Beating the Adaptive Bandit with High Probability

Jacob Abernethy
Computer Science Division
UC Berkeley
jake@cs.berkeley.edu

Alexander Rakhlin
Department of Statistics
University of Pennsylvania
rakhlin@wharton.upenn.edu

***Abstract*— We provide a principled way of proving $\tilde{O}(\sqrt{T})$ high-probability guarantees for partial-information (bandit) problems over arbitrary convex decision sets. First, we prove a regret guarantee for the full-information problem in terms of "local" norms, both for entropy and self-concordant barrier regularization, unifying these methods. Given one of such algorithms as a black-box, we can convert a bandit problem into a full-information problem using a sampling scheme. The main result states that a high-probability $\tilde{O}(\sqrt{T})$ bound holds whenever the black-box, the sampling scheme, and the estimates of missing information satisfy a number of conditions, which are relatively easy to check. At the heart of the method is a construction of linear upper bounds on confidence intervals. As applications of the main result, we provide the first known efficient algorithm for the sphere with an $\tilde{O}(\sqrt{T})$ high-probability bound. We also derive the result for the n-simplex, improving the $O(\sqrt{nT\log(nT)})$ bound of Auer et al [3] by replacing the $\log T$ term with $\log\log T$ and closing the gap to the lower bound of $\Omega(\sqrt{nT})$. While $\tilde{O}(\sqrt{T})$ high-probability bounds should hold for general decision sets through our main result, construction of linear upper bounds depends on the particular geometry of the set; we believe that the sphere example already exhibits the necessary ingredients. The guarantees we obtain hold for adaptive adversaries (unlike the in-expectation results of [1]) and the algorithms are efficient, given that the linear upper bounds on confidence can be computed.**

I. Introduction

The problem of Online Convex Optimization, in which a player attempts to minimize his regret against a possibly adversarial sequence of convex cost functions, is now quite well-understood. The more recent research trend has been to consider various limited-information versions of this problem. In particular, the "bandit" version of Online Linear Optimization (OLO) has received much attention in the past few years.

To be precise, the problem we are interested in can be phrased as a repeated game between the player and the adversary. At each round, the player picks a decision from the allowed convex set of moves, and the adversary simultaneously picks a linear cost function from her allowed set of moves. Unlike the well-understood OLO game, in the bandit version only the cost of the decision is revealed to the player, not the cost function itself. The adversary, on the other hand, is aware of the complete history. The aim of the player is to minimize *regret*, the cumulative cost incurred over the course of the game minus the cumulative cost of the best fixed decision.

The scarcity of information revealed to the player makes the problem difficult. The first efficient algorithm with an $\tilde{O}(\sqrt{T})$ guarantee on the regret for optimization over arbitrary convex sets was recently obtained in [1]. This guarantee was shown to hold *in expectation* and the question of obtaining guarantees *in high probability* was left open. In this paper, we develop a general framework for obtaining high-probability statements for bandit problems. We aim to provide a clean picture, building upon the mechanism employed in [3], [5]. We also simplify the proof of [1] for the regret of regularization with a self-concordant barrier and put it into the context of a general class of regret bounds based on local norms.

A reader surveying the literature on bandit optimization can easily get confused trying to distinguish between the results. Thus, we first itemize some recent papers according to the following criteria: (a) *efficient* algorithm vs *inefficient* algorithm, (b) *arbitrary* convex set vs *simplex* or the *set of flows* in a graph, (c) *optimal* $\tilde{O}(\sqrt{T})$ vs *suboptimal* (e.g. $O\left(T^{2/3}\right)$) guarantee, (d) *in-expectation* vs *high-probability* guarantee, and (e) whether the result holds for an *adaptive* adversary or only an *oblivious* one. For all the results we are aware of (including the ones in this paper), a high-probability guarantee on the regret naturally covers the case of an adaptive adversary. This is not necessarily true for the in-expectation results.

