes the final prediction  $x_t$  by adding certain curvature-adaptive and shrinking exploration to  $y_t$ :  $x_t = y_t + H_t^{-1/2} u_t$ , where  $H_t = \nabla^2 \psi(y_t) + \eta \sigma t I$  and  $u_t$  is chosen from the unit sphere  $\mathbb{S}^d$  uniformly at random ( $x_t \in \mathcal{X}$  is guaranteed by the property of self-concordant barriers). Finally, with the feedback  $f_t(x_t)$ , the gradient estimator is constructed as  $g_t = d \cdot f_t(x_t) H_t^{1/2} u_t$ , which can be shown to be an unbiased and low-variance estimator of the gradient of some smoothed version of  $f_t$  at  $y_t$ .

#### 3.1. Proposed Algorithm and Main Theorem

We are now ready to describe our algorithm. Following ([Bartlett et al., 2007](#)), our first step is also to consider learning over the  $\ell_2$ -regularized loss functions:  $\tilde{f}_t(x) = f_t(x) + \frac{\lambda_t}{2} \|x\|_2^2$  with an adaptively chosen  $\lambda_t > 0$  (note that with the bandit feedback  $f_t(x_t)$ , we can also evaluate  $\tilde{f}_t(x_t)$ ). While [Bartlett et al. \(2007\)](#) apply gradient descent, the standard and optimal algorithm for strongly-convex losses with gradient feedback, here we naturally apply the algorithm of [Hazan and Levy \(2014\)](#) to this sequence of regularized loss functions instead. Since  $\tilde{f}_t$  is  $(\sigma_t + \lambda_t)$ -strongly convex, following

Eq. (3) and adopting a decreasing learning rate  $\eta_t$  shows that we should maintain the auxiliary sequence  $y_1, \dots, y_T$  according to

$$y_{t+1} = \operatorname{argmin}_{x \in \mathcal{X}} \left\{ \sum_{s=1}^t \left( \langle g_s, x \rangle + \frac{\sigma_s + \lambda_s}{2} \|x - y_s\|_2^2 \right) + \frac{1}{\eta_{t+1}} \psi(x) \right\}, \quad (4)$$

where similarly  $g_t = d \cdot \tilde{f}_t(x_t) H_t^{1/2} u_t$  for  $x_t = y_t + H_t^{-1/2} u_t$ ,  $H_t = \nabla^2 \psi(y_t) + \eta_t (\sigma_{1:t-1} + \lambda_{1:t-1}) \mathbf{I}$ , and  $u_t$  chosen randomly from the unit sphere.

While this forms a natural and basic framework of our algorithm, there is in fact a critical issue when analyzing such a barrier-regularized FTRL algorithm due to the decreasing learning rate, which was never encountered in the literature as far as we know since all related works using this framework adopt a fixed learning rate (see e.g., (Abernethy et al., 2008; Saha and Tewari, 2011; Hazan and Kale, 2011; Rakhlin and Sridharan, 2013; Hazan and Levy, 2014; Bubeck et al., 2019)) or an increasing learning rate (Lee et al., 2020). More specifically, to bound the stability  $\langle y_t - y_{t+1}, g_t \rangle$  of the algorithm, all analysis for barrier-regularized FTRL implicitly or explicitly requires bounding the *Newton decrement*, which in our context is  $\|\nabla G_t(y_t)\|_{\nabla^{-2}G_t(y_t)}^2$  for  $G_t$  being the objective function in the FTRL update Eq. (4). To simplify this term, note that since  $\psi$  is a barrier and  $y_t$  minimizes  $G_{t-1}$ , we have

$$\nabla G_{t-1}(y_t) = 0 = \sum_{s=1}^{t-1} (g_s + (\sigma_s + \lambda_s)(y_t - y_s)) + \frac{1}{\eta_t} \nabla \psi(y_t).$$

Further combining this with  $\nabla G_t(y_t) = \sum_{s=1}^t g_s + \sum_{s=1}^{t-1} (\sigma_s + \lambda_s)(y_t - y_s) + \frac{1}{\eta_{t+1}} \nabla \psi(y_t)$  shows  $\nabla G_t(y_t) = g_t + (\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}) \nabla \psi(y_t)$ . Now, if  $\eta_t$  is fixed for all  $t$ , then the Newton decrement simply becomes  $\|g_t\|_{\nabla^{-2}G_t(y_t)}^2 \preceq \|g_t\|_{\nabla^{-2}G_{t-1}(y_t)}^2$ , which by the definition of  $g_t$  is directly bounded by  $\eta_t d^2$ . However, with decreasing learning rates, the extra term contributes to a term of order  $(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t})^2 \|\nabla \psi(y_t)\|_{\nabla^{-2}G_t(y_t)}^2$ , which could be prohibitively large unfortunately.<sup>5</sup>

To resolve this issue, our key observation is:  $\|\nabla \psi(y_t)\|_{\nabla^{-2}G_t(y_t)}^2 \preceq \eta_{t+1} \|\nabla \psi(y_t)\|_{\nabla^{-2}\psi(y_t)}^2$ , and  $\|\nabla \psi(y_t)\|_{\nabla^{-2}\psi(y_t)}^2$  is always bounded by  $\nu$  as long as the self-concordant barrier  $\psi$  is also *logarithmically homogeneous* (see Appendix C for definition and Lemma 29 for this property). A logarithmically homogeneous self-concordant barrier is also called a *normal barrier* for short, and it is only defined for a cone (recall that our feasible set  $\mathcal{X}$ , on the other hand, is always bounded, thus not a cone). Fortunately, this issue has been addressed in a recent work by Lee et al. (2020) on achieving high probability regret bounds for adversarial linear bandits. Their motivation is different, that is, to bound the variance of the learner's loss, related to  $\|y_t\|_{\nabla^2 \psi(y_t)}^2$  using our notation, and this turns out to be bounded when  $\psi$  is a normal barrier — in fact,  $\|y_t\|_{\nabla^2 \psi(y_t)}^2$  and  $\|\nabla \psi(y_t)\|_{\nabla^{-2}\psi(y_t)}^2$  are exactly the same in this case! Their solution regarding  $\mathcal{X}$  not being a cone is to first lift it to  $\mathbb{R}^{d+1}$  and find a normal barrier of the conic hull of this lifted domain (which always exists), then perform FTRL over the lifted domain with this normal barrier regularizer. We extend their idea from linear bandits to convex bandits, formally described below (see also Algorithm 1 for the pseudocode).

5. Instead of FTRL, one might wonder if using the highly related Online Mirror Descent framework could solve the issue caused by decreasing learning rates. We point out that while this indeed addresses the issue for bounding the stability term, it on the other hand introduces a similar issue for the regularization penalty term  $\sum_{t=2}^T (\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}) D_\psi(x, y_t)$ .

---

**Algorithm 1** Adaptive Smooth BCO with Heterogeneous Strong Convexity
 

---

**Input:** smoothness parameter  $\beta$  and a  $\nu$ -self-concordant barrier  $\psi$  for the feasible domain  $\mathcal{X}$ .

**Define:** lifted feasible set  $\mathcal{X} = \{\mathbf{x} = (x, 1) \mid x \in \mathcal{X}\}$ .

**Define:**  $\Psi$  is a normal barrier of the conic hull of  $\mathcal{X}$ :  $\Psi(\mathbf{x}) = \Psi(x, b) = 400(\psi(x/b) - 2\nu \ln b)$ .

**Define:**  $\rho = 512\nu(1 + 32\sqrt{\nu})^2$ .

**Initialize:**  $\lambda_0 = \max\{(\beta + 1)\rho\nu^{-1}, d^2(\beta + 1)\}$  and  $\eta_1 = \frac{1}{2d}\sqrt{\frac{\beta+1}{\lambda_0} + \frac{\nu}{T \log T}}$ .

**Initialize:**  $\mathbf{y}_1 = (y_1, 1) = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \Psi(\mathbf{x})$ .

**for**  $t = 1, 2, \dots, T$  **do**

- 1     Compute  $\mathbf{H}_t = \nabla^2 \Psi(\mathbf{y}_t) + \eta_t (\sigma_{1:t-1} + \lambda_{0:t-1}) \mathbf{I}$ .
- 2     Draw  $\mathbf{u}_t$  uniformly at random from  $\mathbb{S}^{d+1} \cap (\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1})^\perp$ .      $\triangleright \mathbf{w}^\perp$ : space orthogonal to  $\mathbf{w}$
- 3     Compute  $\mathbf{x}_t = (x_t, 1) = \mathbf{y}_t + \mathbf{H}_t^{-\frac{1}{2}} \mathbf{u}_t$ , play the point  $x_t$ , and observe  $f_t(x_t)$  and  $\sigma_t$ .
- 4     Compute regularization coefficient  $\lambda_t \in (0, 1)$ .      $\triangleright$  See Eq. (6) and related discussions
- 5     Compute gradient estimator  $\mathbf{g}_t = d(f_t(x_t) + \frac{\lambda_t}{2} \|x_t\|_2^2) \mathbf{H}_t^{\frac{1}{2}} \mathbf{u}_t$ .
- 6     Compute learning rate  $\eta_{t+1} = \frac{1}{2d}\sqrt{\frac{\beta+1}{\sigma_{1:t} + \lambda_{0:t}} + \frac{\nu}{T \log T}}$ .
- 7     Update  $\mathbf{y}_{t+1} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \left\{ \sum_{s=1}^t (\mathbf{g}_s^\top \mathbf{x} + \frac{\sigma_s + \lambda_s}{2} \|\mathbf{x} - \mathbf{y}_s\|_2^2) + \frac{\lambda_0}{2} \|\mathbf{x}\|_2^2 + \frac{1}{\eta_{t+1}} \Psi(\mathbf{x}) \right\}$ .

**end**

---

**Lifted domain and normal barrier.** To make the dimension of a vector/matrix self-evident, we use bold letters to represent vectors in  $\mathbb{R}^{d+1}$  and matrices in  $\mathbb{R}^{(d+1) \times (d+1)}$ . Define the lifted domain as  $\mathcal{X} = \{\mathbf{x} = (x, 1) \mid x \in \mathcal{X}\} \subseteq \mathbb{R}^{d+1}$ , which simply appends an additional coordinate with constant value 1 to all points in  $\mathcal{X}$ . Define the conic hull of this set as  $\mathcal{K} = \{(x, b) \mid x \in \mathbb{R}^d, b \geq 0, \frac{1}{b}x \in \mathcal{X}\}$ . Our algorithm requires using a normal barrier over  $\mathcal{K}$  as a regularizer (which always exists). While any such normal barrier works, we simply use a canonical one constructed from a  $\nu$ -self concordant barrier  $\psi$  of  $\mathcal{X}$ , defined via  $\Psi(\mathbf{x}) = \Psi(x, b) = 400\psi(x/b) - 2\nu \ln b$  and proven to be a  $\Theta(\nu)$ -normal barrier over  $\mathcal{K}$  in (Nesterov and Nemirovskii, 1994, Proposition 5.1.4). This also shows that our algorithm requires no more than that of Hazan and Levy (2014).

Algorithm 1 then performs FTRL in the lifted domain using  $\Psi$  as the regularizer to maintain an auxiliary sequence  $\mathbf{y}_1, \dots, \mathbf{y}_T$ ; see Line 7. This follows the earlier update rule in Eq. (4), except for an additional  $\ell_2$  regularization term  $\frac{\lambda_0}{2} \|\mathbf{x}\|_2^2$  added for technical reasons. With  $\mathbf{y}_t$  at hand, we compute Hessian matrix  $\mathbf{H}_t = \nabla^2 \Psi(\mathbf{y}_t) + \eta_t (\sigma_{1:t-1} + \lambda_{0:t-1}) \mathbf{I}$  similarly as before (Line 1). What is slightly different now is the exploration (Line 3): we sample  $\mathbf{u}_t$  uniformly at random from the set  $\mathbb{S}^{d+1} \cap (\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1})^\perp$  where  $\mathbf{w}^\perp$  denotes the space orthogonal to  $\mathbf{w}$ , and then obtain a point  $\mathbf{x}_t = \mathbf{y}_t + \mathbf{H}_t^{-\frac{1}{2}} \mathbf{u}_t$ . It can be shown that  $\mathbf{x}_t$  is always on the intersection of the lifted domain  $\mathcal{X}$  and the surface of some ellipsoid centered at  $\mathbf{y}_t$  — we refer the reader to (Lee et al., 2020, Figure 1) for a pictorial illustration and their description for how to sample  $\mathbf{u}_t$  efficiently. Since  $\mathbf{x}_t \in \mathcal{X}$ , it is in the form of  $(x_t, 1)$ , where  $x_t$  will be the final decision of the algorithm.

Upon receiving  $f_t(x_t)$  and  $\sigma_t$ , Algorithm 1 computes the  $\ell_2$  regularization coefficient  $\lambda_t$  in some way (to be discussed soon), gradient estimator  $\mathbf{g}_t$  as in earlier discussion (Line 5), learning rate  $\eta_{t+1}$  as in Line 6, and finally  $\mathbf{y}_{t+1}$  via the aforementioned FTRL.



**Guarantees and  $\ell_2$  regularization coefficient tuning.** We now present some guarantees of our algorithm that hold regardless of the tuning of  $\{\lambda_t\}_{t=1}^T$ . First, we show that except for the last coordinate,  $\mathbf{g}_t$  is an unbiased estimator of a smoothed version of  $\tilde{\mathbf{f}}_t(\mathbf{y}_t)$ , where  $\tilde{\mathbf{f}}_t(\mathbf{y}) = f_t(\mathbf{y}_{[1:d]}) + \frac{\lambda_t}{2} \|\mathbf{y}_{[1:d]}\|_2^2$ . This is a non-trivial generalization of (Lee et al., 2020, Lemma B.9) from linear functions to convex functions. See Appendix A.1 for the proof.

**Lemma 4** For each  $t \in [T]$ , we have  $\mathbb{E}_t[\mathbf{g}_{t,i}] = \nabla \hat{\mathbf{f}}_t(\mathbf{y}_t)_{[i]}$  for all  $i \in [d]$ , where  $\hat{\mathbf{f}}_t$  is the smoothed version of  $\tilde{\mathbf{f}}_t$  defined as  $\hat{\mathbf{f}}_t(\mathbf{x}) \triangleq \mathbb{E}_{\mathbf{b}}[\tilde{\mathbf{f}}_t(\mathbf{x} + \mathbf{H}_t^{-1/2}\mathbf{b})]$ , where  $\mathbf{b}$  is uniformly sampled from  $\mathbb{B}^{d+1} \cap (\mathbf{H}_t^{-1/2}\mathbf{e}_{d+1})^\perp$ .

Thanks to the unbiasedness of the gradient estimators and the crucial properties of normal barrier, we prove the following regret guarantee of Algorithm 1 in Appendix A.3.

**Lemma 5** With any regularization coefficients  $\{\lambda_t\}_{t=1}^T \in (0, 1)$ , Algorithm 1 guarantees:

$$\text{REG} = \tilde{\mathcal{O}} \left( d\sqrt{\nu T} + \lambda_{1:T-1} + \sum_{t=1}^{T-1} \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{0:t}}} \right), \quad (5)$$

if loss functions  $\{f_t\}_{t=1}^T$  are all  $\beta$ -smooth and  $T \geq \rho$  (a constant defined in Algorithm 1).

We are now in the position to specify the tuning of the regularization coefficients. Based on the bound in Eq. (5), we propose to balance the last two terms by picking  $\lambda_t \in (0, 1)$  such that:

$$\lambda_t = \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{0:t}}}, \quad (6)$$

which must exist since when  $\lambda_t = 0$ , the left-hand side is smaller than the right-hand side, while when  $\lambda_t = 1$ , the left-hand side is larger than the right-hand side by the definition of  $\lambda_0$ . Note that unlike the AOGD tuning in Eq. (2), our tuning leads to a cubic equation of  $\lambda_t$ , which does not admit a closed-form. However, the earlier argument on its existence clearly also implies that it can be computed via a simple and efficient binary search (using information available at the end of round  $t$ ). Our next lemma is in the same vein as (Bartlett et al., 2007, Lemma 3.1), which shows that our adaptive tuning is almost as good as the optimal tuning (that knows all  $\sigma_t$ 's ahead of time). This lemma turns out to be crucial to achieve our guarantees that adapt to the environment with heterogeneous curvatures.

**Lemma 6** Define  $\mathcal{B}(\{\lambda_s\}_{s=1}^t) \triangleq \lambda_{1:t} + \sum_{\tau=1}^t \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:\tau} + \lambda_{0:\tau}}}$ , with  $\lambda_0$  defined in Algorithm 1. Then the sequence  $\{\lambda_t\}_{t=1}^T$  attained by solving Eq. (6) satisfies for all  $t \in [T]$ :

$$\mathcal{B}(\{\lambda_s\}_{s=1}^t) \leq 2 \min_{\{\lambda_s^*\}_{s=1}^t \geq 0} \mathcal{B}(\{\lambda_s^*\}_{s=1}^t). \quad (7)$$

The proof of Lemma 6 is deferred in Appendix A.4. Combining Lemma 5 and Lemma 6, we obtain the final regret guarantee in Theorem 7, whose proof can be found in Appendix A.5.

**Theorem 7** Algorithm 1 with adaptive tuning Eq. (6) ensures for any sequence  $\lambda_1^*, \dots, \lambda_T^* \geq 0$ :

$$\text{REG} = \tilde{\mathcal{O}} \left( d\sqrt{\nu T} + \lambda_{1:T-1}^* + \sum_{t=1}^{T-1} \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{0:t}^*}} \right), \quad (8)$$

when loss functions  $\{f_t\}_{t=1}^T$  are all  $\beta$ -smooth and  $T \geq \rho$  (a constant defined in Algorithm 1).

We leave the discussion on the many implications of this general regret bound to the next subsection, and make a final remark on the per-round computational complexity of [Algorithm 1](#). Note that although the objective function in the FTRL update ([Line 7](#)) contains  $\mathcal{O}(T)$  terms, it is clear that by storing and updating some statistics (such as  $\sum_{s=1}^t \mathbf{g}_s$  and  $\sum_{s=1}^t (\sigma_s + \lambda_s) \mathbf{y}_s$ ), one can evaluate its function value and gradient in time independent of  $T$ . Approximating solving the FTRL update (up to precision  $1/\text{poly}(T)$ ) via for example the interior point method thus only requires  $\mathcal{O}(\text{poly}(d \log T))$  time. This is more efficient than the method of [Bubeck et al. \(2021\)](#), which requires  $\mathcal{O}(\text{poly}(d \log T)T)$  time per round even when the domain is a polytope.