With respect to these parameters,

- Auer et al [3] obtained an efficient algorithm for the simplex, with an optimal guarantee which holds in high probability.
- McMahan and Blum [13] and Flaxman et al [11] obtained efficient algorithms for an arbitrary convex set with suboptimal guarantees which hold in expectation against an adaptive adversary.
- Awerbuch and Kleinberg [4] obtained an efficient algorithm for the set of flows with a suboptimal guarantee which holds in expectation against an adaptive adversary.
- György et al [12] obtained an efficient algorithm for the set of flows with a suboptimal guarantee which holds in high probability.[1]
- Dani et al [9] obtained an inefficient algorithm for an arbitrary set, with an optimal guarantee which holds in expectation against an oblivious adversary. The algorithm can be implemented efficiently for the set of flows.
- Bartlett et al [5] extended the result of [9] to obtain an inefficient algorithm for an arbitrary set, with an optimal

[1]The authors also obtained an optimal guarantee for the set of flows in the setting where the lengths of all edges on the chosen path are revealed. This does not match the bandit problem considered in this paper.

guarantee which holds in high probability. The algorithm cannot be (in a straightforward way) implemented efficiently for the set of flows.
- Abernethy et al [1] exhibited an efficient algorithm for an arbitrary convex set, with an optimal guarantee which holds in expectation against an oblivious adversary.
- In this paper, we obtain an efficient algorithm for a sphere and simplex with an optimal guarantee which holds in high probability (and, thus, against an adaptive adversary). Analogous results can be obtained for other convex sets; however, such results would have to be considered on the per-case basis, as the specific geometry of the set plays an important role for obtaining an efficient algorithm with an optimal high-probability guarantee.

This paper is organized as follows. In Section II, we discuss full-information algorithms which will be used as black-boxes for bandit optimization. In Section II-B we prove the known regret guarantees which arise from regularization with a strongly convex function. We argue that these guarantees are not strong enough to be used for bandit optimization and, in Section II-C, we introduce a notion of "local" norms. We prove general regret guarantees with respect to these norms for regularization with a self-concordant barrier and, for the case of the n-simplex, with the entropy function. This allows us to have a unified analysis of bandit optimization with either of these two methods as a black-box. Section III discusses the method of using a randomized algorithm for converting a full-information algorithm into a bandit algorithm. We discuss the advantages of "high-probability" results over the "in-expectation" results and explain why the straightforward way of applying concentration inequalities does not work. Section IV contains the main results of the paper. We state the main result, Theorem 4.1, and then apply it to various settings in the subsequent sections. The multiarmed bandit setting (the simplex case) is considered in Section V-A, and we improve upon the result of Auer et al [3] by removing the $\log T$ factor. We provide a solution for the sphere in Section V-B. In passing, we mention how the "in-expectation" result for general convex sets of [1] immediately follows Theorem 2.3. Another sampling scheme for general bodies is suggested, although we do not go into the details. The proof of our main result, Theorem 4.1, is given in Section VI. It is based on lemmas whose proofs can be found in the technical report [2].

II. Full-Information Algorithms

In this paper, we strive to obtain the most general results possible. To this end, bandit algorithms in Section IV will take as a sub-routine an abstract full-information black-box for regret minimization. We devote the present section to describing known guarantees for some full-information algorithms, as well as to developing a new family of guarantees under "local norms". The latter are suited to the study of bandit optimization.

To make things concrete, the full-information setting is that of *online linear optimization*, which is phrased as the following game between the learner (player) and the environment (adversary). Let $\mathcal{K} \subseteq \mathbb{R}^n$ be a closed convex set.

At each time step $t = 1$ to T,
- Player chooses $\mathbf{x}_t \in \mathcal{K}$
- Adversary independently chooses $\mathbf{f}_t \in \mathbb{R}^n$
- Player suffers loss $\mathbf{f}_t^\mathsf{T}\mathbf{x}_t$ and observes $\mathbf{f}_t$

The aim of the player (algorithm) is to minimize the *regret* against any "comparator" $\mathbf{u} \in \mathcal{K}$

$$R_T(\mathbf{u}) := \sum_{t=1}^{T} \mathbf{f}_t^\mathsf{T}\mathbf{x}_t - \sum_{t=1}^{T} \mathbf{f}_t^\mathsf{T}\mathbf{u}.$$

A. Algorithms

Let $\mathcal{R}(\mathbf{x})$ be a convex function. We consider the following family (with respect to the choice of $\mathcal{R}$) of Follow the Regularized Leader algorithms:

Algorithm 1 Follow the Regularized Leader (FTRL)

Input: $\eta > 0$. On the first round, play $\mathbf{x}_1 := \arg\min_{\mathbf{x}\in\mathcal{K}} \mathcal{R}(\mathbf{x})$. On round $t+1$, play

$$\mathbf{x}_{t+1} := \arg\min_{\mathbf{x}\in\mathcal{K}} \left[\eta \sum_{s=1}^{t} \mathbf{f}_s^\mathsf{T}\mathbf{x} + \mathcal{R}(\mathbf{x}) \right]. \tag{1}$$

Without loss of generality, we assume that $\mathcal{R}$ takes its minimum at 0, since $\arg\min$ is the same modulo constant shifts of $\mathcal{R}$. We begin with a well-known fact, whose easy induction proof can be found e.g. in [16].