### 3.2. Implications of [Theorem 7](#)

In the following, we investigate several special cases and present direct corollaries of [Theorem 7](#) to demonstrate that our algorithm not only matches/improves existing results for homogeneous settings, but also leads to interesting intermediate results in some heterogeneous settings. Note that since our regret bound in [Theorem 7](#) holds for *any* choice of the sequence  $\{\lambda_t^*\}_{t=1}^T \geq 0$ , in each case below we will simply provide a specific sequence of  $\{\lambda_t^*\}_{t=1}^T$  that leads to a favorable guarantee. For simplicity, we also directly replace  $\nu$  with  $\mathcal{O}(d)$  (in both our bounds and previous results) since it is well known that any convex set admits an  $\mathcal{O}(d)$ -self-concordant barrier ([Nesterov and Nemirovskii, 1994](#)).

First, consider the case when no functions have strong convexity, that is,  $\sigma_t = 0$  for all  $t$ . This degenerates to the same homogeneous setting as ([Saha and Tewari, 2011](#)), where their algorithm achieves  $\tilde{\mathcal{O}}(d^{3/2}\sqrt{T} + \beta^{1/3}dT^{2/3})$  regret. As a side product of our adaptive  $\ell_2$ -regularization, our algorithm manages to achieve even better dependency on the dimension  $d$ . Indeed, by choosing  $\lambda_1^* = (1 + \beta)^{1/3}d^{2/3}T^{2/3}$  and  $\lambda_t^* = 0$  for all  $t \geq 2$  in [Theorem 7](#), we obtain the following corollary.

**Corollary 8 (Smooth BCO without strong convexity)** *When  $f_t$  is  $\beta$ -smooth and 0-strongly convex for all  $t \in [T]$ , [Algorithm 1](#) achieves  $\tilde{\mathcal{O}}(d^{3/2}\sqrt{T} + (1 + \beta)^{1/3}d^{2/3}T^{2/3})$  regret.*

Second, we consider the case when all functions are  $\sigma$ -strongly convex for some constant  $\sigma > 0$ , which degenerates to the same homogeneous setting as ([Hazan and Levy, 2014](#)). By picking  $\lambda_t^* = 0$  for all  $t \geq 0$  in [Theorem 7](#), our algorithm achieves the same result as theirs.

**Corollary 9 (Smooth BCO with  $\sigma$ -strong convexity)** *When  $f_t$  is  $\beta$ -smooth and  $\sigma$ -strongly convex, i.e.,  $\sigma_t = \sigma > 0$  for all  $t \in [T]$ , [Algorithm 1](#) achieves  $\tilde{\mathcal{O}}(d^{3/2}\sqrt{T} + d\sqrt{T(1+\beta)/\sigma})$  regret.*

Third, we investigate an intermediate setting with a mixture of  $\sigma$ -strongly convex and 0-strongly convex functions. Specifically, suppose that there are  $M$  functions with no strong convexity, and the rest are  $\sigma$ -strongly convex. According to [Theorem 7](#), the worst case scenario for our algorithm is when these  $M$  functions appear in the first  $M$  rounds, while the best scenarios is when they are in the last  $M$  rounds. Considering these two extremes and picking  $\{\lambda_t^*\}_{t=1}^T$  correspondingly, we obtain the following corollary (see [Appendix A.6](#) for the proof).

**Corollary 10 (Smooth BCO with a mixture of convex and  $\sigma$ -strongly convex functions)** *Suppose that  $\{f_t\}_{t=1}^T$  are  $\beta$ -smooth and  $T - M$  of them are  $\sigma$ -strongly convex. Then [Algorithm 1](#) guarantees*

$$\text{REG} = \tilde{\mathcal{O}} \left( d^{\frac{2}{3}}(1 + \beta)^{\frac{1}{3}}M^{\frac{2}{3}} + d^{\frac{3}{2}}\sqrt{T} + d\sqrt{\frac{(1 + \beta)(T - M)}{\sigma}} \right).$$

If these  $T - M$  functions appear in the first  $T - M$  rounds, then the bound is further improved to

$$\text{REG} = \tilde{\mathcal{O}} \left( d^{\frac{3}{2}} \sqrt{T} + dT \sqrt{\frac{1 + \beta}{\sigma(T - M)}} \right).$$

To better interpret these bounds, we consider how large  $M$  can be (that is, how many functions without strong convexity we can tolerate) to still ensure  $\tilde{\mathcal{O}}(\sqrt{T})$  regret — in the general case (the first bound of the corollary), we see that we can tolerate  $M = \mathcal{O}(T^{3/4})$ , while in the best case (the second bound), we can even tolerate  $M$  being any constant fraction of  $T$ ! On the other hand, a naive method of discarding all functions without strong convexity can only tolerate  $M = \tilde{\mathcal{O}}(\sqrt{T})$ .

Finally, following (Bartlett et al., 2007) we consider a situation with decaying strong convexity:  $\sigma_t = t^{-\alpha}$  for some  $\alpha \in [0, 1]$ . We prove the following corollary; see Appendix A.6 for the proof.

**Corollary 11 (Smooth BCO with decaying strong convexity)** *When  $f_t$  is  $\beta$ -smooth and  $\sigma_t$ -strongly convex with  $\sigma_t = t^{-\alpha}$  for some  $\alpha \in [0, 1]$ , Algorithm 1 guarantees*

$$\text{REG} = \begin{cases} \tilde{\mathcal{O}} \left( d^{\frac{3}{2}} \sqrt{T} + d\sqrt{1 + \beta} T^{\frac{1+\alpha}{2}} \right) & \alpha \in [0, \frac{1}{3} - \frac{2}{3} \log_T d - \frac{1}{3} \log_T(1 + \beta)], \\ \tilde{\mathcal{O}} \left( d^{\frac{3}{2}} \sqrt{T} + (1 + \beta)^{\frac{1}{3}} d^{\frac{2}{3}} T^{\frac{2}{3}} \right) & \alpha \in [\frac{1}{3} - \frac{2}{3} \log_T d - \frac{1}{3} \log_T(1 + \beta), 1]. \end{cases}$$

#### 4. Lipschitz Bandit Convex Optimization with Heterogeneous Strong Convexity

In this section, we consider a similar setting where instead of assuming smoothness, we assume that functions  $\{f_t\}_{t=1}^T$  are known to be  $L$ -Lipschitz (Definition 2). The strong convexity parameter  $\sigma_t$  of function  $f_t$  is still only revealed at the end of round  $t$ . We extend our algorithm to this case and present it in Algorithm 2, which differs from Algorithm 1 only in the tuning of the learning rate  $\eta_t$  (see Line 6) and the regularization coefficient  $\lambda_t$  (see Line 4). These tunings are different because of the different structures in the setting, but their design follows the same idea as before. Similar to Theorem 7, we prove the following theorem (see Appendix B for the proof).

**Theorem 12** *Algorithm 2 ensures for any sequence  $\lambda_1^*, \dots, \lambda_T^* \geq 0$ :*

$$\text{REG} = \tilde{\mathcal{O}} \left( d^{\frac{4}{3}} \nu T^{\frac{1}{3}} + \lambda_{0:T}^* + \sum_{t=1}^T \frac{d^{\frac{2}{3}} (L + 1)^{\frac{2}{3}}}{(\sigma_{1:t-1} + \lambda_{0:t-1}^*)^{\frac{1}{3}}} \right), \quad (10)$$

when all the functions are  $L$ -Lipschitz and  $T \geq \rho'$  (a constant defined in Algorithm 2).

Similar to Section 3.2, we now discuss the implications of this theorem in several special cases, demonstrating that our algorithm not only improves existing results in the homogeneous settings as a side product of the better regularization technique, but also achieves favorable guarantees in some heterogeneous settings. Again, we plug in  $\nu = \mathcal{O}(d)$  for simplicity.

First, we consider the case when no functions have strong convexity, which degenerates to the same homogeneous setting studied in (Kleinberg, 2004; Flaxman et al., 2005; Luo, 2017). Among these results, the best regret bound is  $\tilde{\mathcal{O}}(\sqrt{L} d^{3/4} T^{3/4})$  (Luo, 2017). By picking  $\lambda_1^* = \sqrt{d(L + 1)} T^{3/4}$  and  $\lambda_t^* = 0$  for all  $t \geq 2$  in Theorem 12, we achieve the following guarantee with improved dimension dependency.

---

**Algorithm 2** Adaptive Lipschitz BCO with Heterogeneous Strong Convexity
 

---

**Input:** Lipschitz parameter  $L$  and a  $\nu$ -self-concordant barrier  $\psi$  for the feasible domain  $\mathcal{X}$ .

**Define:** lifted feasible set  $\mathcal{X} = \{x = (x, 1) \mid x \in \mathcal{X}\}$ .

**Define:**  $\Psi$  is a normal barrier of the conic hull of  $\mathcal{X}$ :  $\Psi(x) = \Psi(x, b) = 400(\psi(x/b) - 2\nu \ln b)$ .

**Define:**  $\rho' = \frac{2^{16}(16\sqrt{\nu}d^{1/3}(4L+1)^{1/3}+(L+1)^{2/3})^3}{d}$ .

**Initialize:**  $\lambda_0 = \max\{\rho', d^2(L+1)^2\}$  and  $\eta_1 = (L+1)^{\frac{2}{3}}d^{-\frac{4}{3}}(\frac{1}{\lambda_0} + \frac{1}{T})^{\frac{1}{3}}$ .

**Initialize:**  $y_1 = (y_1, 1) = \operatorname{argmin}_{x \in \mathcal{X}} \Psi(x)$ .

**for**  $t = 1, 2, \dots, T$  **do**

- 1    Define  $H_t = \nabla^2 \Psi(y_t) + \eta_t (\sigma_{1:t-1} + \lambda_{0:t-1}) I$ .
- 2    Draw  $u_t$  uniformly at random from  $\mathbb{S}^{d+1} \cap (H_t^{-\frac{1}{2}} e_{d+1})^\perp$ .     $\triangleright w^\perp$ : space orthogonal to  $w$
- 3    Compute  $x_t = (x_t, 1) = y_t + H_t^{-\frac{1}{2}} u_t$ , play the point  $x_t$ , and observe  $f_t(x_t)$  and  $\sigma_t$ .
- 4    Compute regularization coefficient  $\lambda_t \in (0, 1)$  as the solution of the following equation

$$\lambda_t = \frac{d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\sigma_{1:t} + \lambda_{0:t})^{\frac{1}{3}}}. \quad (9)$$

- 5    Compute gradient estimator  $g_t = d(f_t(x_t) + \frac{\lambda_t}{2}\|x_t\|_2^2) H_t^{\frac{1}{2}} u_t$ .
- 6    Compute learning rate  $\eta_{t+1} = d^{-\frac{4}{3}}(L+1)^{\frac{2}{3}} \cdot \left(\frac{1}{\sigma_{1:t} + \lambda_{0:t}} + \frac{1}{T}\right)^{\frac{1}{3}}$ .
- 7    Update  $y_{t+1} = \operatorname{argmin}_{x \in \mathcal{X}} \left\{ \sum_{s=1}^t (g_s^\top x + \frac{\sigma_s + \lambda_s}{2}\|x - y_s\|_2^2) + \frac{\lambda_0}{2}\|x\|_2^2 + \frac{1}{\eta_{t+1}}\Psi(x) \right\}$ .

**end**

---

**Corollary 13 (Lipschitz BCO without strong convexity)** *When  $f_t$  is  $L$ -Lipschitz and 0-strongly convex for all  $t \in [T]$ , Algorithm 2 achieves  $\tilde{O}(\sqrt{d(L+1)}T^{3/4})$  regret.*

Second, we consider the case when all loss functions are  $\sigma$ -strongly convex. This degenerates to the homogeneous setting studied in (Agarwal et al., 2010), where they achieve  $\tilde{O}(d^{4/3}L^{2/3}\sigma^{-1/3}T^{2/3})$  regret.<sup>6</sup> Once again, by picking  $\lambda_t^* = 0$  for all  $t \geq 1$  in Theorem 12, we obtain the following result with improved dimension dependency.

**Corollary 14 (Lipschitz BCO with  $\sigma$ -strong convexity)** *When  $f_t$  is  $L$ -Lipschitz and  $\sigma$ -strongly convex, i.e.,  $\sigma_t = \sigma$  for all  $t \in [T]$ , Algorithm 2 achieves  $\tilde{O}(d^{2/3}(L+1)^{2/3}\sigma^{-1/3}T^{2/3})$  regret.*

Third, we consider the case with a mixture of 0-strongly convex and  $\sigma$ -strongly convex functions.

**Corollary 15 (Lipschitz BCO with a mixture of convex and  $\sigma$ -strongly convex functions)** *Suppose that  $\{f_t\}_{t=1}^T$  are  $L$ -Lipschitz and  $T - M$  of them are  $\sigma$ -strongly convex. Then Algorithm 2 guarantees*

$$\text{REG} = \left( \sqrt{d(L+1)}M^{3/4} + d^{2/3}(L+1)^{2/3}\sigma^{-1/3}(T-M)^{2/3} \right).$$

---

6. The bound stated in their paper has  $d^{2/3}$  dependency on the dimension, but that is under a different assumption on  $\mathcal{X}$ . Translating their setting to ours via a reshaping trick (Flaxman et al., 2005, Section 3.2) leads to the  $d^{4/3}$  dependency.

If these  $T - M$  functions appear in the first  $T - M$  rounds, then the bound is further improved to

$$\text{REG} = \tilde{\mathcal{O}} \left( \frac{d^{2/3}(L+1)^{2/3}T}{\sigma^{1/3}(T-M)^{1/3}} \right).$$

The proof can be found in [Appendix B.2](#). Similar to the discussion in [Section 3.2](#), we consider how large  $M$  can be to still ensure  $\tilde{\mathcal{O}}(T^{2/3})$  regret — according to the first bound, we can always tolerate  $M = \mathcal{O}(T^{8/9})$ , while in the best case (the second bound), we can tolerate  $M$  being any constant fraction of  $T$ . These are again much stronger compared to the naive method of discarding all functions without strong convexity, which can only tolerate  $M = \tilde{\mathcal{O}}(T^{2/3})$ .

Finally, we consider the example with  $\sigma_t = t^{-\alpha}$  again. See [Appendix B.2](#) for the proof.

**Corollary 16 (Lipschitz BCO with decaying strong convexity)** *When  $f_t$  is  $L$ -Lipschitz and  $\sigma_t$ -strongly convex with  $\sigma_t = t^{-\alpha}$  for some  $\alpha \in [0, 1]$ , [Algorithm 2](#) guarantees*

$$\text{REG} = \begin{cases} \tilde{\mathcal{O}}(d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}T^{\frac{2+\alpha}{3}}) & \alpha \in [0, \frac{1}{4} - \frac{1}{2} \log_T(L+1) - \frac{1}{2} \log_T d], \\ \tilde{\mathcal{O}}(\sqrt{d(L+1)}T^{\frac{3}{4}}) & \alpha \in [\frac{1}{4} - \frac{1}{2} \log_T(L+1) - \frac{1}{2} \log_T d, 1]. \end{cases}$$

## 5. Conclusion

Our work initiates the study of bandit convex optimization with heterogeneous curvature and proposes strong algorithms and guarantees that automatically adapt to individual curvature of each loss function. As the first step in this direction, we have assumed homogeneous smoothness or Lipschitzness and only considered heterogeneous strong convexity. Extending the heterogeneity to the other curvature parameters is an immediate next step. Moreover, it is worth investigating an even more challenging setting where the individual curvature information is not revealed to the learner at the end of each round, or at least has to be learned via other weaker and indirect feedback (such as some rough and potentially incorrect estimation of the curvature).

## Acknowledgments

Peng Zhao is supported by National Science Foundation of China (61921006). Haipeng Luo and Mengxiao Zhang are supported by NSF Award IIS1943607.

## References

- Jacob Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT)*, pages 263–274, 2008.
- Jacob D Abernethy, Elad Hazan, and Alexander Rakhlin. Interior-point methods for full-information and bandit online learning. *IEEE Transactions on Information Theory*, 58(7):4164–4175, 2012.
- Alekh Agarwal, Ofer Dekel, and Lin Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Proceedings of the 23rd Conference on Learning Theory (COLT)*, pages 28–40, 2010.

- Alekh Agarwal, Haipeng Luo, Behnam Neyshabur, and Robert E Schapire. Corraling a band of bandit algorithms. In *Proceedings of the 30th Conference on Learning Theory (COLT)*, pages 12–38, 2017.
- Peter L. Bartlett, Elad Hazan, and Alexander Rakhlin. Adaptive online gradient descent. In *Advances in Neural Information Processing Systems 20 (NIPS)*, pages 65–72, 2007.
- Sébastien Bubeck and Ronen Eldan. Multi-scale exploration of convex functions and bandit convex optimization. In *Proceedings of the 29th Conference on Learning Theory (COLT)*, volume 49, pages 583–589, 2016.
- Sébastien Bubeck, Nicolo Cesa-Bianchi, and Sham M Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*, pages 41.1–41.14, 2012.
- Sébastien Bubeck, Yuanzhi Li, Haipeng Luo, and Chen-Yu Wei. Improved path-length regret bounds for bandits. In *Proceedings of the 32nd Conference on Learning Theory (COLT)*, pages 508–528, 2019.
- Sébastien Bubeck, Ronen Eldan, and Yin Tat Lee. Kernel-based methods for bandit convex optimization. *Journal of ACM*, 68(4), 2021.
- Sébastien Bubeck, Ofer Dekel, Tomer Koren, and Yuval Peres. Bandit convex optimization:  $\sqrt{T}$  regret in one dimension. In *Proceedings of the 28th Conference on Learning Theory (COLT)*, pages 266–278, 2015.
- Varsha Dani, Thomas P Hayes, and Sham M Kakade. The price of bandit information for online optimization. In *Advances in Neural Information Processing Systems 20 (NIPS)*, pages 345–352, 2008.
- Ofer Dekel, Ronen Eldan, and Tomer Koren. Bandit smooth convex optimization: Improving the bias-variance tradeoff. In *Advances in Neural Information Processing Systems 28 (NIPS)*, pages 2926–2934, 2015.
- Abraham Flaxman, Adam Tauman Kalai, and H. Brendan 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 (SODA)*, pages 385–394, 2005.
- Elad Hazan and Satyen Kale. Better algorithms for benign bandits. *Journal of Machine Learning Research*, 12:1287–1311, 2011.
- Elad Hazan and Kfir Y. Levy. Bandit convex optimization: Towards tight bounds. In *Advances in Neural Information Processing Systems 27 (NIPS)*, pages 784–792, 2014.
- Elad Hazan and Yuanzhi Li. An optimal algorithm for bandit convex optimization. *arXiv preprint arXiv:1603.04350*, 2016.
- Xiaowei Hu, Prashanth L. A., András György, and Csaba Szepesvári. (bandit) convex optimization with biased noisy gradient oracles. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 819–828, 2016.

- Shinji Ito. An optimal algorithm for bandit convex optimization with strongly-convex and smooth loss. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 2229–2239, 2020.
- Robert D. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems 17 (NIPS)*, pages 697–704, 2004.
- Tor Lattimore. Improved regret for zeroth-order adversarial bandit convex optimisation. *Mathematical Statistics and Learning*, 2(3):311–334, 2020.
- 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. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*, pages 15522–15533, 2020.
- Haipeng Luo. Lecture note 18, Introduction to Online Learning. 2017. URL <https://haipeng-luo.net/courses/CSCI699/lecture18.pdf>.
- Teodor Vanislavov Marinov and Julian Zimmert. The pareto frontier of model selection for general contextual bandits. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, 2021.
- Arkadi S Nemirovski and Michael J Todd. Interior-point methods for optimization. *Acta Numerica*, 17:191–234, 2008.
- Yurii Nesterov and Arkadii Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, 1994.
- Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Proceedings of the 26th Conference On Learning Theory (COLT)*, pages 993–1019, 2013.
- Aadirupa Saha, Nagarajan Natarajan, Praneeth Netrapalli, and Prateek Jain. Optimal regret algorithm for pseudo-1d bandit convex optimization. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, pages 9255–9264, 2021.
- Ankan Saha and Ambuj Tewari. Improved regret guarantees for online smooth convex optimization with bandit feedback. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 636–642, 2011.
- Arun Sai Suggala, Pradeep Ravikumar, and Praneeth Netrapalli. Efficient bandit convex optimization: Beyond linear losses. In *Proceedings of the 34th Conference on Learning Theory (COLT)*, pages 4008–4067, 2021.
- Scott Yang and Mehryar Mohri. Optimistic bandit convex optimization. In *Advances in Neural Information Processing Systems 29 (NIPS)*, pages 2289–2297, 2016.

## Appendix A. Omitted Details for Section 3

### A.1. Proof of Lemma 4

In this section, we prove one of our key lemmas (Lemma 4), which shows that the lifted gradient estimator constructed in Line 5 of Algorithm 1 is an unbiased estimator in the first  $d$  dimensional coordinates.

**Proof** Fix any  $t \in [T]$  and let  $\mathbf{w} = \mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1} / \|\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1}\|_2$ . As  $\mathbf{u} \sim \mathbb{S}^{d+1} \cap (\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1})^\perp$  and  $\mathbf{b} \sim \mathbb{B}^{d+1} \cap (\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1})^\perp$ , there exists a transformation matrix  $\mathbf{M} \in \mathbb{R}^{d \times (d+1)}$  that satisfies  $\mathbf{M}\mathbf{w} = 0$ ,  $\mathbf{M}^\top \mathbf{M} = \mathbf{I} - \mathbf{w}\mathbf{w}^\top$  and  $\mathbf{M}\mathbf{M}^\top = \mathbf{I}$ , such that  $\mathbf{u} = \mathbf{M}^\top v$ ,  $\mathbf{b} = \mathbf{M}^\top b$  where  $v$  is uniformly drawn from  $\mathbb{S}^d$  and  $b$  is uniformly drawn from  $\mathbb{B}^d$ . In fact, the  $d$  row vectors of  $\mathbf{M}$  together with  $\mathbf{w}$  forms a set of unit orthogonal base in the  $(d+1)$ -dimensional space.

Recall the two following functions whose feasible domain is in  $(d+1)$ -dimensional space.

$$\mathbf{f}_t(\mathbf{x}) \triangleq f_t(\mathbf{x}_{[1:d]}), \quad (11)$$

$$\tilde{\mathbf{f}}_t(\mathbf{x}) \triangleq \mathbf{f}_t(\mathbf{x}) + \frac{\lambda_t}{2} \|\mathbf{x}_{[1:d]}\|_2^2 = f_t(\mathbf{x}_{[1:d]}) + \frac{\lambda_t}{2} \|\mathbf{x}_{[1:d]}\|_2^2 \quad (12)$$

Then we define the following functions in the  $d$ -dimensional space. Let  $J_t : \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $J_t(x) = \tilde{\mathbf{f}}_t(\mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top x)$  and  $\hat{J}_t(x) = \mathbb{E}_{b \sim \mathbb{B}^d} [J_t(x+b)]$ . In addition, we denote  $\hat{\mathbf{y}}_t = (y_t, 0) \in \mathbb{R}^{d+1}$  that appends an additional constant value 0 to  $y_t$  in the  $(d+1)$ -th coordinate. Then by the definition of  $\mathbf{g}_t$ , we have

$$\begin{aligned} \mathbb{E}_t[\mathbf{g}_t] &= d \mathbb{E}_{\mathbf{u} \sim \mathbb{S}^{d+1} \cap (\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1})^\perp} \left[ \tilde{\mathbf{f}}_t(\mathbf{y}_t + \mathbf{H}_t^{-\frac{1}{2}} \mathbf{u}) \mathbf{H}_t^{\frac{1}{2}} \mathbf{u} \right] \\ &= d \mathbb{E}_{v \sim \mathbb{S}^d} \left[ J_t(\mathbf{M} \mathbf{H}_t^{\frac{1}{2}} \hat{\mathbf{y}}_t + v) \mathbf{H}_t^{\frac{1}{2}} \mathbf{M}^\top v \right] \\ &= d \mathbf{H}_t^{\frac{1}{2}} \mathbf{M}^\top \mathbb{E}_{v \sim \mathbb{S}^d} \left[ J_t(\mathbf{M} \mathbf{H}_t^{\frac{1}{2}} \hat{\mathbf{y}}_t + v) v \right] \\ &= \mathbf{H}_t^{\frac{1}{2}} \mathbf{M}^\top \nabla \hat{J}_t(\mathbf{M} \mathbf{H}_t^{\frac{1}{2}} \hat{\mathbf{y}}_t), \end{aligned}$$

where the final equality is due to Lemma 5 of Flaxman et al. (2005). The second equality is because of the following reasoning. Note that by the definition of  $J_t$  and properties of  $\mathbf{M}$ , we have

$$\begin{aligned} J_t(\mathbf{M} \mathbf{H}_t^{\frac{1}{2}} \hat{\mathbf{y}}_t + v) &= \tilde{\mathbf{f}}_t(\mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top (\mathbf{M} \mathbf{H}_t^{\frac{1}{2}} \hat{\mathbf{y}}_t + v)) \\ &= \tilde{\mathbf{f}}_t(\mathbf{H}_t^{-\frac{1}{2}} (\mathbf{I} - \mathbf{w}\mathbf{w}^\top) \mathbf{H}_t^{\frac{1}{2}} \hat{\mathbf{y}}_t + \mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top v) \\ &= \tilde{\mathbf{f}}_t((\mathbf{I} - \mathbf{H}_t^{-\frac{1}{2}} \mathbf{w}\mathbf{w}^\top \mathbf{H}_t^{\frac{1}{2}}) \hat{\mathbf{y}}_t + \mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top v). \end{aligned}$$

In addition, by the definition of  $\mathbf{w}$ , we have

$$\mathbf{I} - \mathbf{H}_t^{-\frac{1}{2}} \mathbf{w}\mathbf{w}^\top \mathbf{H}_t^{\frac{1}{2}} = \mathbf{I} - \frac{\mathbf{H}_t^{-1} \mathbf{e}_{d+1} \mathbf{e}_{d+1}^\top}{\|\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1}\|_2^2}$$



Note that the second term has all entries 0 except for the last column, i.e., the  $(d+1)$ -th one. Therefore we have  $\mathbf{H}_t^{-\frac{1}{2}} \mathbf{w} \mathbf{w}^\top \mathbf{H}_t^{\frac{1}{2}} \hat{\mathbf{y}}_t = \mathbf{0}$ , which leads to

$$J_t(\mathbf{M} \mathbf{H}_t^{\frac{1}{2}} \hat{\mathbf{y}}_t + v) = \tilde{\mathbf{f}}_t(\hat{\mathbf{y}}_t + \mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top v) = \tilde{\mathbf{f}}_t(\mathbf{y}_t + \mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top v),$$

where the last equality is because  $\tilde{\mathbf{f}}_t(\mathbf{x})$  has no dependence on the last coordinate of  $\mathbf{x}$ .

Furthermore, according to the definition of  $\hat{J}_t(x)$ , we have

$$\begin{aligned} \nabla \hat{J}_t(x) &= (\mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top)^\top \mathbb{E}_{b \sim \mathbb{B}^d} \left[ \nabla \tilde{\mathbf{f}}_t(\mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top (x + b)) \right] \\ &= (\mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top)^\top \mathbb{E}_{b \sim \mathbb{B}^d} \left[ \nabla \tilde{\mathbf{f}}_t(\mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top x + \mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top b) \right] \\ &= (\mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top)^\top \mathbb{E}_{b \sim \mathbb{B}^{d+1} \cap (\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1})^\perp} \left[ \nabla \tilde{\mathbf{f}}_t(\mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top x + \mathbf{H}_t^{-\frac{1}{2}} b) \right] \\ &= \mathbf{M} \mathbf{H}_t^{-\frac{1}{2}} \nabla \hat{\mathbf{f}}_t(\mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top x), \end{aligned}$$

where the fourth equality is by the definition of  $\hat{\mathbf{f}}_t$ . Therefore, we get

$$\begin{aligned} \mathbb{E}_t[\mathbf{g}_t] &= \mathbf{H}_t^{\frac{1}{2}} \mathbf{M}^\top \nabla \hat{J}_t(\mathbf{M} \mathbf{H}_t^{\frac{1}{2}} \hat{\mathbf{y}}_t) \\ &= \mathbf{H}_t^{\frac{1}{2}} \mathbf{M}^\top \mathbf{M} \mathbf{H}_t^{-\frac{1}{2}} \nabla \hat{\mathbf{f}}_t(\mathbf{H}_t^{-\frac{1}{2}} \mathbf{M}^\top \mathbf{M} \mathbf{H}_t^{\frac{1}{2}} \hat{\mathbf{y}}_t) \\ &= (\mathbf{I} - \mathbf{H}_t^{\frac{1}{2}} \mathbf{w} \mathbf{w}^\top \mathbf{H}_t^{-\frac{1}{2}}) \nabla \hat{\mathbf{f}}_t((\mathbf{I} - \mathbf{H}_t^{-\frac{1}{2}} \mathbf{w} \mathbf{w}^\top \mathbf{H}_t^{\frac{1}{2}}) \hat{\mathbf{y}}_t). \end{aligned} \quad (13)$$

Since  $\mathbf{w} = \mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1} \cdot \|\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1}\|_2^{-1}$ , we have

$$\mathbf{I} - \mathbf{H}_t^{\frac{1}{2}} \mathbf{w} \mathbf{w}^\top \mathbf{H}_t^{-\frac{1}{2}} = \mathbf{I} - \frac{\mathbf{e}_{d+1} \mathbf{e}_{d+1}^\top \mathbf{H}_t^{-1}}{\|\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1}\|_2^2}, \quad \mathbf{I} - \mathbf{H}_t^{-\frac{1}{2}} \mathbf{w} \mathbf{w}^\top \mathbf{H}_t^{\frac{1}{2}} = \mathbf{I} - \frac{\mathbf{H}_t^{-1} \mathbf{e}_{d+1} \mathbf{e}_{d+1}^\top}{\|\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1}\|_2^2}.$$

This shows that for any  $\mathbf{x} \in \mathbb{R}^{d+1}$ ,  $((\mathbf{I} - \mathbf{H}_t^{\frac{1}{2}} \mathbf{w} \mathbf{w}^\top \mathbf{H}_t^{-\frac{1}{2}}) \mathbf{x})_{[1:d]} = \mathbf{x}_{[1:d]}$  and also we have  $(\mathbf{I} - \mathbf{H}_t^{-\frac{1}{2}} \mathbf{w} \mathbf{w}^\top \mathbf{H}_t^{\frac{1}{2}}) \hat{\mathbf{y}}_t = \hat{\mathbf{y}}_t$  because the  $(d+1)$ -th coordinate of  $\hat{\mathbf{y}}_t$  is 0. Combining the above with Eq. (13) yields the following result:

$$\begin{aligned} \mathbb{E}_t[\mathbf{g}_t] &= \left( \mathbf{I} - \frac{\mathbf{e}_{d+1} \mathbf{e}_{d+1}^\top \mathbf{H}_t^{-1}}{\|\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1}\|_2^2} \right) \nabla \hat{\mathbf{f}}_t(\hat{\mathbf{y}}_t) \\ &= \left( \mathbf{I} - \frac{\mathbf{e}_{d+1} \mathbf{e}_{d+1}^\top \mathbf{H}_t^{-1}}{\|\mathbf{H}_t^{-\frac{1}{2}} \mathbf{e}_{d+1}\|_2^2} \right) \left[ \nabla \hat{\mathbf{f}}_t(\mathbf{y}_t)_{[1:d]}; 0 \right] = \left[ \nabla \hat{\mathbf{f}}_t(\mathbf{y}_t)_{[1:d]}; * \right], \end{aligned}$$

where  $* \in \mathbb{R}$  denotes the last coordinate of the expectation of the gradient estimator that can be calculated according to the context. Note that the last step is true by noting that  $\hat{\mathbf{f}}_t$  is defined as a smoothed function of  $\tilde{\mathbf{f}}_t$  that is irrelevant to the  $(d+1)$ -th coordinate.

Hence, we show that the first  $d$  dimensions of the estimator constructed in Line 5 are unbiased and finish the proof.  $\blacksquare$

## A.2. Stability Lemma

In this section, we prove the following lemma which shows the stability of the dynamics of our algorithm. We point out that this stability lemma is the main technical reason that we introduce the lifting idea.

**Lemma 17** *Consider the following FTRL update:*

$$\mathbf{y}_{t+1} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \left\{ \sum_{s=1}^t \left( \mathbf{g}_s^\top \mathbf{x} + \frac{\sigma_s + \lambda_s}{2} \|\mathbf{x} - \mathbf{y}_s\|_2^2 \right) + \frac{\lambda_0}{2} \|\mathbf{x}\|_2^2 + \frac{1}{\eta_{t+1}} \Psi(\mathbf{x}) \right\},$$

where  $\Psi(\mathbf{x}) = \Psi(x, b) = 400(\psi(x/b) - 2\nu \ln b)$  is a normal barrier of the conic hull of  $\mathcal{X}$  defined by  $\operatorname{con}(\mathcal{X}) = \{\mathbf{0}\} \cup \{(w, b) \mid \frac{w}{b} \in \mathcal{X}, w \in \mathbb{R}^d, b > 0\}$ , and  $\psi$  is a  $\nu$ -self-concordant barrier of  $\mathcal{X} \subseteq \mathbb{R}^d$ ,  $\mathbf{H}_t = \nabla^2 \Psi(\mathbf{y}_t) + \eta_t(\sigma_{1:t-1} + \lambda_{0:t-1})$ ,  $\mathbf{u}_s$  is uniformly sampled from  $\mathbb{S}^{d+1} \cap (\mathbf{H}_s^{-\frac{1}{2}} \mathbf{e}_{d+1})^\perp$  and  $\mathbf{g}_s = d(f_s(x_s) + \frac{\lambda_s}{2} \|x_s\|_2^2) \mathbf{H}_s^{\frac{1}{2}} \mathbf{u}_s$  for  $s \in [t]$ ,

Suppose that the following two conditions hold: (1) the sequence of learning rates  $\{\eta_t\}_{t=1}^T$  is non-increasing and satisfies  $\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} \leq C(\lambda_t + \sigma_t)^p$  for some  $C > 0$  and  $p > 0$ ; (2)  $\sigma_t \leq \gamma$  holds for some  $\gamma > 0$  and  $\eta_1 \leq \frac{1}{32(d+16\sqrt{\nu}C(\gamma+1)^p)}$ ,  $\lambda_t \in (0, 1)$  holds for all  $t \in [T]$ , and  $\lambda_0 > 0$ . Then, we have  $\|\mathbf{y}_t - \mathbf{y}_{t+1}\|_{\mathbf{H}_t} \leq \frac{1}{2}$ .

**Proof** Define the objective of FTRL update to be  $\mathbf{F}_{t+1}(\mathbf{x}) = \sum_{s=1}^t \ell_s(\mathbf{x}) + \mathbf{R}_{t+1}(\mathbf{x})$  with  $\mathbf{R}_{t+1}(\mathbf{x}) = \frac{\lambda_0}{2} \|\mathbf{x}\|_2^2 + \frac{1}{\eta_{t+1}} \Psi(\mathbf{x})$  and  $\ell_s(\mathbf{x}) \triangleq \langle \mathbf{g}_s, \mathbf{x} \rangle + \frac{\sigma_s + \lambda_s}{2} \|\mathbf{x} - \mathbf{y}_s\|_2^2$ . Therefore, we have  $\mathbf{y}_{t+1} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \mathbf{F}_{t+1}(\mathbf{x})$ . Define  $\Psi_t(\mathbf{x}) \triangleq \Psi(\mathbf{x}) + \frac{\eta_t \lambda_0}{2} \|\mathbf{x}\|_2^2 + \eta_t \sum_{s=1}^{t-1} \frac{\sigma_s + \lambda_s}{2} \|\mathbf{x} - \mathbf{y}_s\|_2^2$ . With this definition, we have  $\mathbf{H}_t = \nabla^2 \Psi_t(\mathbf{y}_t)$  and  $\mathbf{F}_{t+1}(\mathbf{x}) = \sum_{s=1}^t \langle \mathbf{g}_s, \mathbf{x} \rangle + \frac{1}{\eta_{t+1}} \Psi_{t+1}(\mathbf{x})$ . Moreover, according to the definition of self-concordant function (see [Definition 23](#)), we know that  $\Psi_t$  is also a self-concordant function.