Proposition 2.1: The *regret* of Algorithm 1, relative to a comparator $\mathbf{u} \in \mathcal{K}$, can be upper bounded as

$$R_T(\mathbf{u}) \le \sum_{t=1}^{T} \mathbf{f}_t^\mathsf{T}(\mathbf{x}_t - \mathbf{x}_{t+1}) + \eta^{-1}\mathcal{R}(\mathbf{u}). \tag{2}$$

The FTRL algorithm is closely related to the following Mirror Descent-style algorithm [8], [16].

Algorithm 2 Mirror Descent with Projections

On the first round, play $\mathbf{x}_1 := \arg\min_{\mathbf{x}\in\mathcal{K}} \mathcal{R}(\mathbf{x})$. On round $t+1$, compute

$$\tilde{\mathbf{x}}_{t+1} := \arg\min_{\mathbf{x}\in\mathbb{R}^n} \ \eta\mathbf{f}_t^\mathsf{T}\mathbf{x} + D_{\mathcal{R}}(\mathbf{x}, \mathbf{x}_t)$$

and then play the *projected* point

$$\mathbf{x}_{t+1} := \arg\min_{\mathbf{x}\in\mathcal{K}} \ D_{\mathcal{R}}(\mathbf{x}, \tilde{\mathbf{x}}_{t+1})$$

This algorithm is given in two steps although it can be described in one. Indeed, the point $\mathbf{x}_{t+1}$ can simply be obtained as the solution to

$$\arg\min_{\mathbf{x}\in\mathcal{K}} \ \eta\mathbf{f}_t^\mathsf{T}\mathbf{x} + D_{\mathcal{R}}(\mathbf{x}, \mathbf{x}_t).$$

However, we emphasize the unprojected point $\tilde{\mathbf{x}}_{t+1}$ as it gives us an occasionally more useful regret bound:

Proposition 2.2: The *regret* of Algorithm 2, relative to a comparator $\mathbf{u} \in \mathcal{K}$, can be upper bounded as

$$R_T(\mathbf{u}) \leq \sum_{t=1}^{T} \mathbf{f}_t^{\mathsf{T}}(\mathbf{x}_t - \tilde{\mathbf{x}}_{t+1}) + \eta^{-1}\mathcal{R}(\mathbf{u}). \tag{3}$$

The analogue of Proposition 2.1 also holds:

$$R_T(\mathbf{u}) \leq \sum_{t=1}^{T} \mathbf{f}_t^{\mathsf{T}}(\mathbf{x}_t - \mathbf{x}_{t+1}) + \eta^{-1}\mathcal{R}(\mathbf{u}). \tag{4}$$

We also note that the two algorithms coincide if $\mathcal{R}$ is a barrier. We refer to [16] for the proofs of these facts.

B. Regret Bounds with Respect to "Fixed" Norms

The regret bounds stated in Propositions 2.1 and 2.2 are not ultimately satisfying. In particular, it is not immediately obvious whether the terms $\mathbf{f}_t^{\mathsf{T}}(\mathbf{x}_t - \mathbf{x}_{t+1})$ are small. Notice that the point $\mathbf{x}_{t+1}$ depends on both $\mathbf{f}_t$ as well as on the behavior of $\mathcal{R}$. It would be much more appealing if we could remove the dependence on the points $\mathbf{x}_t$ and have the regret depend solely on the Adversary's choices $\mathbf{f}_t$ and our choice of regularizer.

This can indeed be achieved if we require certain conditions on our regularizer. The typical approach is to require that $\mathcal{R}$ is *strongly convex* with respect to some norm $\|\cdot\|$, which implies that

$$\begin{aligned} \|\mathbf{x}_t - \mathbf{x}_{t+1}\|^2 &\leq \langle \nabla R(\mathbf{x}_t) - \nabla R(\mathbf{x}_{t+1}), \mathbf{x}_t - \mathbf{x}_{t+1} \rangle \\ &\leq \|\nabla R(\mathbf{x}_t) - \nabla R(\mathbf{x}_{t+1})\|^* \|\mathbf{x}_t - \mathbf{x}_{t+1}\|. \end{aligned} \tag{5}$$

where $\|\cdot\|^*$ is the norm dual to $\|\cdot\|$, and the last step follows by Hölder's Inequality. Hence, strong convexity of $\mathcal{R}$ implies

$$\|\mathbf{x}_t - \mathbf{x}_{t+1}\| \leq \|\nabla R(\mathbf{x}_t) - \nabla R(\mathbf{x}_{t+1})\|^*,$$

making possible the following result.