Because of the convexity of  $\mathbf{F}_{t+1}$ , in order to prove the desired conclusion, it suffices to show that for any  $\mathbf{y}' \in \mathcal{X}$  satisfying  $\|\mathbf{y}' - \mathbf{y}_t\|_{\mathbf{H}_t} = \frac{1}{2}$ , we have  $\mathbf{F}_{t+1}(\mathbf{y}') \geq \mathbf{F}_{t+1}(\mathbf{y}_t)$ . To this end, we first calculate  $\nabla \mathbf{F}_{t+1}(\mathbf{y}_t)$  and the Hessian of  $\mathbf{F}_{t+1}$  as follows:

$$\begin{aligned} \nabla \mathbf{F}_{t+1}(\mathbf{y}_t) &= \sum_{s=1}^t \nabla \ell_s(\mathbf{y}_t) + \lambda_0 \mathbf{y}_t + \frac{1}{\eta_{t+1}} \nabla \Psi(\mathbf{y}_t) \\ &= \sum_{s=1}^t \mathbf{g}_s + \sum_{s=1}^t (\sigma_s + \lambda_s) (\mathbf{y}_t - \mathbf{y}_s) + \lambda_0 \mathbf{y}_t + \frac{1}{\eta_{t+1}} \nabla \Psi(\mathbf{y}_t) \\ &= \nabla \mathbf{F}_t(\mathbf{y}_t) + \mathbf{g}_t + \left( \frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} \right) \nabla \Psi(\mathbf{y}_t). \end{aligned} \tag{14}$$

$$\begin{aligned} \nabla^2 \mathbf{F}_{t+1}(\mathbf{x}) &= \frac{1}{\eta_{t+1}} \nabla^2 \Psi_{t+1}(\mathbf{x}) = \frac{1}{\eta_{t+1}} \nabla^2 \Psi(\mathbf{x}) + \left( \lambda_0 \mathbf{I} + \sum_{s=1}^t (\sigma_s + \lambda_s) \mathbf{I} \right) \\ &\succeq \frac{1}{\eta_t} \nabla^2 \Psi(\mathbf{x}) + \left( \lambda_0 \mathbf{I} + \sum_{s=1}^{t-1} (\sigma_s + \lambda_s) \mathbf{I} \right) = \frac{1}{\eta_t} \nabla^2 \Psi_t(\mathbf{x}), \end{aligned} \tag{15}$$

where the inequality is because  $\eta_t \geq \eta_{t+1}$ . Based on the above, using Taylor's expansion of  $\mathbf{F}_{t+1}$  at  $\mathbf{y}_t$ , we know that there exists  $\boldsymbol{\xi}_t$  that lies in the line segment of  $\mathbf{y}_t$  and  $\mathbf{y}'$  such that:

$$\begin{aligned}
 & \mathbf{F}_{t+1}(\mathbf{y}') \\
 &= \mathbf{F}_{t+1}(\mathbf{y}_t) + \mathbf{h}^\top \nabla \mathbf{F}_{t+1}(\mathbf{y}_t) + \frac{1}{2} \|\mathbf{h}\|_{\nabla^2 \mathbf{F}_{t+1}(\boldsymbol{\xi}_t)}^2 \quad (\mathbf{h} \triangleq \mathbf{y}' - \mathbf{y}_t \text{ and } \boldsymbol{\xi}_t \in [\mathbf{y}_t, \mathbf{y}']) \\
 &\geq \mathbf{F}_t(\mathbf{y}_t) + \mathbf{h}^\top \nabla \mathbf{F}_t(\mathbf{y}_t) + \left( \frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} \right) \nabla \Psi(\mathbf{y}_t)^\top \mathbf{h} + \mathbf{g}_t^\top \mathbf{h} + \frac{1}{2\eta_t} \|\mathbf{h}\|_{\nabla^2 \Psi_t(\boldsymbol{\xi}_t)}^2 \\
 &\quad \text{(by Eq. (14) and Eq. (15))} \\
 &\geq \mathbf{F}_t(\mathbf{y}_t) + \mathbf{h}^\top \nabla \mathbf{F}_t(\mathbf{y}_t) + \left( \frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} \right) \nabla \Psi(\mathbf{y}_t)^\top \mathbf{h} + \mathbf{g}_t^\top \mathbf{h} + \frac{1}{2\eta_t} \|\mathbf{h}\|_{\mathbf{H}_t}^2 \cdot (1 - \|\mathbf{y}_t - \boldsymbol{\xi}_t\|_{\mathbf{H}_t})^2 \\
 &\quad (\nabla^2 \Psi_t(\mathbf{y}_t) = \mathbf{H}_t, \Psi_t \text{ is a self-concordant function and by Lemma 25)} \\
 &\geq \mathbf{F}_t(\mathbf{y}_t) - \left( \frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} \right) \left| \nabla \Psi(\mathbf{y}_t)^\top \mathbf{h} \right| - \|\mathbf{g}_t\|_{\mathbf{H}_t}^* \|\mathbf{h}\|_{\mathbf{H}_t} + \frac{1}{2\eta_t} \|\mathbf{h}\|_{\mathbf{H}_t}^2 \cdot (1 - \|\mathbf{y}_t - \boldsymbol{\xi}_t\|_{\mathbf{H}_t})^2 \\
 &\quad \text{(first-order optimality of } \mathbf{y}_t) \\
 &\geq \mathbf{F}_t(\mathbf{y}_t) - \left( \frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} \right) \left| \nabla \Psi(\mathbf{y}_t)^\top \mathbf{h} \right| - \frac{1}{2} \|\mathbf{g}_t\|_{\mathbf{H}_t}^* + \frac{1}{32\eta_t} \quad (\|\mathbf{h}\|_{\mathbf{H}_t} = \frac{1}{2}, \|\mathbf{y}_t - \boldsymbol{\xi}_t\|_{\mathbf{H}_t} \leq \frac{1}{2}) \\
 &\geq \mathbf{F}_t(\mathbf{y}_t) - C(\sigma_t + \lambda_t)^p \left| \nabla \Psi(\mathbf{y}_t)^\top \mathbf{h} \right| - \frac{1}{2} \|\mathbf{g}_t\|_{\mathbf{H}_t}^* + \frac{1}{32\eta_t} \quad \text{(by definition of } \eta_t) \\
 &\geq \mathbf{F}_t(\mathbf{y}_t) - C(\gamma + 1)^p \left| \nabla \Psi(\mathbf{y}_t)^\top \mathbf{h} \right| - \frac{1}{2} \|\mathbf{g}_t\|_{\mathbf{H}_t}^* + \frac{1}{32\eta_1}. \\
 &\quad (\sigma_t \leq \gamma \text{ and } \lambda_t \in (0, 1), \{\eta_t\}_{t=1}^T \text{ is monotonically non-increasing.})
 \end{aligned}$$

Furthermore, note that the third term, which is the gradient local norm, can be upper bounded by

$$\begin{aligned}
 \|\mathbf{g}_t\|_{\mathbf{H}_t}^{*2} &= d^2 \left( \mathbf{f}_t(\mathbf{x}_t) + \frac{\lambda_t}{2} \|\mathbf{x}_t\|_2^2 \right)^2 (\|\mathbf{H}_t^{\frac{1}{2}} \mathbf{u}\|_{\mathbf{H}_t}^*)^2 \\
 &\leq d^2 \left( 1 + \frac{\lambda_t}{2} \right)^2 \mathbf{u}^\top \mathbf{H}_t^{\frac{1}{2}} \mathbf{H}_t^{-1} \mathbf{H}_t^{\frac{1}{2}} \mathbf{u} \leq 4d^2.
 \end{aligned} \tag{16}$$

For the second term, we have

$$\begin{aligned}
 \left| \nabla \Psi(\mathbf{y}_t)^\top \mathbf{h} \right| &\leq \|\nabla \Psi(\mathbf{y}_t)\|_{\nabla^{-2} \Psi(\mathbf{y}_t)} \|\mathbf{h}\|_{\nabla^2 \Psi(\mathbf{y}_t)} \\
 &\leq \|\nabla \Psi(\mathbf{y}_t)\|_{\nabla^{-2} \Psi(\mathbf{y}_t)} \|\mathbf{h}\|_{\nabla^2 \Psi_t(\mathbf{y}_t)} \quad (\nabla^2 \Psi_t(\mathbf{y}) \succeq \nabla^2 \Psi(\mathbf{y})) \\
 &= \|\mathbf{y}_t\|_{\nabla^2 \Psi(\mathbf{y}_t)} \|\mathbf{h}\|_{\nabla^2 \Psi_t(\mathbf{y}_t)} = \frac{\sqrt{\bar{\nu}}}{2}.
 \end{aligned}$$

The last two equations make use of the properties of  $\bar{\nu}$ -normal barrier (see Lemma 29):  $\nabla^2 \Psi(\mathbf{y}_t) \mathbf{y}_t = -\nabla \Psi(\mathbf{y}_t)$  and  $\|\mathbf{y}_t\|_{\nabla^2 \Psi(\mathbf{y}_t)}^2 = \bar{\nu}$ , as well as the fact that  $\|\mathbf{h}\|_{\mathbf{H}_t} = \frac{1}{2}$ . Moreover, the constructed normal barrier satisfies that  $\bar{\nu} = 800\nu$  (see Lemma 28). Therefore, we have

$$\mathbf{F}_{t+1}(\mathbf{y}') \geq \mathbf{F}_{t+1}(\mathbf{y}_t) - \frac{\sqrt{800\nu}}{2} C(\gamma + 1)^p - d + \frac{1}{32\eta_1} \geq \mathbf{F}_{t+1}(\mathbf{y}_t),$$

where the last step is due to the setting of  $\eta_1 \leq \frac{1}{32(d+16\sqrt{\nu}C(\gamma+1)^p)}$ . Hence, we complete the proof.  $\blacksquare$

To apply [Lemma 17](#), when all the functions are  $\beta$ -smooth (see [Definition 1](#)), we can choose  $\gamma = \beta$  to satisfy the condition  $\sigma_t \leq \gamma$ ; when all the functions are  $L$ -Lipschitz (see [Definition 2](#)), we show in [Lemma 31](#) that choosing  $\gamma = 4L$  satisfies the condition of  $\sigma_t \leq \gamma$ .

### A.3. Proof of [Lemma 5](#)

To bound the expected regret, we decompose the cumulative regret with respect to  $x \in \mathcal{X}$  in the following way using the functions in the lifted domain defined in [Lemma 4](#), [Eq. \(11\)](#) and [Eq. \(12\)](#):

$$\begin{aligned}
 & \mathbb{E} \left[ \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(x) \right] \\
 &= \mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}_t) - \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}) \right] \\
 &= \mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}_t) - \sum_{t=1}^T \mathbf{f}_t(\tilde{\mathbf{x}}) \right] + \mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\tilde{\mathbf{x}}) - \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}) \right] \\
 &= \underbrace{\mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}_t) - \sum_{t=1}^T \mathbf{f}_t(\mathbf{y}_t) \right]}_{\text{EXPLORATION}} + \underbrace{\mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\mathbf{y}_t) - \sum_{t=1}^T \tilde{\mathbf{f}}_t(\mathbf{y}_t) \right]}_{\text{REGULARIZATION I}} + \underbrace{\mathbb{E} \left[ \sum_{t=1}^T \tilde{\mathbf{f}}_t(\mathbf{y}_t) - \sum_{t=1}^T \hat{\mathbf{f}}_t(\mathbf{y}_t) \right]}_{\text{SMOOTH I}} \\
 &\quad + \underbrace{\mathbb{E} \left[ \sum_{t=1}^T \hat{\mathbf{f}}_t(\mathbf{y}_t) - \sum_{t=1}^T \hat{\mathbf{f}}_t(\tilde{\mathbf{x}}) \right]}_{\text{REG TERM}} + \underbrace{\mathbb{E} \left[ \sum_{t=1}^T \hat{\mathbf{f}}_t(\tilde{\mathbf{x}}) - \sum_{t=1}^T \tilde{\mathbf{f}}_t(\tilde{\mathbf{x}}) \right]}_{\text{SMOOTH II}} + \underbrace{\mathbb{E} \left[ \sum_{t=1}^T \tilde{\mathbf{f}}_t(\tilde{\mathbf{x}}) - \sum_{t=1}^T \mathbf{f}_t(\tilde{\mathbf{x}}) \right]}_{\text{REGULARIZE II}} \\
 &\quad + \underbrace{\mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\tilde{\mathbf{x}}) - \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}) \right]}_{\text{COMPARATOR BIAS}}, \tag{17}
 \end{aligned}$$

where in the second equality, we define  $\tilde{\mathbf{x}} \triangleq (1 - \frac{1}{T}) \mathbf{x} + \frac{1}{T} \cdot \mathbf{y}_1$ , where  $\mathbf{y}_1 = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \Psi(\mathbf{x})$ . Note that both  $\tilde{\mathbf{x}}$  and  $\mathbf{y}_1$  belong to the shrunk lifted feasible set  $\tilde{\mathcal{X}} = \{\mathbf{x} = (x, 1) \mid x \in \mathcal{X}, \pi_{\mathbf{y}_1}(\mathbf{x}) \leq 1 - \frac{1}{T}\}$ . We remind the readers the notations  $\tilde{x} = \tilde{\mathbf{x}}_{[1:d]}$  and  $y_1 = \mathbf{y}_1_{[1:d]}$ , and we have  $\tilde{\mathbf{x}} = (\tilde{x}, 1)$  and  $\mathbf{y}_1 = (y_1, 1)$ .

We now bound the each term of the regret decomposition in [Eq. \(17\)](#) individually. First, for the two terms [REGULARIZATION I](#) and [REGULARIZATION II](#), we have for any  $\mathbf{x} \in \mathcal{X}$ ,

$$\mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}) - \sum_{t=1}^T \tilde{\mathbf{f}}_t(\mathbf{x}) \right] = \mathbb{E} \left[ \sum_{t=1}^T f_t(x) - \sum_{t=1}^T \tilde{f}_t(x) \right] \leq \sum_{t=1}^T \frac{\lambda_t}{2}, \tag{18}$$

which essentially is the bias due to introducing the regularization term.

Second, consider the two terms [SMOOTH I](#) and [SMOOTH II](#). According to the definition of  $\tilde{\mathbf{f}}_t$  shown in [Lemma 4](#), we know that  $\tilde{\mathbf{f}}_t$  is  $(\beta + \lambda_t)$ -smooth. Using the fact that perturbation  $\mathbf{b}$  has mean  $\mathbf{0}$ , we can bound the two term as follows: for any  $\mathbf{x} \in \mathcal{X}$ ,

$$\mathbb{E}_{\mathbf{b}} \left[ \sum_{t=1}^T \tilde{\mathbf{f}}_t(\mathbf{x} + \mathbf{H}_t^{-\frac{1}{2}} \mathbf{b}) - \sum_{t=1}^T \tilde{\mathbf{f}}_t(\mathbf{x}) \right] \leq \sum_{t=1}^T \frac{\beta + \lambda_t}{2} \left\| \mathbf{H}_t^{-\frac{1}{2}} \mathbf{b} \right\|_2^2$$

$$\leq \sum_{t=1}^T \frac{d(\beta + \lambda_t)}{\sqrt{(\beta + 1)(\sigma_{1:t-1} + \lambda_{0:t-1})}}, \quad (19)$$

where the second inequality is because  $\mathbf{H}_t \succeq \eta_t(\sigma_{1:t-1} + \lambda_{0:t-1})\mathbf{I}$  and  $\eta_t = \frac{1}{2d} \sqrt{\frac{\beta+1}{\sigma_{1:t-1} + \lambda_{0:t-1}} + \frac{\nu}{T \log T}}$ .

Third, by definition of  $\mathbf{y}_t$  and the  $\beta$ -smoothness of function  $f_t$ , EXPLORATION term can be bounded by

$$\mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}_t) - \sum_{t=1}^T \mathbf{f}_t(\mathbf{y}_t) \right] \leq \sum_{t=1}^T \frac{\beta}{2} \left\| \mathbf{H}_t^{-\frac{1}{2}} \mathbf{u}_t \right\|_2^2 \leq \sum_{t=1}^T \frac{d\beta}{\sqrt{(\beta + 1)(\sigma_{1:t-1} + \lambda_{0:t-1})}}. \quad (20)$$

Fourth, for COMPARATOR BIAS, according to the definition of  $\tilde{\mathbf{x}}$  and using the convexity property of  $\mathbf{f}_t$ , we have

$$\begin{aligned} \mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\tilde{\mathbf{x}}) - \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}) \right] &\leq \mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t \left( \frac{1}{T} \mathbf{y}_1 + \left(1 - \frac{1}{T}\right) \mathbf{x} \right) - \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}) \right] \\ &\leq \mathbb{E} \left[ \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t(\mathbf{y}_1) - \frac{1}{T} \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}) \right] \leq 2. \end{aligned} \quad (21)$$

Therefore, it suffices to further bound the REG TERM, which is the expected regret over the smoothed version of the lifted online functions. The following lemma proves the upper bound for the REG TERM. We remark that bounding this REG TERM is the most challenging part of the proof and is also the technical reason for us to lift the domain.

**Lemma 18** *When loss functions  $\{f_t\}_{t=1}^T$  are all  $\beta$ -smooth, if  $T \geq \rho$  (a constant defined in [Algorithm 1](#)), [Algorithm 1](#) guarantees that*

$$\text{REG TERM} \leq \tilde{\mathcal{O}} \left( d\sqrt{\nu T} + \sum_{t=1}^T \frac{d\sqrt{\beta + 1}}{\sqrt{\sigma_{1:t-1} + \lambda_{0:t-1}}} \right). \quad (22)$$

**Proof** According to the definition of REG TERM, we have

$$\mathbb{E} \left[ \sum_{t=1}^T \hat{\mathbf{f}}_t(\mathbf{y}_t) - \sum_{t=1}^T \hat{\mathbf{f}}_t(\tilde{\mathbf{x}}) \right] \leq \mathbb{E} \left[ \sum_{t=1}^T \left( \nabla \hat{\mathbf{f}}_t(\mathbf{y}_t)^\top (\mathbf{y}_t - \tilde{\mathbf{x}}) - \frac{\sigma_t + \lambda_t}{2} \|\mathbf{y}_t - \tilde{\mathbf{x}}\|_2^2 \right) \right] \quad (23)$$

$$= \mathbb{E} \left[ \sum_{t=1}^T \mathbf{g}_t^\top (\mathbf{y}_t - \tilde{\mathbf{x}}) - \frac{\sigma_t + \lambda_t}{2} \|\mathbf{y}_t - \tilde{\mathbf{x}}\|_2^2 \right] \quad (24)$$

$$= \mathbb{E} \left[ \sum_{t=1}^T \ell_t(\mathbf{y}_t) - \sum_{t=1}^T \ell_t(\tilde{\mathbf{x}}) \right]. \quad (25)$$

In above, [Eq. \(23\)](#) holds owing to the  $(\sigma_t + \lambda_t)$ -strong-convexity of  $\hat{\mathbf{f}}_t$  (actually only in the first  $d$  dimension but it is enough as  $\mathbf{y}_t$  and  $\tilde{\mathbf{x}}$  have the same last coordinate); [Eq. \(24\)](#) is true because [Lemma 4](#) ensures that  $\mathbf{g}_t$  is an unbiased estimator of  $\nabla \hat{\mathbf{f}}_t(\mathbf{y}_t)$  in the first  $d$  coordinates and meanwhile  $\mathbf{y}_t - \tilde{\mathbf{x}}$  has the last coordinate 0. The last step shown in [Eq. \(25\)](#) is by introducing the surrogate loss

$\ell_t : \mathcal{X} \mapsto \mathbb{R}$ , defined as  $\ell_t(\mathbf{x}) \triangleq \langle \mathbf{g}_t, \mathbf{x} \rangle + \frac{\sigma_t + \lambda_t}{2} \|\mathbf{x} - \mathbf{y}_t\|_2^2$ . Note that according to this construction, we have  $\nabla \ell_t(\mathbf{y}_t) = \mathbf{g}_t$ .