Proposition 2.3: When $\mathcal{R}$ is strongly convex with respect to the norm $\|\cdot\|$, then for Algorithms 1 and 2 we have the following regret bound[2]:

$$R_T(\mathbf{u}) \leq \eta \sum_{t=1}^{T} \|\mathbf{f}_t\|^{*2} + \eta^{-1}\mathcal{R}(\mathbf{u}).$$

Proof: For the case of FTRL (Algorithm 1), when $\mathcal{R}$ is a barrier function (and thus $\mathbf{x}_t$ is always attained on the interior of $\mathcal{K}$) it is a convenient fact that $\nabla R(\mathbf{x}_t) - \nabla R(\mathbf{x}_{t+1}) = \eta \mathbf{f}_t$. Applying Hölder's inequality in the statement of Proposition 2.1 leads to the desired result. If $\mathcal{R}$ is not a barrier, an application of the Kolmogorov criterion (see [7], Theorem 2.4.2) for generalized projections at step (5) yields the statement of the Proposition. For Algorithm 2, the proof is a bit more involved, but is well-known (see e.g. [6]). Again, we refer the reader to [16], [17] for details. ∎

The easiest way to see Proposition 2.3 at work is to assume that $\mathbf{f}_t \in B_p$ and $\mathcal{K} \subseteq B_q$, the unit zero-centered balls with respect to ℓ_p and ℓ_q norms, where (p, q) is a dual pair. When faced with the particular choice of (ℓ_∞, ℓ_1) pair of norms, the natural choice of regularization is the unnormalized negative *entropy function*

$$\mathcal{R}(\mathbf{x}) = \sum_i (\mathbf{x}[i] \log \mathbf{x}[i] - \mathbf{x}[i]) + (1 + \log n), \tag{6}$$

defined over the positive orthant. Here the $1 + \log n$ term ensures that $\min \mathcal{R} = 0$ over the n-simplex $\mathcal{K}$. It is easy to see that this regularization function leads to the so-called "exponential weights":

$$\mathbf{x}_{t+1}[i] = \frac{\exp\left(\eta \sum_{s=1}^{t} \mathbf{f}_t[i]\right)}{\sum_{j=1}^{n} \exp\left(-\eta \sum_{s=1}^{t} \mathbf{f}_t[j]\right)},$$

and indeed this is true for *both* Algorithm 1 and Algorithm 2. For the future, it is useful to note that the unprojected updated $\tilde{\mathbf{x}}_{t+1}$ has the very simple "unnormalized form":

$$\tilde{\mathbf{x}}_{t+1}[i] = \mathbf{x}_t[i] \exp(-\eta \mathbf{f}_t[i]). \tag{7}$$

It is well-known that the entropy function has the useful property of strong convexity with respect to the ℓ_1 norm. We can thus apply Proposition 2.3 to obtain:

$$R_T(\mathbf{u}) \leq \eta \sum_{t=1}^{T} \|\mathbf{f}_t\|_\infty^2 + \eta^{-1} \log N.$$

where the $\log N$ arises by taking $\mathcal{R}(\cdot)$ at any corner of the n-simplex. In the "expert setting" it is typical to assume that $\|\mathbf{f}_t\|_\infty \leq 1$, and so setting $\eta = \sqrt{(\log N)/T}$ appropriately we obtain

$$R_T(\mathbf{u}) \leq \eta T + \eta^{-1} \log N = 2\sqrt{T \log N}.$$

C. Regret Bounds with Respect to "Local" Norms

The analysis of Proposition 2.3 is the typical approach, and indeed it can be shown that the above bound for exponential weights is very tight, i.e. within a small constant factor from optimal. On the other hand, there are times when we cannot make the assumption that $\mathbf{f}_t$ is bounded with respect to a fixed norm. This is particularly relevant in the bandit setting, when we will be estimating the functions $\mathbf{f}_t$ yet our estimates will blow up depending on the location of the point $\mathbf{x}_t$. In such cases, to obtain tighter bounds, it will be necessary to measure the size of $\mathbf{f}_t$ with respect to a *changing norm*. While it may not be obvious at present, the ideal choice of norm is *the inverse Hessian of $\mathcal{R}$ at the point $\mathbf{x}_t$*.

From now on, define $\|\mathbf{z}\|_{\mathbf{x}} := \sqrt{\mathbf{z}^{\mathsf{T}} \nabla^2 \mathcal{R}(\mathbf{x}) \mathbf{z}}$, where $\mathbf{z} \in \mathbb{R}^n$ is arbitrary and where $\mathcal{R}$ is assumed to be the regularizer in question. The dual of this norm $\|\mathbf{z}\|_{\mathbf{x}}^*$ is identically the norm with respect to the inverse Hessian, i.e. $\|\mathbf{z}\|_{\mathbf{x}}^* := \sqrt{\mathbf{z}^{\mathsf{T}} \nabla^2 \mathcal{R}(\mathbf{x})^{-1} \mathbf{z}}$. Our goal will now be to obtain bounds of the form

$$R_T(\mathbf{u}) \leq \eta \sum_{t=1}^{T} (\|\mathbf{f}_t\|_{\mathbf{x}_t}^*)^2 + \eta^{-1}\mathcal{R}(\mathbf{u}). \tag{8}$$

Let us introduce the following shorthand: $\|\mathbf{z}\|_t := \|\mathbf{z}\|_{\mathbf{x}_t}$ for the norm defined with respect to $\mathbf{x}_t$.

[2]We also mention that a more refined proof leads to a constant of $\frac{1}{2}$ instead of 1 in front of the $\eta \sum_{t=1}^{T} \|\mathbf{f}_t\|^{*2}$ term.

For the case when $\mathcal{R}(\mathbf{x}) = \|\mathbf{x}\|_2^2$ (leading to the "online gradient descent" algorithm), this bound is easy: since $\nabla^2\mathcal{R}(\mathbf{x}) = I_n$, and $\mathcal{R}$ is strongly convex with respect to the ℓ_2 norm, we already know that

$$R_T(\mathbf{u}) \leq \eta \sum_{t=1}^T \|\mathbf{f}_t\|_2^2 + \eta^{-1}\mathcal{R}(\mathbf{u}) = \eta \sum_{t=1}^T (\|\mathbf{f}_t\|_t^*)^2 + \eta^{-1}\mathcal{R}(\mathbf{u}).$$

1) Regret guarantee for the entropy regularizer.: For the entropic regularization case mentioned above, proving a regret bound with respect to the local norm $\|\cdot\|_{\mathbf{x}}$ requires a little bit more work. First notice that $\nabla^2\mathcal{R}(\mathbf{x}) = \mathrm{diag}(\mathbf{x}[1]^{-1}, \ldots, \mathbf{x}[n]^{-1})$, and that $1 - e^{-x} \leq x$ for all real x. Next, using Eq. (7),

$$\|\mathbf{x}_t - \tilde{\mathbf{x}}_{t+1}\|_t = \sqrt{\sum_{i=1}^n (\mathbf{x}_t[i] - \tilde{\mathbf{x}}_{t+1}[i])^2 / \mathbf{x}_t[i]}$$

$$= \sqrt{\sum_{i=1}^n \mathbf{x}_t[i](1 - e^{-\eta \mathbf{f}_t[i]})^2} \leq \eta \sqrt{\sum_{i=1}^n \mathbf{x}_t[i]\mathbf{f}_t[i]^2} = \eta \|\mathbf{f}_t\|_t^*.$$

Now we make special use of Proposition 2.2. By Hölder's Inequality,

$$R_T(\mathbf{u}) \leq \eta \sum_{t=1}^T (\|\mathbf{f}_t\|_t^*)^2 + \eta^{-1}\mathcal{R}(\mathbf{u}).$$

It can be verified that Algorithms 1 and 2 produce the same $\mathbf{x}_t$ when $\mathcal{R}$ is the entropy function and $\mathcal{K}$ is the simplex. Thus, we have proved the following Theorem.

Theorem 2.1: The exponential weights algorithm (either Algorithm 1 or Algorithm 2) enjoys the following bound in terms of "local" norms:

$$R_T(\mathbf{u}) \leq \eta \sum_{t=1}^T \left(\|\mathbf{f}_t\|_t^*\right)^2 + \eta^{-1}\mathcal{R}(\mathbf{u}).$$

As a side remark, we mention that one can prove the same guarantee (with a slightly worse constant) by starting from Eq. (2) instead of Eq. (3). A lemma which can be found in the Appendix of [2], implies that

$$\begin{aligned}\|\mathbf{x}_t - \mathbf{x}_{t+1}\|_t^2 &= \sum_{i=1}^n \mathbf{x}_t[i]\left(1 - \frac{e^{-\eta\mathbf{f}_t[i]}}{\sum_{j=1}^n \mathbf{x}_t[j]e^{-\eta\mathbf{f}_t[j]}}\right)^2 \\ &= \frac{\sum_{i=1}^n \mathbf{x}_t[i]\left(e^{-\eta\mathbf{f}_t[i]}\right)^2}{\left(\sum_{j=1}^n \mathbf{x}_t[j]e^{-\eta\mathbf{f}_t[j]}\right)^2} - 1 \\ &\leq \beta \sum_{i=1}^n \mathbf{x}_t[i](\eta\mathbf{f}_t[i])^2 = \beta(\|\eta\mathbf{f}_t\|_t^*)^2\end{aligned}$$

for a small constant β.