In addition, our FTRL update rule can be written in the following two forms:

$$\begin{aligned} \mathbf{y}_{t+1} &= \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \left\{ \sum_{s=1}^t \left( \langle \mathbf{g}_s, \mathbf{x} \rangle + \frac{\sigma_s + \lambda_s}{2} \|\mathbf{y}_s - \mathbf{x}\|_2^2 \right) + \frac{\lambda_0}{2} \|\mathbf{x}\|_2^2 + \frac{1}{\eta_{t+1}} \Psi(\mathbf{x}) \right\} \\ &= \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \left\{ \sum_{s=1}^t \ell_s(\mathbf{x}) + \frac{\lambda_0}{2} \|\mathbf{x}\|_2^2 + \frac{1}{\eta_{t+1}} \Psi(\mathbf{x}) \right\} \\ &= \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \left\{ \sum_{s=1}^t \langle \mathbf{g}_s, \mathbf{x} \rangle + \frac{1}{\eta_{t+1}} \Psi_{t+1}(\mathbf{x}) \right\}, \end{aligned}$$

where  $\Psi_{t+1}(\mathbf{x}) = \Psi(\mathbf{x}) + \eta_{t+1} \left( \frac{\lambda_0}{2} \|\mathbf{x}\|_2^2 + \sum_{s=1}^t \frac{\sigma_s + \lambda_s}{2} \|\mathbf{x} - \mathbf{y}_s\|_2^2 \right)$ . As discussed in [Lemma 17](#)  $\Psi_{t+1}$  is still a self-concordant function and moreover  $\mathbf{H}_{t+1} = \nabla^2 \Psi_{t+1}(\mathbf{y}_{t+1})$ .

Recall the definition  $\mathbf{R}_{t+1}(\mathbf{x}) = \frac{\lambda_0}{2} \|\mathbf{x}\|_2^2 + \frac{1}{\eta_{t+1}} \Psi(\mathbf{x})$  and  $\mathbf{F}_{t+1}(\mathbf{x}) = \sum_{s=1}^t \ell_s(\mathbf{x}) + \mathbf{R}_{t+1}(\mathbf{x})$ . Denote by  $\mathbf{R}'_{t+1}(\mathbf{x}) = \frac{\lambda_0}{2} \|\mathbf{x}\|_2^2 + \frac{1}{\eta_{t+1}} (\Psi(\mathbf{x}) - \Psi(\mathbf{y}_1))$  the (shifted) regularizer and by  $\mathbf{Q}_{t+1}(\mathbf{x}) = \sum_{s=1}^t \ell_s(\mathbf{x}) + \mathbf{R}'_{t+1}(\mathbf{x})$  and (shifted) FTRL objective. Therefore, we have  $\mathbf{F}_{t+1}(\mathbf{x}) = \mathbf{Q}_{t+1}(\mathbf{x}) + \frac{1}{\eta_{t+1}} \Psi(\mathbf{y}_1)$ . Then,  $\mathbf{y}_{t+1} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \mathbf{F}_{t+1}(\mathbf{x}) = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \mathbf{Q}_{t+1}(\mathbf{x})$  according to the FTRL update rule, and we have

$$\begin{aligned} & \sum_{t=1}^T \ell_t(\mathbf{y}_t) - \sum_{t=1}^T \ell_t(\tilde{\mathbf{x}}) \\ & \leq \mathbf{R}'_{T+1}(\tilde{\mathbf{x}}) - \mathbf{R}'_1(\mathbf{y}_1) + \sum_{t=1}^T \nabla \ell_t(\mathbf{y}_t)^\top (\mathbf{y}_t - \mathbf{y}_{t+1}) \\ & \quad - \sum_{t=1}^T D_{\mathbf{Q}_t + \ell_t}(\mathbf{y}_{t+1}, \mathbf{y}_t) + \sum_{t=1}^T \left( \mathbf{R}'_t(\mathbf{y}_{t+1}) - \mathbf{R}'_{t+1}(\mathbf{y}_{t+1}) \right) \\ & \leq \mathbf{R}'_{T+1}(\tilde{\mathbf{x}}) - \mathbf{R}'_1(\mathbf{y}_1) + \sum_{t=1}^T \mathbf{g}_t^\top (\mathbf{y}_t - \mathbf{y}_{t+1}) - \sum_{t=1}^T D_{\mathbf{Q}_t + \ell_t}(\mathbf{y}_{t+1}, \mathbf{y}_t) \\ & \leq \mathbf{R}'_{T+1}(\tilde{\mathbf{x}}) - \mathbf{R}'_1(\mathbf{y}_1) + \sum_{t=1}^T \mathbf{g}_t^\top (\mathbf{y}_t - \mathbf{y}_{t+1}) - \sum_{t=1}^T D_{\mathbf{Q}_t}(\mathbf{y}_{t+1}, \mathbf{y}_t). \end{aligned}$$

In above, the first inequality is due to the standard FTRL analysis as shown in [Lemma 30](#); the second inequality is true because the surrogate loss satisfies that  $\nabla \ell_t(\mathbf{y}_t) = \mathbf{g}_t$  and  $0 \leq \mathbf{R}'_t(\mathbf{x}) \leq \mathbf{R}'_{t+1}(\mathbf{x})$  holds for any  $\mathbf{x} \in \mathcal{X}$  as the learning rate is monotonically non-increasing and  $\mathbf{y}_1 = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \Psi(\mathbf{x})$ . The last inequality follows from  $\nabla^2 \ell_t(\mathbf{x}) = (\sigma_t + \lambda_t) \mathbf{I}$  and the following inequality:

$$\begin{aligned} D_{\mathbf{Q}_t + \ell_t}(\mathbf{y}_{t+1}, \mathbf{y}_t) &= D_{\mathbf{Q}_t}(\mathbf{y}_{t+1}, \mathbf{y}_t) + D_{\ell_t}(\mathbf{y}_{t+1}, \mathbf{y}_t) \\ &= D_{\mathbf{Q}_t}(\mathbf{y}_{t+1}, \mathbf{y}_t) + \frac{\sigma_t + \lambda_t}{2} \|\mathbf{y}_{t+1} - \mathbf{y}_t\|_2^2 \geq D_{\mathbf{Q}_t}(\mathbf{y}_{t+1}, \mathbf{y}_t). \end{aligned}$$

In addition, by Taylor expansion, we know that  $D_{\mathbf{Q}_t}(\mathbf{y}_{t+1}, \mathbf{y}_t) = \frac{1}{2} \|\mathbf{y}_{t+1} - \mathbf{y}_t\|_{\nabla^2 \mathbf{Q}_t(\boldsymbol{\xi}_t)}^2$  for some  $\boldsymbol{\xi}_t \in [\mathbf{y}_t, \mathbf{y}_{t+1}]$ , and  $\nabla^2 \mathbf{Q}_t(\mathbf{x}) = \nabla^2 \mathbf{F}_t(\mathbf{x}) = \frac{1}{\eta_t} \nabla^2 \Psi_t(\mathbf{x})$  as shown in the first equality of Eq. (15). Therefore, combining all above, we get that

$$\begin{aligned} & \sum_{t=1}^T \ell_t(\mathbf{y}_t) - \sum_{t=1}^T \ell_t(\tilde{\mathbf{x}}) \\ & \leq \frac{\lambda_0}{2} \|\tilde{\mathbf{x}}\|_2^2 + \frac{\Psi(\tilde{\mathbf{x}}) - \Psi(\mathbf{y}_1)}{\eta_{T+1}} + \sum_{t=1}^T \mathbf{g}_t^\top (\mathbf{y}_t - \mathbf{y}_{t+1}) - \sum_{t=1}^T \frac{1}{2\eta_t} \|\mathbf{y}_{t+1} - \mathbf{y}_t\|_{\nabla^2 \Psi_t(\boldsymbol{\xi}_t)}^2. \end{aligned} \quad (26)$$

In the following, we proceed to analyze the crucial terms  $\mathbf{g}_t^\top (\mathbf{y}_t - \mathbf{y}_{t+1})$  and  $\|\mathbf{y}_{t+1} - \mathbf{y}_t\|_{\nabla^2 \Psi_t(\boldsymbol{\xi}_t)}^2$ . For the first term, by Holder's inequality, we have

$$\mathbf{g}_t^\top (\mathbf{y}_t - \mathbf{y}_{t+1}) \leq \|\mathbf{g}_t\|_{\mathbf{H}_t}^* \cdot \|\mathbf{y}_t - \mathbf{y}_{t+1}\|_{\mathbf{H}_t}. \quad (27)$$

The second term is more involved to analyze. To do this, we first verify that the conditions required in Lemma 17 are indeed satisfied. First, it is direct to see that  $\{\eta_t\}_{t=1}^T$  is non-increasing and

$$\begin{aligned} \frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} & \leq 2d \left( \frac{1}{\sqrt{\frac{\beta+1}{\lambda_{0:t} + \sigma_{1:t}} + \frac{\nu}{T \log T}}} - \frac{1}{\sqrt{\frac{\beta+1}{\lambda_{0:t-1} + \sigma_{1:t-1}} + \frac{\nu}{T \log T}}} \right) \\ & \leq \frac{2d}{\sqrt{\beta+1}} \left( \sqrt{\lambda_{0:t} + \sigma_{1:t}} - \sqrt{\lambda_{0:t-1} + \sigma_{1:t-1}} \right) \\ & \leq \frac{2d\sqrt{\lambda_t + \sigma_t}}{\sqrt{\beta+1}}, \end{aligned}$$

where the second inequality is because  $(a+c)^{-1/2} - (b+c)^{-1/2}$  is decreasing in  $c$  when  $a \leq b$ . Therefore, this satisfies that  $\eta_{t+1}^{-1} - \eta_t^{-1} \leq C(\lambda_t + \sigma_t)^p$  with  $C = 2d(\beta+1)^{-1/2}$  and  $p = \frac{1}{2}$ . In addition, note that as  $T \geq \rho = 512\nu(1 + 32\sqrt{\nu})^2$  and  $\lambda_0 \geq (\beta+1)\rho\nu^{-1}$ , we have

$$\eta_1 = \frac{1}{2d} \sqrt{\frac{\beta+1}{\lambda_0} + \frac{\nu}{T \log T}} \leq \frac{1}{2d} \sqrt{\frac{2\nu}{\rho}} = \frac{1}{32d(1 + 32\sqrt{\nu})} = \frac{1}{32(d + 16\sqrt{\nu}C(\gamma+1)^p)},$$

with  $\gamma = \beta$ . Therefore, according to Lemma 17, we show that  $\|\mathbf{y}_t - \mathbf{y}_{t+1}\|_{\mathbf{H}_t} \leq \frac{1}{2}$ . Then, due to the nice properties of optimization with self-concordant functions (see Lemma 25), we obtain that

$$\|\mathbf{y}_{t+1} - \mathbf{y}_t\|_{\nabla^2 \Psi_t(\boldsymbol{\xi}_t)} \geq \|\mathbf{y}_{t+1} - \mathbf{y}_t\|_{\nabla^2 \Psi_t(\mathbf{y}_t)} \cdot (1 - \|\mathbf{y}_{t+1} - \boldsymbol{\xi}_t\|_{\nabla^2 \Psi_t(\mathbf{y}_t)}) \geq \frac{1}{2} \|\mathbf{y}_{t+1} - \mathbf{y}_t\|_{\mathbf{H}_t}, \quad (28)$$

where the last inequality makes use of the result  $\|\mathbf{y}_t - \mathbf{y}_{t+1}\|_{\mathbf{H}_t} \leq \frac{1}{2}$  as well as the fact that  $\nabla^2 \Psi_t(\mathbf{y}_t) = \mathbf{H}_t$ . Plugging inequalities Eq. (27) and Eq. (28) to the regret upper bound achieves

$$\begin{aligned} & \sum_{t=1}^T \ell_t(\mathbf{y}_t) - \sum_{t=1}^T \ell_t(\tilde{\mathbf{x}}) \\ & \leq \frac{\lambda_0}{2} \|\tilde{\mathbf{x}}\|_2^2 + \frac{\Psi(\tilde{\mathbf{x}}) - \Psi(\mathbf{y}_1)}{\eta_{T+1}} + \sum_{t=1}^T \mathbf{g}_t^\top (\mathbf{y}_t - \mathbf{y}_{t+1}) - \sum_{t=1}^T \frac{1}{2\eta_t} \|\mathbf{y}_{t+1} - \mathbf{y}_t\|_{\nabla^2 \Psi_t(\boldsymbol{\xi}_t)}^2 \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{\lambda_0}{2} \|\tilde{\mathbf{x}}\|_2^2 + \frac{\Psi(\tilde{\mathbf{x}}) - \Psi(\mathbf{y}_1)}{\eta_{T+1}} + \sum_{t=1}^T \left( \|\mathbf{g}_t\|_{\mathbf{H}_t}^* \cdot \|\mathbf{y}_t - \mathbf{y}_{t+1}\|_{\mathbf{H}_t} - \frac{1}{8\eta_t} \|\mathbf{y}_{t+1} - \mathbf{y}_t\|_{\mathbf{H}_t}^2 \right) \\
 &\leq \frac{\lambda_0}{2} \|\tilde{\mathbf{x}}\|_2^2 + \frac{\Psi(\tilde{\mathbf{x}}) - \Psi(\mathbf{y}_1)}{\eta_{T+1}} + \sum_{t=1}^T 2\eta_t \|\mathbf{g}_t\|_{\mathbf{H}_t}^{*2} \tag{29}
 \end{aligned}$$

$$\leq \mathcal{O}(1) + \frac{\Psi(\tilde{\mathbf{x}}) - \Psi(\mathbf{y}_1)}{\eta_{T+1}} + \sum_{t=1}^T 8\eta_t d^2, \tag{Eq. (16)}$$

$$\begin{aligned}
 &\leq \mathcal{O}(1) + \frac{\Psi(\tilde{\mathbf{x}}) - \Psi(\mathbf{y}_1)}{\eta_{T+1}} + \sum_{t=1}^T \frac{8d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t-1} + \lambda_{0:t-1}}} + \mathcal{O}\left(d\sqrt{\nu T \log T}\right), \\
 &\leq \mathcal{O}(1) + \mathcal{O}\left(d\sqrt{\nu T \log T}\right) + \sum_{t=1}^T \frac{8d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t-1} + \lambda_{0:t-1}}}, \tag{30}
 \end{aligned}$$

where Eq. (30) holds because of the following three facts. First, as both  $\tilde{\mathbf{x}}$  and  $\mathbf{y}_1$  belong to the shrunk lifted domain  $\tilde{\mathcal{X}} = \{\mathbf{x} \mid \pi_{\mathbf{y}_1}(\mathbf{x}) \leq 1 - \frac{1}{T}\}$ , based on Lemma 24, we have that  $0 \leq \Psi(\mathbf{x}) - \Psi(\mathbf{y}_1) \leq \bar{\nu} \ln\left(\frac{1}{1 - \pi_{\mathbf{y}_1}(\mathbf{x})}\right) \leq \bar{\nu} \log T$  holds for all  $\mathbf{x} \in \tilde{\mathcal{X}}$ . Second, as demonstrated in Lemma 28, the normal barrier we choose in Algorithm 1 ensures that  $\bar{\nu} = 800\nu = \mathcal{O}(\nu)$ . Third,  $\eta_{T+1} \geq \frac{1}{2d} \sqrt{\frac{\nu}{T \log T}}$ . This finishes the proof of Lemma 18.  $\blacksquare$

Now we are ready to prove our main lemma (Lemma 5). Below we restate the lemma for convenience.

**Lemma 19** *With any regularization coefficients  $\{\lambda_t\}_{t=1}^T \in (0, 1)$ , Algorithm 1 guarantees:*

$$\text{REG} = \tilde{\mathcal{O}} \left( d\sqrt{\nu T} + \lambda_{1:T-1} + \sum_{t=1}^{T-1} \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{0:t}}} \right), \tag{31}$$

if loss functions  $\{f_t\}_{t=1}^T$  are all  $\beta$ -smooth and  $T \geq \rho$  (a constant defined in Algorithm 1).

**Proof** Combining all the above terms in Eq. (18), Eq. (19), Eq. (20), Eq. (21), and Eq. (22) as well as the decomposition in Eq. (17), we obtain the following expected regret upper bound:

$$\begin{aligned}
 &\mathbb{E} \left[ \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(x) \right] \\
 &\stackrel{(17)}{\leq} \mathbb{E} \left[ \sum_{t=1}^T \hat{\mathbf{f}}_t(\mathbf{y}_t) - \sum_{t=1}^T \hat{\mathbf{f}}_t(\tilde{\mathbf{x}}) \right] + 2 \\
 &\quad + \mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}_t) - \sum_{t=1}^T \mathbf{f}_t(\mathbf{y}_t) \right] + \mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\mathbf{y}_t) - \sum_{t=1}^T \tilde{\mathbf{f}}_t(\mathbf{y}_t) \right] + \mathbb{E} \left[ \sum_{t=1}^T \tilde{\mathbf{f}}_t(\mathbf{y}_t) - \sum_{t=1}^T \hat{\mathbf{f}}_t(\mathbf{y}_t) \right] \\
 &\quad + \mathbb{E} \left[ \sum_{t=1}^T \hat{\mathbf{f}}_t(\tilde{\mathbf{x}}) - \sum_{t=1}^T \tilde{\mathbf{f}}_t(\tilde{\mathbf{x}}) \right] + \mathbb{E} \left[ \sum_{t=1}^T \tilde{\mathbf{f}}_t(\tilde{\mathbf{x}}) - \sum_{t=1}^T \mathbf{f}_t(\tilde{\mathbf{x}}) \right] \\
 &\leq \mathcal{O} \left( \lambda_{1:T} + d\sqrt{\nu T \log T} + \sum_{t=1}^T \frac{d(\beta + \lambda_t)}{\sqrt{\beta + 1} \sqrt{\sigma_{1:t-1} + \lambda_{0:t-1}}} + \sum_{t=1}^T \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t-1} + \lambda_{0:t-1}}} \right)
 \end{aligned}$$



$$\begin{aligned}
 &\leq \tilde{\mathcal{O}} \left( \lambda_{1:T} + d\sqrt{\nu T} + \sum_{t=1}^T \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t-1} + \lambda_{0:t-1}}} \right) && (\lambda_t \in (0, 1)) \\
 &\leq \tilde{\mathcal{O}} \left( \lambda_{1:T-1} + d\sqrt{\nu T} + \sum_{t=1}^{T-1} \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{0:t}}} \right), && (32)
 \end{aligned}$$

where the last step hold because our choice of regularization coefficients  $\lambda_t \in (0, 1)$  for  $t \in [T]$  and the input parameter  $\lambda_0 \geq 1$ , which finishes the proof.  $\blacksquare$

#### A.4. Proof of Lemma 6

**Proof** We prove the claim Eq. (7) by induction, whose proof technique is similar to (Bartlett et al., 2007, Lemma 3.1).