2) Regret guarantee for the self-concordant regularizer.: It was shown in [1] that, for the case of linear bandit optimization, the regularization function must have the property that it curves strongly near the boundary. Indeed, it was observed that the Hessian of $\mathcal{R}$ must behave roughly as inverse distance $1/d$, or even inverse squared distance $1/d^2$, to the boundary. Indeed, the entropy function discussed above possesses the former property on the n-simplex, but functions with this $1/d$ growth property are not readily available for general convex sets. To obtain a function whose Hessian grows as $1/d^2$ is much easier: the *self-concordant barrier*, commonly known as "log barrier", is the central object of study in Interior Point Methods. In particular, self-concordant barriers always exist and can be efficiently computed for many known bodies (see, e.g., [14]).

For a convex set with linear constraints, the typical choice of a self-concordant barrier is simply the sum of negative log distance to each boundary. That is, if the set is defined by $A\mathbf{x} \leq \mathbf{b}$, then we would let $\mathcal{R}(\mathbf{x}_t) = \sum_i -\log(b_i - \mathbf{e}_i^\intercal A\mathbf{x})$. It is true that, up to a constant, $\mathcal{R}$ is strongly convex with respect to the ℓ_2 norm, and we can then easily prove a bound in terms of $\sum_t \|\mathbf{f}_t\|_2^2$. On the other hand, it is precisely the case of bandit linear optimization for which it is useful to bound the regret in terms of the local norms $\|\mathbf{f}_t\|_{\mathbf{x}_t}^*$ as in (8). It was shown in [1] that the Hessian of a self-concordant barrier not only plays a crucial role in bounding the regret, but also gives a handle on the local geometry through the notion of a Dikin ellipsoid. We refer the reader to [1] for more information on the Dikin ellipsoid and its relation to sampling.

As before, we can use Hölder's in equality to bound

$$\mathbf{f}_t^\intercal(\mathbf{x}_t - \tilde{\mathbf{x}}_{t+1}) \leq \|\mathbf{f}_t\|_t^* \|\mathbf{x}_t - \mathbf{x}_{t+1}\|_t,$$

and now, as in the previous section, we would like to replace $\|\mathbf{x}_t - \mathbf{x}_{t+1}\|_t$ with the dual norm $\eta\|\mathbf{f}_t\|_t^*$. While it is not immediately obvious how this should be accomplished, we can appeal to several nice results about self-concordant functions which makes our job easy. Define the objective of Algorithm 1 as $\Phi_t(\mathbf{x}) = \eta\sum_{s=1}^t \mathbf{f}_t^\intercal\mathbf{x} + \mathcal{R}(\mathbf{x})$. Since the barrier $\mathcal{R}$ goes to infinity at the boundary of the set $\mathcal{K}$, we have that $\mathbf{x}_{t+1}$ is the unconstrained minimizer of Φ_t.

To begin our short journey to the land of Interior Point Methods, define the *Newton decrement* for Φ_t as

$$\lambda(\mathbf{x}, \Phi_t) := \|\nabla\Phi_t(\mathbf{x})\|_{\mathbf{x}}^* = \|\nabla^2\Phi_t(\mathbf{x})^{-1}\nabla\Phi_t(\mathbf{x})\|_{\mathbf{x}}$$

and note that since $\mathcal{R}$ is self-concordant then so is Φ_t. The above quantity can be used to measure roughly how far a point is from the global optimum:

Theorem 2.2 (e.g. [14]): For any self-concordant function g, whenever $\lambda(\mathbf{x}, g) \leq 1/2$, we have

$$\|\mathbf{x} - \arg\min g\|_{\mathbf{x}} \leq 2\lambda(\mathbf{x}, g)$$

where the local norm $\|\cdot\|_{\mathbf{x}}$ is defined with respect to g, i.e. $\|\mathbf{y}\|_{\mathbf{x}} := \sqrt{\mathbf{y}^\intercal(\nabla^2 g(\mathbf{x}))\mathbf{y}}$.

We can immediately apply this theorem using the objective Φ_t and the point $\mathbf{x}_t$. Recalling that $\nabla^2\Phi_t = \nabla^2\mathcal{R}$, we see that, under the conditions of the Theorem,

$$\|\mathbf{x}_t - \mathbf{x}_{t+1}\|_t = \|\mathbf{x}_t - \arg\min\Phi_t\|_t \leq 2\lambda(\mathbf{x}_t, \Phi_t) = 2\eta\|\mathbf{f}_t\|_t^*$$

The last equality holds because, as is easy to check, $\nabla\Phi_t(\mathbf{x}_t) = \eta\mathbf{f}_t$. We therefore have

Theorem 2.3: Suppose for all $t \in \{1 \dots T\}$ we have $\eta \|\mathbf{f}_t\|_t^* \le \frac{1}{2}$, and $\mathcal{R}(\cdot)$ is self-concordant. Then

$$R_T(\mathbf{u}) \le 2\eta \sum_{t=1}^{T} \left[\|\mathbf{f}_t\|_t^*\right]^2 + \eta^{-1}\mathcal{R}(\mathbf{u}).$$

Given Theorem 2.3, the result of Abernethy, Hazan, and Rakhlin [1] follows immediately, as we show in Section V-C.

III. Bandit Feedback

In the bandit version of online linear optimization, the function $\mathbf{f}_t$ is not revealed to us except for its value at $\mathbf{x}_t$. The mechanism employed by all algorithms known to the authors is to construct a biased or unbiased estimate $\tilde{\mathbf{f}}_t$ of the vector $\mathbf{f}_t$ from the single number revealed to us and feed it to the black box full-information algorithm. In order to construct $\tilde{\mathbf{f}}_t$, the algorithm has to randomly sample $\mathbf{y}_t$ around $\mathbf{x}_t$ instead of deterministically playing $\mathbf{x}_t$. Hence, the template bandit algorithm is: at round t to predict $\mathbf{y}_t$ such that $\mathbb{E}\mathbf{y}_t \approx \mathbf{x}_t$, obtain $\mathbf{f}_t^\intercal \mathbf{y}_t$, construct $\tilde{\mathbf{f}}_t$, feed it into the black box, and obtain the new $\mathbf{x}_{t+1}$. The particular method for sampling $\mathbf{y}_t$ and constructing $\tilde{\mathbf{f}}_t$ will be called *the sampling scheme*.

The regret of the above procedure, relative to a comparator $\mathbf{u}$, is

$$R_T(\mathbf{u}) = \sum_{t=1}^{T} \mathbf{f}_t^\intercal(\mathbf{y}_t - \mathbf{u}).$$

However, the guarantees for the black-box are for a different quantity, which we denote as

$$\tilde{R}_T(\mathbf{u}) = \sum_{t=1}^{T} \tilde{\mathbf{f}}_t^\intercal(\mathbf{x}_t - \mathbf{u}).$$

Let $\mathbb{E}_t$ denote the conditional expectation, given the random variables for time steps $1 \dots t-1$. If it is the case that $\mathbb{E}_t\tilde{\mathbf{f}}_t = \mathbf{f}_t$ and $\mathbb{E}_t\mathbf{y}_t = \mathbf{x}_t$, then for any *fixed* $\mathbf{u}$,

$$\mathbb{E}_t\left[\tilde{\mathbf{f}}_t^\intercal(\mathbf{x}_t - \mathbf{u})\right] = \mathbb{E}_t\left[\mathbf{f}_t^\intercal(\mathbf{y}_t - \mathbf{u})\right]. \tag{9}$$

We conclude that $\mathbb{E}R_T(\mathbf{u}) = \mathbb{E}\tilde{R}(\mathbf{u})$. Hence, *expected regret* against a fixed $\mathbf{u}$ can be bounded through the expected regret of the black-box.

There are two downsides to the above argument. The first is that an "in expectation" result is much weaker than the corresponding "high probability" statement as the variance of the quantities involved can be (and, in fact, is) very large. It is not very satisfying to say that the regret is of the correct order in expectation but has fluctuations of a higher order of magnitude. The second weakness is in the fact that $\mathbf{u}$ is fixed and, therefore, cannot depend on the random moves of the player; in other words, the adversary must be oblivious. Both of the downsides are overcome by proving a high probability guarantee.