Consider the base case when  $t = 1$ . For simplicity, we define  $\lambda_0^* = \lambda_0$ . If  $\lambda_1 \leq \lambda_1^* \triangleq \operatorname{argmin}_{\lambda_1 \geq 0} \mathcal{B}(\lambda_1)$ , we have  $\mathcal{B}(\lambda_1) = \lambda_1 + \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:1} + \lambda_{0:1}}} = 2\lambda_1 \leq 2\lambda_1^* \leq 2\mathcal{B}(\lambda_1^*)$ , where the second equality is true because of the condition in Eq. (6). Otherwise, we have  $\mathcal{B}(\lambda_1) = \frac{2d\sqrt{\beta+1}}{\sqrt{\sigma_{1:1} + \lambda_{0:1}}} \leq \frac{2d\sqrt{\beta+1}}{\sqrt{\sigma_{1:1} + \lambda_{0:1}^*}} \leq 2\mathcal{B}(\lambda_1^*)$ . Combining both scenarios verifies the base case.

Suppose we have  $\mathcal{B}(\{\lambda_s\}_{s=1}^{t-1}) \leq 2 \min_{\{\lambda'_s\}_{s=1}^{t-1} \geq 0} \mathcal{B}(\{\lambda'_s\}_{s=1}^{t-1})$ . With a slight abuse of notation, we set  $\{\lambda_s^*\}_{s=1}^t = \operatorname{argmin}_{\{\lambda'_s\}_{s=1}^t \geq 0} \mathcal{B}(\{\lambda'_s\}_{s=1}^t)$ . Similarly, if  $\lambda_{1:t} \leq \lambda_{1:t}^*$ , we have

$$\mathcal{B}(\{\lambda_s\}_{s=1}^t) = \lambda_{1:t} + \sum_{s=1}^t \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:s} + \lambda_{0:s}}} = \lambda_{1:t} + \sum_{s=1}^t \lambda_s = 2\lambda_{1:t} \leq 2\lambda_{1:t}^* \leq 2\mathcal{B}(\{\lambda_s^*\}_{s=1}^t).$$

Otherwise, we have

$$\lambda_t + \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{0:t}}} = \frac{2d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{0:t}}} \leq \frac{2d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{0:t}^*}} \leq 2 \left( \lambda_t^* + \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{1:t}^*}} \right).$$

Using the induction hypothesis, we have

$$\begin{aligned}
 \mathcal{B}(\{\lambda_s\}_{s=1}^t) &= \mathcal{B}(\{\lambda_s\}_{s=1}^{t-1}) + \lambda_t + \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{0:t}}} \\
 &\leq \min_{\{\lambda'_s\}_{s=1}^{t-1} \geq 0} 2\mathcal{B}(\{\lambda'_s\}_{s=1}^{t-1}) + 2 \left( \lambda_t^* + \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{1:t}^*}} \right) \\
 &\leq 2\mathcal{B}(\{\lambda_s^*\}_{s=1}^{t-1}) + 2 \left( \lambda_t^* + \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{1:t}^*}} \right) \\
 &= 2\mathcal{B}(\{\lambda_s^*\}_{s=1}^t),
 \end{aligned}$$

where the first inequality is because of the induction hypothesis. Combining both cases, we have that  $\mathcal{B}(\{\lambda_s\}_{s=1}^t) \leq 2\mathcal{B}(\{\lambda_s^*\}_{s=1}^t)$ .  $\blacksquare$

### A.5. Proof of Theorem 7

**Proof** By Lemma 5 we have

$$\text{REG} \leq \tilde{\mathcal{O}} \left( \lambda_{1:T-1} + d\sqrt{\nu T} + \sum_{t=1}^{T-1} \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{0:t}}} \right),$$

which holds for any sequence of regularization coefficients  $\lambda_1, \dots, \lambda_T \in (0, 1)$ . Moreover, due to the specific calculation of regularization coefficients (see Eq. (6)) and from Lemma 6, we immediately achieve for any  $\lambda_1^*, \dots, \lambda_T^* \geq 0$ ,

$$\text{REG} \leq \tilde{\mathcal{O}} \left( \lambda_{1:T-1}^* + d\sqrt{\nu T} + \sum_{t=1}^{T-1} \frac{d\sqrt{\beta+1}}{\sqrt{\sigma_{1:t} + \lambda_{0:t}^*}} \right),$$

which finishes the proof of Theorem 7.  $\blacksquare$

### A.6. Proofs for Implications of Theorem 7

In this section, we provide the proofs of implications in Section 3.2.

**Proof** [of Corollary 8] Since Theorem 7 holds for any non-negative sequence of  $\{\lambda_t^*\}_{t=1}^T$ , in particular, we choose  $\lambda_1^* = (1 + \beta)^{\frac{1}{3}} d^{\frac{2}{3}} T^{\frac{2}{3}}$  and  $\lambda_t^* = 0$  for all  $t \geq 2$ , then with  $\nu = \mathcal{O}(d)$ , we obtain that

$$\text{REG} \stackrel{(8)}{\leq} \tilde{\mathcal{O}} \left( (1 + \beta)^{\frac{1}{3}} d^{\frac{2}{3}} T^{\frac{2}{3}} + d^{\frac{3}{2}} \sqrt{T} + (1 + \beta)^{\frac{1}{3}} d^{\frac{2}{3}} T^{\frac{2}{3}} \right) = \tilde{\mathcal{O}} \left( d^{\frac{3}{2}} \sqrt{T} + (1 + \beta)^{\frac{1}{3}} d^{\frac{2}{3}} T^{\frac{2}{3}} \right),$$

where the last step holds due to  $\nu = \mathcal{O}(d)$  (see Lemma 26).  $\blacksquare$

**Proof** [of Corollary 9] Since Theorem 7 holds for any non-negative sequence of  $\{\lambda_t^*\}_{t=1}^T$ , in particular, we choose  $\lambda_t^* = 0$  for all  $t \geq 1$  and, then with  $\nu = \mathcal{O}(d)$ , we obtain that

$$\text{REG} \stackrel{(8)}{\leq} \tilde{\mathcal{O}} \left( d\sqrt{\nu T} + d\sqrt{\frac{(1 + \beta)T}{\sigma}} \right) = \tilde{\mathcal{O}} \left( d^{\frac{3}{2}} \sqrt{T} + d\sqrt{\frac{T(1 + \beta)}{\sigma}} \right),$$

which ends the proof.  $\blacksquare$

**Proof** [of Corollary 10] In the first environment where there are  $M$  rounds in which the loss function is 0-strongly convex, to make the right hand side of Eq. (8) the largest, we have  $\sigma_\tau = 0$  when  $\tau \in [M]$  and  $\sigma_\tau = \sigma$  when  $\tau > M$ . Set  $\lambda_t^* = 0$  for all  $t \geq 2$ . According to Eq. (8) shown in Theorem 7 and the choice  $\nu = \mathcal{O}(d)$ , we have

$$\begin{aligned} \text{REG} &\leq \tilde{\mathcal{O}} \left( \lambda_1^* + d\sqrt{\nu T} + \sum_{t=1}^{T-1} \frac{d\sqrt{1 + \beta}}{\sqrt{\lambda_1^* + \sigma_{1:t}}} \right) \\ &\leq \tilde{\mathcal{O}} \left( \lambda_1^* + d^{\frac{3}{2}} \sqrt{T} + \frac{d\sqrt{1 + \beta}M}{\sqrt{\lambda_1^*}} + d\sqrt{1 + \beta} \min \left\{ \frac{(T - M)}{\sqrt{\lambda_1^*}}, \sqrt{\frac{T - M}{\sigma}} \right\} \right) \end{aligned}$$

$$\leq \tilde{\mathcal{O}} \left( d^{\frac{2}{3}}(1+\beta)^{\frac{1}{3}}M^{\frac{2}{3}} + d^{\frac{3}{2}}\sqrt{T} + d\sqrt{1+\beta} \min \left\{ \frac{T-M}{(1+\beta)^{\frac{1}{6}}d^{\frac{1}{3}}M^{\frac{1}{3}}}, \sqrt{\frac{T-M}{\sigma}} \right\} \right),$$

(choosing  $\lambda_1^* = (1+\beta)^{\frac{1}{3}}d^{\frac{2}{3}}M^{\frac{2}{3}}$ )

which leads to the first regret bound. Next, we consider the second environment where the first  $T - M$  loss functions are  $\sigma$ -strongly convex. Similarly, we choose  $\lambda_t^* = 0$  for  $t \geq 2$  and we have our regret bounded as follows:

$$\begin{aligned} \text{REG} &\leq \tilde{\mathcal{O}} \left( \lambda_1^* + d\sqrt{\nu T} + \sum_{t=1}^{T-1} \frac{d\sqrt{1+\beta}}{\sqrt{\lambda_1^* + \sigma_{1:t}}} \right) \\ &\leq \tilde{\mathcal{O}} \left( \lambda_1^* + d^{\frac{3}{2}}\sqrt{T} + \frac{dM\sqrt{1+\beta}}{\sqrt{(T-M)\sigma + \lambda_1^*}} + d\sqrt{1+\beta} \min \left\{ \frac{T-M}{\sqrt{\lambda_1^*}}, \sqrt{\frac{T-M}{\sigma}} \right\} \right). \end{aligned}$$

When  $T - M = \Theta(T)$ , we have

$$\begin{aligned} \text{REG} &\leq \tilde{\mathcal{O}} \left( \lambda_1^* + d^{\frac{3}{2}}\sqrt{T} + \frac{dM\sqrt{1+\beta}}{\sqrt{T\sigma + \lambda_1^*}} + d\sqrt{1+\beta}\sqrt{\frac{T}{\sigma}} \right) \\ &\leq \tilde{\mathcal{O}} \left( \lambda_1^* + d^{\frac{3}{2}}\sqrt{T} + \frac{dM\sqrt{1+\beta}}{\sqrt{T\sigma}} + d\sqrt{1+\beta}\sqrt{\frac{T}{\sigma}} \right) \\ &\leq \tilde{\mathcal{O}} \left( d^{\frac{3}{2}}\sqrt{T} + \frac{dT\sqrt{1+\beta}}{\sqrt{\sigma(T-M)}} \right), \end{aligned}$$

where the last inequality is by choosing  $\lambda_1^* = 0$ . When  $T - M = o(T)$ , we have  $M = \Theta(T)$ . Furthermore, if  $\lambda_1^* \leq \sigma(T - M)$ , we have  $\frac{T-M}{\sqrt{\lambda_1^*}} \geq \sqrt{\frac{T-M}{\sigma}}$  and therefore,

$$\begin{aligned} \text{REG} &\leq \tilde{\mathcal{O}} \left( \lambda_1^* + d^{\frac{3}{2}}\sqrt{T} + \frac{dT\sqrt{1+\beta}}{\sqrt{(T-M)\sigma}} + d\sqrt{1+\beta}\sqrt{\frac{T-M}{\sigma}} \right) \\ &\leq \tilde{\mathcal{O}} \left( d^{\frac{3}{2}}\sqrt{T} + \frac{dT\sqrt{\beta+1}}{\sqrt{\sigma(T-M)}} \right), \end{aligned}$$

where the last inequality is by choosing  $\lambda_1^* = 0$ . On the other hand, if  $\lambda_1^* \geq \sigma(T - M)$ , we have

$$\text{REG} \leq \tilde{\mathcal{O}} \left( \lambda_1^* + d^{\frac{3}{2}}\sqrt{T} + \frac{dT\sqrt{1+\beta}}{\sqrt{\lambda_1^*}} \right) \leq \tilde{\mathcal{O}} \left( \sigma(T-M) + d^{\frac{2}{3}}(1+\beta)^{\frac{1}{3}}T^{\frac{2}{3}} + d^{\frac{3}{2}}\sqrt{T} \right),$$

where the last inequality is by choosing  $\lambda_1^* = \max \left\{ \sigma(T-M), d^{\frac{2}{3}}(1+\beta)^{\frac{1}{3}}T^{\frac{2}{3}} \right\}$ . Combining the above bounds, we have

$$\text{REG} \leq \tilde{\mathcal{O}} \left( d^{\frac{3}{2}}\sqrt{T} + \min \left\{ \frac{dT\sqrt{1+\beta}}{\sqrt{\sigma(T-M)}}, \sigma(T-M) + d^{\frac{2}{3}}(1+\beta)^{\frac{1}{3}}T^{\frac{2}{3}} \right\} \right),$$

which finishes the proof. ■

**Proof** [of [Corollary 11](#)] Since [Theorem 7](#) holds for any sequence of  $\{\lambda_t^*\}_{t=1}^T$ , in particular, we choose  $\lambda_t^* = 0$  for all  $t \geq 2$  and set  $\lambda_1^* = (1 + \beta)^{\mu_0} d^{\mu_1} \cdot T^{\mu_2}$  with  $\mu_2 < 1$ , then we obtain that

$$\begin{aligned} \text{REG} &\leq \tilde{\mathcal{O}} \left( \lambda_1^* + d\sqrt{\nu T} + d\sqrt{\beta + 1} \sum_{t=1}^{T-1} \frac{1}{\sqrt{t^{1-\alpha} + \lambda_1^*}} \right) \\ &\leq \tilde{\mathcal{O}} \left( \lambda_1^* + d^{\frac{3}{2}}\sqrt{T} + d\sqrt{\beta + 1} \min \left\{ T^{\frac{1}{2} + \frac{\alpha}{2}}, \frac{T}{\sqrt{\lambda_1^*}} \right\} \right) \\ &= \tilde{\mathcal{O}} \left( (1 + \beta)^{\mu_0} d^{\mu_1} T^{\mu_2} + d^{\frac{3}{2}}\sqrt{T} + d\sqrt{\beta + 1} \min \left\{ T^{\frac{1+\alpha}{2}}, (1 + \beta)^{-\frac{\mu_0}{2}} d^{-\frac{\mu_1}{2}} T^{1-\frac{\mu_2}{2}} \right\} \right). \end{aligned}$$

First, the above bound can be upper bounded by

$$\text{REG} \leq \tilde{\mathcal{O}} \left( d^{\frac{3}{2}}\sqrt{T} + (1 + \beta)^{\mu_0} d^{\mu_1} T^{\mu_2} + (1 + \beta)^{\frac{1-\mu_0}{2}} d^{1-\frac{\mu_1}{2}} T^{1-\frac{\mu_2}{2}} \right) \leq \tilde{\mathcal{O}} \left( d^{\frac{3}{2}}\sqrt{T} + (1 + \beta)^{\frac{1}{3}} d^{\frac{2}{3}} T^{\frac{2}{3}} \right),$$

where the last inequality is by choosing  $\mu_0 = \frac{1}{3}$ ,  $\mu_1 = \frac{2}{3}$  and  $\mu_2 = \frac{2}{3}$ . Furthermore, when  $\alpha \in [0, \frac{1}{3} - \frac{1}{3} \log_T(\beta + 1) - \frac{2}{3} \log_T d]$ , we have  $T^{\frac{1+\alpha}{2}} \leq (1 + \beta)^{-\frac{1}{6}} d^{-\frac{1}{3}} T^{\frac{2}{3}}$ . Therefore, set  $\mu_0 = \mu_1 = \mu_2 = 0$  and we have

$$\text{REG} \leq \tilde{\mathcal{O}} \left( d^{\frac{3}{2}}\sqrt{T} + d\sqrt{1 + \beta} T^{\frac{1+\alpha}{2}} \right).$$

Combining both situations finishes the proof. ■

## Appendix B. Omitted Details for [Section 4](#)

In this section, we show the proof in the Lipschitz BCO setting. Specifically, we show the proof for the main theorem of Lipschitz BCO in [Appendix B.1](#) and show the proofs for the implications of [Theorem 12](#) in [Appendix B.2](#).

### B.1. Proof of [Theorem 12](#)

Following the same regret decomposition as [Eq. \(17\)](#), we decompose the regret into the following terms where  $\tilde{\mathbf{x}}$  is defined the same as the one in [Eq. \(17\)](#).

$$\begin{aligned} &\mathbb{E} \left[ \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(x) \right] \\ &= \underbrace{\mathbb{E} \left[ \sum_{t=1}^T f_t(\mathbf{x}_t) - \sum_{t=1}^T f_t(\mathbf{y}_t) \right]}_{\text{EXPLORATION}} + \underbrace{\mathbb{E} \left[ \sum_{t=1}^T f_t(\mathbf{y}_t) - \sum_{t=1}^T \tilde{f}_t(\mathbf{y}_t) \right]}_{\text{REGULARIZATION I}} + \underbrace{\mathbb{E} \left[ \sum_{t=1}^T \tilde{f}_t(\mathbf{y}_t) - \sum_{t=1}^T \hat{f}_t(\mathbf{y}_t) \right]}_{\text{SMOOTH I}} \\ &\quad + \underbrace{\mathbb{E} \left[ \sum_{t=1}^T \hat{f}_t(\mathbf{y}_t) - \sum_{t=1}^T \hat{f}_t(\tilde{\mathbf{x}}) \right]}_{\text{REG TERM}} + \underbrace{\mathbb{E} \left[ \sum_{t=1}^T \hat{f}_t(\tilde{\mathbf{x}}) - \sum_{t=1}^T \tilde{f}_t(\tilde{\mathbf{x}}) \right]}_{\text{SMOOTH II}} + \underbrace{\mathbb{E} \left[ \sum_{t=1}^T \tilde{f}_t(\tilde{\mathbf{x}}) - \sum_{t=1}^T f_t(\tilde{\mathbf{x}}) \right]}_{\text{REGULARIZATION II}} \end{aligned}$$

$$+ \underbrace{\mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\tilde{\mathbf{x}}) - \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}) \right]}_{\text{COMPARATOR BIAS}}, \quad (33)$$

For terms REGULARIZATION I and REGULARIZATION II, we bound them in the same way as shown in Eq. (18): for any  $\mathbf{x} \in \mathcal{X}$ ,

$$\mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}) - \sum_{t=1}^T \tilde{\mathbf{f}}_t(\mathbf{x}) \right] \leq \sum_{t=1}^T \frac{\lambda_t}{2}. \quad (34)$$

For terms SMOOTH I and SMOOTH II, instead of using the smoothness property in Appendix A, we use the Lipschitzness of  $\tilde{\mathbf{f}}_t$  and bound the two terms as follows:

$$\begin{aligned} & \mathbb{E}_{\mathbf{b}} \left[ \sum_{t=1}^T \tilde{\mathbf{f}}_t(\mathbf{x} + \mathbf{H}_t^{-\frac{1}{2}} \mathbf{b}) - \sum_{t=1}^T \tilde{\mathbf{f}}_t(\mathbf{x}) \right] \\ & \leq \sum_{t=1}^T (L + \lambda_t) \left\| \mathbf{H}_t^{-\frac{1}{2}} \mathbf{b} \right\|_2 \leq \sum_{t=1}^T \frac{L + 1}{\sqrt{\eta_t (\sigma_{1:t-1} + \lambda_{0:t-1})}} \leq \sum_{t=1}^T \frac{d^{\frac{2}{3}} (L + 1)^{\frac{2}{3}}}{(\sigma_{1:t-1} + \lambda_{0:t-1})^{\frac{1}{3}}}, \end{aligned} \quad (35)$$

where the last inequality is by the definition of  $\eta_t \geq d^{-\frac{4}{3}} (L + 1)^{\frac{2}{3}} (\sigma_{1:t-1} + \lambda_{0:t-1})^{-\frac{1}{3}}$ .

For term EXPLORATION, we again use the Lipschitzness of  $\mathbf{f}_t$  and have

$$\begin{aligned} & \mathbb{E} \left[ \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}_t) - \sum_{t=1}^T \mathbf{f}_t(\mathbf{y}_t) \right] \\ & \leq \mathbb{E} \left[ \sum_{t=1}^T L \|\mathbf{x}_t - \mathbf{y}_t\|_2 \right] \leq \sum_{t=1}^T L \left\| \mathbf{H}_t^{-\frac{1}{2}} \mathbf{u}_t \right\|_2 \leq \frac{d^{\frac{2}{3}} (L + 1)^{\frac{2}{3}}}{(\sigma_{1:t-1} + \lambda_{0:t-1})^{\frac{1}{3}}}. \end{aligned} \quad (36)$$

For term COMPARATOR BIAS, as shown in Eq. (21), we have

$$\text{COMPARATOR BIAS} \leq 2. \quad (37)$$

Next, we show the following lemma bounding REG TERM.

**Lemma 20** *When loss functions  $\{f_t\}_{t=1}^T$  are all  $L$ -Lipschitz, if  $T \geq \rho'$  (a constant defined in Algorithm 2), Algorithm 2 guarantees that*

$$\text{REG TERM} \leq \tilde{\mathcal{O}} \left( d^{\frac{4}{3}} \nu T^{\frac{1}{3}} + \sum_{t=1}^T \frac{d^{\frac{2}{3}} (L + 1)^{\frac{2}{3}}}{(\sigma_{1:t-1} + \lambda_{0:t-1})^{\frac{1}{3}}} \right). \quad (38)$$

**Proof** Similar to the analysis in Lemma 18, we first verify the conditions in Lemma 17 are satisfied. It is direct to see that  $\{\eta_t\}_{t=1}^T$  is non-increasing and

$$\frac{1}{\eta_t} - \frac{1}{\eta_{t+1}} \leq d^{\frac{4}{3}} (L + 1)^{-\frac{2}{3}} \left( \left( \frac{1}{\sigma_{1:t} + \lambda_{0:t}} + \frac{1}{T} \right)^{-\frac{1}{3}} - \left( \frac{1}{\sigma_{1:t-1} + \lambda_{0:t-1}} + \frac{1}{T} \right)^{-\frac{1}{3}} \right)$$

$$\begin{aligned} &\leq d^{\frac{4}{3}}(L+1)^{-\frac{2}{3}} \left( (\sigma_{1:t} + \lambda_{0:t})^{\frac{1}{3}} - (\sigma_{1:t-1} + \lambda_{0:t-1})^{\frac{1}{3}} \right) \\ &\leq d^{\frac{4}{3}}(L+1)^{-\frac{2}{3}} (\sigma_t + \lambda_t)^{\frac{1}{3}}. \end{aligned}$$

Therefore,  $\eta_{t+1}^{-1} - \eta_t^{-1} \leq C(\sigma_t + \lambda_t)^p$  with  $C = d^{\frac{4}{3}}(L+1)^{-\frac{2}{3}}$  and  $p = \frac{1}{3}$ . Also, because of [Lemma 31](#), choosing  $\gamma = 4L$  ensures that  $\sigma_t \leq \gamma$  for all  $t \in [T]$ . Moreover, because of the choice of  $\lambda_0$  and  $T \geq \lambda_0$ , we have

$$\begin{aligned} \eta_1 &= d^{-\frac{4}{3}}(L+1)^{\frac{2}{3}} \left( \frac{1}{\lambda_0} + \frac{1}{T} \right)^{\frac{1}{3}} \\ &\leq d^{-\frac{4}{3}}(L+1)^{\frac{2}{3}} \lambda_0^{-\frac{1}{3}} \cdot 2^{\frac{1}{3}} \\ &\leq d^{-\frac{4}{3}}(L+1)^{\frac{2}{3}} \cdot \frac{d^{\frac{1}{3}}}{32} \cdot \frac{1}{16\sqrt{\nu}d^{\frac{1}{3}}(4L+1)^{\frac{1}{3}} + (L+1)^{\frac{2}{3}}} \\ &= \frac{1}{32d} \cdot \frac{(L+1)^{\frac{2}{3}}}{16\sqrt{\nu}d^{\frac{1}{3}}(4L+1)^{\frac{1}{3}} + (L+1)^{\frac{2}{3}}} \\ &= \frac{1}{32(d + 16\sqrt{\nu}C(4L+1)^p)}. \end{aligned}$$

Therefore, according to [Lemma 17](#),  $\|\mathbf{y}_t - \mathbf{y}_{t+1}\|_{\mathbf{H}_t} \leq \frac{1}{2}$ . In addition, according to [Eq. \(16\)](#), we have  $\|\mathbf{g}_t\|_{\mathbf{H}_t}^* \leq 2d$  for all  $t \in [T]$ . Therefore, [Eq. \(29\)](#) holds. Noticing that  $\eta_t = d^{-\frac{4}{3}}(L+1)^{\frac{2}{3}}(\sigma_{1:t-1} + \lambda_{0:t-1})^{-\frac{1}{3}}$ , and using [Eq. \(25\)](#) and [Eq. \(29\)](#), we have

$$\begin{aligned} \text{REG TERM} &= \mathbb{E} \left[ \sum_{t=1}^T \widehat{\mathbf{f}}_t(\mathbf{y}_t) - \sum_{t=1}^T \widehat{\mathbf{f}}_t(\tilde{\mathbf{x}}) \right] \\ &\leq \mathbb{E} \left[ \sum_{t=1}^T \ell_t(\mathbf{y}_t) - \sum_{t=1}^T \ell_t(\tilde{\mathbf{x}}) \right] \tag{Eq. (25)} \\ &\leq \frac{\lambda_0}{2} \|\tilde{\mathbf{x}}\|_2^2 + \frac{\Psi(\tilde{\mathbf{x}}) - \Psi(\mathbf{y}_1)}{\eta_{T+1}} + \sum_{t=1}^T 2\eta_t \|\mathbf{g}_t\|_{\mathbf{H}_t}^{*2} \tag{Eq. (29)} \\ &\leq \mathcal{O}(1) + \tilde{\mathcal{O}} \left( d^{\frac{4}{3}} \nu T^{\frac{1}{3}} \right) + \sum_{t=1}^T \frac{8d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\sigma_{1:t-1} + \lambda_{0:t-1})^{\frac{1}{3}}}, \end{aligned}$$

where we use the fact that  $\eta_{T+1} \geq d^{-\frac{4}{3}}(L+1)^{\frac{2}{3}}T^{-\frac{1}{3}} \geq d^{-\frac{4}{3}}T^{-\frac{1}{3}}$ . This finishes the proof.  $\blacksquare$

Finally, we combine the above terms and show the following theorem, which holds for an arbitrary sequence of  $\{\lambda_t\}_{t=1}^T$  with  $\lambda_t \in (0, 1)$  for all  $t \in [T]$ , not necessarily satisfying [Eq. \(9\)](#).

**Theorem 21** *With any regularization coefficients  $\{\lambda_t\}_{t=1}^T \in (0, 1)$ , [Algorithm 2](#) guarantees:*

$$\text{REG} \leq \tilde{\mathcal{O}} \left( \sum_{t=1}^T \frac{d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\sigma_{1:t-1} + \lambda_{0:t-1})^{\frac{1}{3}}} + \lambda_{1:T} \right), \tag{39}$$

if loss functions  $\{f_t\}_{t=1}^T$  are all  $L$ -Lipschitz and  $T \geq \rho'$  (a constant defined in [Algorithm 2](#)).

**Proof** Combining [Eq. \(34\)](#), [Eq. \(35\)](#), [Eq. \(36\)](#), [Eq. \(37\)](#) and [Eq. \(38\)](#), we have

$$\mathbb{E} \left[ \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(x) \right] \leq \tilde{\mathcal{O}} \left( \sum_{t=1}^T \frac{d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\sigma_{1:t-1} + \lambda_{0:t-1})^{\frac{1}{3}}} + d^{\frac{4}{3}}\nu T^{\frac{1}{3}} + \lambda_{1:T} \right).$$

■

Next we show that if we choose the adaptive regularization coefficients as shown in [Eq. \(9\)](#), the obtained regret bound is no worse than the one with an optimal tuning of  $\{\lambda_t\}_{t=1}^T$ .

**Lemma 22** *Consider the following objective*

$$\mathcal{B}'(\{\lambda_s\}_{s=1}^t) \triangleq \lambda_{1:t} + \sum_{\tau=1}^t \frac{d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\sigma_{1:\tau} + \lambda_{0:\tau})^{\frac{1}{3}}}, \quad (40)$$

with  $\lambda_0$  defined in [Algorithm 2](#). Then the sequence  $\{\lambda_t\}_{t=1}^T$  attained by solving [Eq. \(9\)](#) satisfies that for all  $t \in [T]$ ,  $\lambda_t \in (0, 1)$  and

$$\mathcal{B}'(\{\lambda_s\}_{s=1}^t) \leq 2 \min_{\{\lambda_s^*\}_{s=1}^t \geq 0} \mathcal{B}'(\{\lambda_s^*\}_{s=1}^t). \quad (41)$$

**Proof** First, we show that there exists a coefficient  $\lambda_t \in (0, 1)$  for all  $t \in [T]$  that satisfies the fixed-point problem [Eq. \(40\)](#). Indeed, we have the following two observations:

- on one hand, when setting  $\lambda_t = 0$ , the LHS of [Eq. \(41\)](#) equals to 0, while the RHS of [Eq. \(41\)](#) is strictly larger than 0;
- on the other hand, when setting  $\lambda_t = 1$ , the LHS of [Eq. \(41\)](#) is equal to 1 but the RHS of [Eq. \(41\)](#) is strictly less than 1 due to the choice of  $\lambda_0 \geq d^2(L+1)^2$ .

Combining both facts shows that there exists a coefficient  $\lambda_t \in (0, 1)$  that satisfies [Eq. \(41\)](#).

We prove this by induction similar to Lemma 3.1 in [\(Bartlett et al., 2007\)](#). Again, we set  $\lambda_0^* = \lambda_0$ .

Consider the case of  $t = 1$ . If  $\hat{\lambda}_1 \leq \lambda_1^*$ , we have  $\mathcal{B}'(\hat{\lambda}_1) = \hat{\lambda}_1 + \frac{d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\sigma_1 + \hat{\lambda}_{0:1})^{\frac{1}{3}}} = 2\hat{\lambda}_1 \leq 2\lambda_1^* \leq 2\mathcal{B}'(\lambda_1^*)$ .

Otherwise, we have  $\mathcal{B}'(\hat{\lambda}_1) = \frac{2d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\sigma_1 + \hat{\lambda}_{0:1})^{\frac{1}{3}}} \leq \frac{2d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\lambda_{0:1}^* + \sigma_1)^{\frac{1}{3}}} \leq 2\mathcal{B}'(\lambda_1^*)$ .

Suppose we have  $\mathcal{B}'(\{\hat{\lambda}_s\}_{s=1}^{t-1}) \leq 2 \min_{\{\lambda_s^*\}_{s=1}^{t-1} \geq 0} \mathcal{B}'(\{\lambda_s^*\}_{s=1}^{t-1})$ . With a slight abuse of notation, we set  $\{\lambda_s^*\}_{s=1}^t = \operatorname{argmin}_{\{\lambda_s^*\}_{s=1}^t \geq 0} \mathcal{B}'(\{\lambda_s^*\}_{s=1}^t)$ . Similarly, if  $\hat{\lambda}_{1:t} \leq \lambda_{1:t}^*$ , we have

$$\mathcal{B}'(\{\hat{\lambda}_s\}_{s=1}^t) = \hat{\lambda}_{1:t} + \sum_{s=1}^t \frac{d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\sigma_{1:s} + \hat{\lambda}_{0:s})^{\frac{1}{3}}} = \hat{\lambda}_{1:t} + \sum_{s=1}^t \hat{\lambda}_s \leq 2\lambda_{1:t}^* \leq 2\mathcal{B}'(\{\lambda_s^*\}_{s=1}^t).$$

Otherwise, we have

$$\hat{\lambda}_t + \frac{d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\sigma_{1:t} + \hat{\lambda}_{0:t})^{\frac{1}{3}}} = \frac{2d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\sigma_{1:t} + \hat{\lambda}_{0:t})^{\frac{1}{3}}} \leq \frac{2d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\sigma_{1:t} + \lambda_{0:t}^*)^{\frac{1}{3}}} \leq 2 \left( \lambda_t^* + \frac{d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(\sigma_{1:t} + \lambda_{1:t}^*)^{\frac{1}{3}}} \right).$$

Using the induction hypothesis, we have  $\mathcal{B}'(\{\hat{\lambda}_s\}_{s=1}^t) \leq 2\mathcal{B}'(\{\lambda_s^*\}_{s=1}^t)$ . ■

Therefore, combining [Theorem 21](#) and [Lemma 22](#) gives the proof of [Theorem 12](#).

## B.2. Proofs for Implications of Theorem 12

In this subsection, we prove the corollaries presented in Section 4.

**Proof** [of Corollary 13] Since Theorem 12 holds for any sequence of  $\{\lambda_t^*\}_{t=1}^T$ , in particular, we choose  $\lambda_1^* = \sqrt{(L+1)dT^{3/4}}$  and  $\lambda_t^* = 0$  for all  $t \geq 2$ , then we obtain that

$$\text{REG} \stackrel{(10)}{\leq} \tilde{\mathcal{O}}\left(\sqrt{d(L+1)T^{3/4}}\right),$$

which completes the proof.  $\blacksquare$

**Proof** [of Corollary 14] Choose  $\lambda_t^* = 0$  for all  $t \geq 1$  and by Theorem 12, we obtain that

$$\text{REG} \stackrel{(10)}{\leq} \tilde{\mathcal{O}}\left(\sum_{t=1}^T \frac{d^{2/3}(L+1)^{2/3}}{\sigma^{1/3}t^{1/3}}\right) = \tilde{\mathcal{O}}((L+1)^{2/3}d^{2/3}T^{2/3}\sigma^{-1/3}),$$

which completes the proof.  $\blacksquare$

**Proof** [of Corollary 15] In the first environment where there are  $M$  rounds such that the loss function is 0-strongly convex. In order to make the right hand side of Eq. (10) the largest, we have  $\sigma_\tau = 0$  when  $\tau \in [M]$  and  $\sigma_\tau = \sigma$  when  $\tau > M$ . Set  $\lambda_t^* = 0$  for all  $t \geq 2$ , then Theorem 12 implies that (omitting the  $\tilde{\mathcal{O}}(d^{4/3}\nu T^{1/3})$  low-order term)

$$\begin{aligned} \text{REG} &\leq \tilde{\mathcal{O}}\left(\lambda_1^* + \sum_{t=1}^T \frac{d^{2/3}(L+1)^{2/3}}{(\sigma_{1:t} + \lambda_1^*)^{1/3}}\right) \\ &\leq \tilde{\mathcal{O}}\left(\lambda_1^* + \lambda_1^{*-1/3} M d^{2/3} (L+1)^{2/3} + d^{2/3} (L+1)^{2/3} \min\left\{\sigma^{-1/3} (T-M)^{2/3}, \lambda_1^{*-1/3} (T-M)\right\}\right) \\ &\leq \tilde{\mathcal{O}}\left(\sqrt{d(L+1)} M^{3/4} + d^{2/3} (L+1)^{2/3} \min\left\{\sigma^{-1/3} (T-M)^{2/3}, \frac{T-M}{d^{1/6} (L+1)^{1/6} M^{1/4}}\right\}\right), \end{aligned}$$

where the last inequality is by choosing  $\lambda_1^* = \sqrt{d(L+1)} M^{3/4}$ . This proves the first result.

Consider the second type of environment where the first  $T - M$  rounds are  $\sigma$ -strongly convex functions and the remaining rounds are 0-strongly convex functions. Still set  $\lambda_t^* = 0$  for all  $t \geq 2$  and we have

$$\begin{aligned} \text{REG} &\leq \tilde{\mathcal{O}}\left(\lambda_1^* + \sum_{t=1}^T \frac{d^{2/3}(L+1)^{2/3}}{(\sigma_{1:t} + \lambda_1^*)^{1/3}}\right) \\ &\leq \tilde{\mathcal{O}}\left(\lambda_1^* + d^{2/3} (L+1)^{2/3} \min\left\{\lambda_1^{*-1/3} (T-M), \sigma^{-1/3} (T-M)^{2/3}\right\} + \frac{d^{2/3} M (L+1)^{2/3}}{(\lambda_1^* + \sigma(T-M))^{1/3}}\right). \end{aligned}$$

When  $T - M = \Theta(T)$ , we have

$$\text{REG} \leq \tilde{\mathcal{O}}\left(\lambda_1^* + d^{2/3} (L+1)^{2/3} \min\left\{\lambda_1^{*-1/3} T, \sigma^{-1/3} T^{2/3}\right\} + \frac{d^{2/3} M (L+1)^{2/3}}{(\lambda_1^* + \sigma T)^{1/3}}\right)$$



$$\begin{aligned} &\leq \tilde{\mathcal{O}} \left( \lambda_1^* + \frac{d^{\frac{2}{3}}T(L+1)^{\frac{2}{3}}}{(\sigma T)^{\frac{1}{3}}} \right) \\ &\leq \tilde{\mathcal{O}} \left( \frac{d^{\frac{2}{3}}T(L+1)^{\frac{2}{3}}}{\sigma^{\frac{1}{3}}(T-M)^{\frac{1}{3}}} \right), \end{aligned}$$

where the last inequality is by choosing  $\lambda_1^* = 0$ . When  $T - M = o(T)$ , we have  $M = \Theta(T)$ . Furthermore, when  $\lambda_1 \leq \sigma(T - M)$ , we have  $\lambda_1^{*-\frac{1}{3}}(T - M) \geq \sigma^{-\frac{1}{3}}(T - M)^{\frac{2}{3}}$  and therefore,

$$\text{REG} \leq \tilde{\mathcal{O}} \left( \lambda_1^* + d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}\sigma^{-\frac{1}{3}}T^{\frac{2}{3}} + \frac{d^{\frac{2}{3}}T(L+1)^{\frac{2}{3}}}{(\lambda_1^* + \sigma(T-M))^{\frac{1}{3}}} \right) \leq \tilde{\mathcal{O}} \left( \frac{d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}T}{\sigma^{\frac{1}{3}}(T-M)^{\frac{1}{3}}} \right),$$

where the last inequality is by choosing  $\lambda_1^* = 0$ . When  $\lambda_1^* \geq \sigma(T - M)$ , we have  $\lambda_1^{*-\frac{1}{3}}(T - M) \leq \sigma^{-\frac{1}{3}}(T - M)^{\frac{2}{3}}$  and therefore,

$$\begin{aligned} \text{REG} &\leq \tilde{\mathcal{O}} \left( \lambda_1^* + d^{\frac{2}{3}}\lambda_1^{*-\frac{1}{3}}(L+1)^{\frac{2}{3}}T + \frac{d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}T}{(\lambda_1^* + \sigma(T-M))^{\frac{1}{3}}} \right) \leq \tilde{\mathcal{O}} \left( \lambda_1^* + d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}\lambda_1^{*-\frac{1}{3}}T \right) \\ &\leq \tilde{\mathcal{O}} \left( \sigma(T-M) + \sqrt{d(L+1)}T^{\frac{3}{4}} \right), \end{aligned}$$

where the last inequality is by choosing  $\lambda_1^* = \max \left\{ \sigma(T - M), \sqrt{d(L+1)}T^{\frac{3}{4}} \right\}$ . Combining the two cases, we have

$$\text{REG} \leq \tilde{\mathcal{O}} \left( \min \left\{ \frac{d^{\frac{2}{3}}T(L+1)^{\frac{2}{3}}}{\sigma^{\frac{1}{3}}(T-M)^{\frac{1}{3}}}, \sigma(T-M) + \sqrt{d(L+1)}T^{\frac{3}{4}} \right\} \right),$$

leading to the second conclusion.  $\blacksquare$

**Proof** [of [Corollary 16](#)] Since [Theorem 12](#) holds for any sequence of  $\{\lambda_t^*\}_{t=1}^T$ , in particular, we choose  $\lambda_t^* = 0$  for all  $t \geq 2$  and set  $\lambda_1^* = (L+1)^{\mu_0}d^{\mu_1}T^{\mu_2}$ , we obtain that (again omitting the low-order term)

$$\begin{aligned} \text{REG} &\leq \tilde{\mathcal{O}} \left( \lambda_1^* + \sum_{t=1}^T \frac{d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}}{(t^{1-\alpha} + \lambda_1^*)^{\frac{1}{3}}} \right) \\ &\leq \tilde{\mathcal{O}} \left( (L+1)^{\mu_0}d^{\mu_1}T^{\mu_2} + d^{\frac{2}{3}}(L+1)^{\frac{2}{3}} \min \left\{ (L+1)^{-\frac{\mu_0}{3}}d^{-\frac{\mu_1}{3}}T^{1-\frac{\mu_2}{3}}, T^{\frac{2+\alpha}{3}} \right\} \right). \end{aligned}$$

First, the above bound can be upper bounded by

$$\text{REG} \leq \tilde{\mathcal{O}} \left( (L+1)^{\mu_0}d^{\mu_1}T^{\mu_2} + d^{\frac{2-\mu_1}{3}}(L+1)^{\frac{2-\mu_0}{3}}T^{1-\frac{\mu_2}{3}} \right) = \tilde{\mathcal{O}} \left( \sqrt{d(L+1)}T^{\frac{3}{4}} \right),$$

where the last equality is true by choosing  $\mu_0 = \frac{1}{2}$ ,  $\mu_1 = \frac{1}{2}$  and  $\mu_2 = \frac{3}{4}$ . Second, when  $\alpha \in [0, \frac{1}{4} - \frac{1}{2} \log_T(L+1) - \frac{1}{2} \log_T d]$ , we choose  $\mu_0 = \mu_1 = \mu_2 = 0$  and have

$$\text{REG} \leq \tilde{\mathcal{O}} \left( d^{\frac{2}{3}}(L+1)^{\frac{2}{3}}T^{\frac{2+\alpha}{3}} \right).$$

Again, we emphasize that the setting of  $\lambda_1^*$  is required in the analysis only and will not affect the algorithmic procedures. Combining both situations finishes the proof.  $\blacksquare$

### Appendix C. Self-concordant Barrier Properties

One of the important technical tools used in this paper is the *self-concordant barrier*, which is widely used in the interior-point method and becomes a central concept in modern convex optimization (Nesterov and Nemirovskii, 1994). The method is introduced to online learning community in the seminal paper of Abernethy et al. (2008) and successfully resolve several important open problems. Below, we list several basic definitions and some important properties, most of which can be found in (Nemirovski and Todd, 2008, Section 2).

**Definition 23 (Self-Concordant Functions)** *Let  $\mathcal{X} \subseteq \mathbb{R}^d$  be a closed convex domain with a nonempty interior  $\text{int}(\mathcal{X})$ . A function  $\mathcal{R} : \text{int}(\mathcal{X}) \mapsto \mathbb{R}$  is called self-concordant on  $\mathcal{X}$  if*

- (i)  $\mathcal{R}$  is a three times continuously differentiable convex function, and approaches infinity along any sequence of points approaching  $\partial\mathcal{X}$ ; and
- (ii)  $\mathcal{R}$  satisfies the differential inequality: for every  $\mathbf{h} \in \mathbb{R}^d$  and  $x \in \text{int}(\mathcal{X})$ ,

$$|\nabla^3\mathcal{R}(x)[\mathbf{h}, \mathbf{h}, \mathbf{h}]| \leq 2(\nabla^2\mathcal{R}(x)[\mathbf{h}, \mathbf{h}])^{\frac{3}{2}},$$

where the third-order differential is defined as

$$\nabla^3\mathcal{R}(x)[\mathbf{h}, \mathbf{h}, \mathbf{h}] \triangleq \frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} \mathcal{R}(x + t_1\mathbf{h} + t_2\mathbf{h} + t_3\mathbf{h}) \Big|_{t_1=t_2=t_3=0}.$$

Given a real  $\nu \geq 1$ ,  $\mathcal{R}$  is called a  $\nu$ -self-concordant barrier ( $\nu$ -SCB) for  $\mathcal{X}$  if  $\mathcal{R}$  is self-concordant on  $\mathcal{X}$  and, in addition, for every  $\mathbf{h} \in \mathbb{R}^d$  and  $x \in \text{int}(\mathcal{X})$ ,

$$|\nabla\mathcal{R}(x)[\mathbf{h}]| \leq \nu^{\frac{1}{2}}(\nabla^2\mathcal{R}(x)[\mathbf{h}, \mathbf{h}])^{\frac{1}{2}}.$$

Given a self-concordant function  $\mathcal{R}$  on  $\mathcal{X}$ , for any  $\mathbf{h} \in \mathbb{R}^d$  the induced local norm is defined as

$$\|\mathbf{h}\|_x \triangleq \|\mathbf{h}\|_{\nabla^2\mathcal{R}(x)} = \sqrt{\mathbf{h}^\top \nabla^2\mathcal{R}(x)\mathbf{h}}, \quad \text{and} \quad \|\mathbf{h}\|_x^* \triangleq \|\mathbf{h}\|_{\nabla^2\mathcal{R}(x)}^* = \sqrt{\mathbf{h}^\top (\nabla^2\mathcal{R}(x))^{-1}\mathbf{h}}. \quad (42)$$

Below, we present several key technical lemmas regarding to the self-concordant functions.

**Lemma 24 (Proposition 2.3.2 of Nesterov and Nemirovskii (1994))** *Let  $\mathcal{R} : \text{int}(\mathcal{X}) \mapsto \mathbb{R}$  be a  $\nu$ -self-concordant barrier over the closed convex set  $\mathcal{X} \in \mathbb{R}^d$ , then for any  $x, y \in \text{int}(\mathcal{X})$ , we have  $\mathcal{R}(y) - \mathcal{R}(x) \leq \nu \log \frac{1}{1-\pi_x(y)}$ , where  $\pi_x(y) \triangleq \inf\{t \geq 0 \mid x + t^{-1}(y-x) \in \mathcal{X}\}$  is the Minkowski function of  $\mathcal{X}$  whose pole is on  $x$ , which is always in  $[0, 1]$ .*

**Lemma 25 (Theorem 2.1.1 of Nesterov and Nemirovskii (1994))** *Let  $\psi$  be a self-concordant function on the closed convex body  $\mathcal{X} \subseteq \mathbb{R}^d$  with non-empty interior, then*

$$\|h\|_{\nabla^2\psi(x')} \geq \|h\|_{\nabla^2\psi(x)} (1 - \|x - x'\|_{\nabla^2\psi(x)}) \quad (43)$$

holds for any  $h \in \mathbb{R}^d$  and any  $x \in \text{int}(\mathcal{X})$  with  $x' \in \mathcal{E}_1(x)$  where  $\mathcal{E}_1(x) \triangleq \{y \in \mathbb{R}^d \mid \|y - x\|_x \leq 1\}$  denotes the Dikin ellipsoid of  $\mathcal{R}$  and we always have  $\mathcal{E}_1(x) \subseteq \mathcal{X}$ .

**Lemma 26 (Theorem 2.5.1 of Nesterov and Nemirovskii (1994))** For each closed convex body  $\mathcal{X} \subseteq \mathbb{R}^d$ , there exists an  $\mathcal{O}(d)$ -self-concordant barrier on  $\mathcal{X}$ .

In the following, we present the definition and properties of a *normal barrier*.

**Definition 27 (Normal Barriers)** Let  $K \subseteq \mathbb{R}^d$  be a closed and proper convex cone and let  $\bar{\nu} \geq 1$ . A function  $\psi : \text{int}(K) \rightarrow \mathbb{R}$  is called a  $\bar{\nu}$ -logarithmically homogeneous self-concordant barrier (or simply a  $\bar{\nu}$ -normal barrier) on  $K$  if it is self-concordant on  $\text{int}(K)$  and is logarithmically homogeneous with parameter  $\bar{\nu}$ , namely

$$\psi(tw) = \psi(w) - \bar{\nu} \ln t, \quad \forall w \in \text{int}(K), t > 0.$$

**Lemma 28 (Proposition 5.1.4 of Nesterov and Nemirovskii (1994))** Suppose  $\psi$  is a  $\nu$ -self-concordant barrier on  $\mathcal{X} \subseteq \mathbb{R}^d$ . Then the function

$$\Psi(w, b) \triangleq 400 \left( \psi\left(\frac{w}{b}\right) - 2\nu \ln b \right)$$

is a  $\bar{\nu}$ -normal barrier on  $\text{con}(\mathcal{X}) \subseteq \mathbb{R}^{d+1}$  with  $\bar{\nu} = 800\nu$ , where  $\text{con}(\mathcal{X}) = \{\mathbf{0}\} \cup \{(w, b) \mid \frac{w}{b} \in \mathcal{X}, w \in \mathbb{R}^d, b > 0\}$  is the conic hull of  $\mathcal{X}$  lifted to  $\mathbb{R}^{d+1}$  (by appending a dummy variable 1 to the last coordinate).

**Lemma 29 (Proposition 2.3.4 of Nesterov and Nemirovskii (1994))** Suppose  $\psi$  is a  $\nu$ -normal barrier on  $\mathcal{X} \subseteq \mathbb{R}^d$ . Then for any  $x, y \in \text{int}(\mathcal{X})$ , we have

- (1)  $\|x\|_{\nabla^2 \psi(x)}^2 = x^\top \nabla^2 \psi(x) x = \nu$ ;
- (2)  $\nabla^2 \psi(x) x = -\nabla \psi(x)$ ;
- (3)  $\psi(y) \geq \psi(x) - \nu \ln \frac{-\langle \nabla \psi(x), y \rangle}{\nu}$ .
- (4)  $\|\nabla \psi(x)\|_{\nabla^{-2} \psi(x)}^2 = \nu$ .

## Appendix D. Additional Lemmas

### D.1. FTRL Lemma

For completeness, we present the following general result for FTRL-type algorithms as follows.

**Lemma 30** Let  $\mathcal{X} \subseteq \mathbb{R}^d$  be a closed and convex feasible set, and denote by  $R_t : \mathcal{X} \mapsto \mathbb{R}$  the convex regularizer and by  $f_t : \mathcal{X} \mapsto \mathbb{R}$  the convex online functions. Denote by  $F_t(x) = R_t(x) + \sum_{s=1}^{t-1} f_s(x)$  and the FTRL update rule is specified as  $x_t \in \text{argmin}_{x \in \mathcal{X}} F_t(x)$ . Then, for any  $u \in \mathcal{X}$  we have

$$\begin{aligned} \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(u) &\leq R_{T+1}(u) - R_1(x_1) + \sum_{t=1}^T \nabla f_t(x_t)^\top (x_t - x_{t+1}) \\ &\quad - \sum_{t=1}^T D_{F_t + f_t}(x_{t+1}, x_t) + \sum_{t=1}^T \left( R_t(x_{t+1}) - R_{t+1}(x_{t+1}) \right), \end{aligned} \tag{44}$$

where  $D_{F_t + f_t}(\cdot, \cdot)$  denotes the Bregman divergence induced by the function  $F_t + f_t$ .

**Proof** It is easy to verify that the following equation holds for any comparator  $u \in \mathcal{X}$ ,

$$\begin{aligned} \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(u) &= R_{T+1}(u) - R_1(x_1) + F_{T+1}(x_{T+1}) - F_{T+1}(u) \\ &\quad + \sum_{t=1}^T \left( F_t(x_t) - F_{t+1}(x_{t+1}) + f_t(x_t) \right). \end{aligned}$$

Moreover, we have

$$\begin{aligned} &F_t(x_t) - F_{t+1}(x_{t+1}) + f_t(x_t) \\ &= F_t(x_t) + f_t(x_t) - (F_t(x_{t+1}) + f_t(x_{t+1})) + R_t(x_{t+1}) - R_{t+1}(x_{t+1}) \\ &= \langle \nabla F_t(x_t) + \nabla f_t(x_t), x_t - x_{t-1} \rangle - D_{F_t+f_t}(x_{t+1}, x_t) + R_t(x_{t+1}) - R_{t+1}(x_{t+1}) \\ &\leq \langle \nabla f_t(x_t), x_t - x_{t-1} \rangle - D_{F_t+f_t}(x_{t+1}, x_t) + R_t(x_{t+1}) - R_{t+1}(x_{t+1}) \end{aligned}$$

where the last inequality holds by the optimality condition of  $x_t \in \operatorname{argmin}_{x \in \mathcal{X}} F_t(x)$ . Hence, combining the above equations finishes the proof.  $\blacksquare$

## D.2. Relations among strong convexity, smoothness and Lipschitzness

In this section, we discuss the relations among strong convexity, smoothness and Lipschitzness. First, we point out a minor technical flaw that appeared in two previous works on BCO (Saha and Tewari, 2011; Hazan and Levy, 2014). In both works, the authors use the statement that a convex function  $f$  that is  $\beta$ -smooth and has bounded value in  $[-1, 1]$  has Lipschitz constant no more than  $2\beta + 1$  when  $\max_{x, x' \in \mathcal{X}} \|x - x'\|_2 \in [2, 4]$ . However, this is not correct as we give the following counter example.

**Example 1** Consider the following function in 2-dimensional space:  $f(x, y) = Gy$  where  $G > 1$  can be arbitrarily large and the first coordinate does not affect the function value. The feasible domain is defined as  $\mathcal{X} = \{(x, y) \mid x \in [-1, 1], y \in [-\frac{1}{G}, \frac{1}{G}]\}$  with diameter in  $[2, 4]$ . It is direct to see that function  $f$  is 0-smooth and has bounded value in  $[-1, 1]$ . However the Lipschitz constant is  $G$ , which can be arbitrarily large.

Hazan and Levy (2014) and Saha and Tewari (2011) use this property to bound the term COMPARTOR BIAS in Eq. (17). We fix that by using the property of convexity.

Next, we discuss the relationship between strong convexity and Lipschitzness. Specifically, the following lemma shows that for a convex function  $f$  that is  $L$ -Lipschitz and defined over a bounded domain with diameter  $D$ , its strong convexity parameter  $\sigma$  is upper bounded by  $\frac{4L}{D}$ .

**Lemma 31** If a convex function  $f : \mathcal{X} \mapsto \mathbb{R}$  is  $L$ -Lipschitz and  $\sigma$ -strongly convex, and has bounded domain diameter  $\max_{x, x' \in \mathcal{X}} \|x - x'\|_2 = D$ , then we have  $\sigma \leq \frac{4L}{D}$ .

In fact, we have for any  $x, y \in \mathcal{X}$

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

Choose  $y = \operatorname{argmin}_{x \in \mathcal{X}} f_t(x)$  and we have

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

Therefore,  $\sigma \leq \frac{2L}{\|x - y\|_2}$  for any  $x \in \mathcal{X}$ , which means that  $\sigma \leq \frac{4L}{D}$ . This is because we can choose  $x_1, x_2 \in \mathcal{X}$  such that  $\|x_1 - x_2\| = D = \max_{x, x' \in \mathcal{X}} \|x - x'\|_2$ . Then we have  $\|x_1 - y\|_2 + \|x_2 - y\|_2 \geq \|x_1 - x_2\|_2 = D$ , which means that either  $\|x_1 - y\| \geq \frac{D}{2}$  or  $\|x_2 - y\|_2 \geq \frac{D}{2}$ . This shows that when  $f_t$  is both  $\sigma$ -strongly convex and  $L$ -Lipschitz, we have  $\sigma \leq \frac{4L}{D}$ .