It is tempting to use the following (incorrect) argument for proving a high-probability bound on $\mathcal{R}_T(\mathbf{u})$ given an $\tilde{O}(\sqrt{T})$ bound on $\mathbb{E}\tilde{R}_T(\mathbf{u})$: To obtain a high-probability bound, fix a $\mathbf{u} \in \mathcal{K}$ and use Azuma-Hoeffding inequality to show an $O(\sqrt{T})$ concentration of $R_T(\mathbf{u})$ around $\mathbb{E}R_T(\mathbf{u})$. Next, replace $\mathbb{E}R_T(\mathbf{u})$ by $\mathbb{E}\tilde{R}_T(\mathbf{u})$, which is $\tilde{O}(\sqrt{T})$, and take a union bound over a discretization of $\mathbf{u}$. The last step only introduces a $\log T$ factor into the bound, as we discuss later. This approach fails[3] for the simple reason that through the martingale difference argument $R_T(\mathbf{u})$ is concentrated around *the sum of conditional expectations* $\sum_{t=1}^{T}\mathbb{E}_t\tilde{\mathbf{f}}_t^\intercal(\mathbf{x}_t - \mathbf{u})$, not the full expectation $\mathbb{E}\sum_t \tilde{\mathbf{f}}_t^\intercal(\mathbf{x}_t - \mathbf{u})$. The sum of conditional expectations of $\mathbf{f}_t^\intercal(\mathbf{y}_t - \mathbf{u})$ terms is indeed equal to the sum of conditional expectations of $\tilde{\mathbf{f}}_t^\intercal(\mathbf{x}_t - \mathbf{u})$ terms. However, we do not know how to bound the latter: the regret guarantee for the black-box comes for the expected regret, not the sum of conditional expectations, thus breaking the argument.

Indeed, for proving high probability bounds, a more refined analysis is needed. We try to convey the big picture in the next section and illustrate it by proving a high-probability bound for the sphere and the simplex, using the regularization with self-concordant barrier and entropy, respectively, as black-boxes.

IV. High Probability Bounds

We now present a template algorithm for bandit optimization. We assume that a full-information black-box algorithm for linear optimization is available to us.

At each time step $t = 1$ to T,

- Decide on the sampling scheme for this round, i.e. construct a distribution for $\mathbf{y}_t$ with $\mathbb{E}_t\mathbf{y}_t \approx \mathbf{x}_t$.
- Draw a sample $\mathbf{y}_t \in \mathcal{K}$ from the distribution and observe the loss $\mathbf{f}_t^\intercal\mathbf{y}_t$.
- Construct $\tilde{\mathbf{f}}_t$ such that $\mathbb{E}_t\tilde{\mathbf{f}}_t = \mathbf{f}_t$.
- Construct a linear bias-function $g_t(\mathbf{u}) = \tilde{\mathbf{g}}_t^\intercal\mathbf{u} + \mu_t$.
- Feed $\tilde{\mathbf{f}}_t - \alpha\tilde{\mathbf{g}}_t$ into the black-box and receive $\mathbf{x}_{t+1}$.

The algorithm requires two parameters, α and η, which in turn depend on various aspects of the problem. The following is the main result of the paper.

Theorem 4.1: Suppose $\mathbf{f}_t \in B_p$ for all t and $\mathcal{K} \subseteq B_q$, where p and q are dual. Let $\alpha = \sqrt{\frac{\log(2\log(T)/\delta')}{nT}}$. Suppose we can find $c_1, c_2, c_3, c_4, c_5, c_6 \ge 0$, such that for all $t \in \{1, \dots, T\}$ all of the following hold:

(A) The black-box full information algorithm enjoys a regret bound of the form

$$R_T(\mathbf{u}) \le c_1\eta \sum_{t=1}^{T} \left[\|\mathbf{f}_t\|_t^*\right]^2 + \eta^{-1}\mathcal{R}(\mathbf{u})$$

with the "local" norm $\|\cdot\|_t$ defined by $\nabla^2\mathcal{R}(\mathbf{x}_t)$.

(B) $\|\mathbb{E}_t\mathbf{y}_t - \mathbf{x}_t\|_q \le c_2\sqrt{\frac{n}{T}}$.

(C) $|\tilde{\mathbf{f}}_t^\intercal\mathbf{u}| \le c_3\sqrt{nT}$ for all $\mathbf{u} \in \mathcal{K}$.

(D) We can construct a linear function $g_t(\mathbf{u}) = \tilde{\mathbf{g}}_t^\intercal\mathbf{u} + \mu_t$ such that

$$(\mathbf{x}_t - \mathbf{u})^\intercal\mathbb{E}_t\tilde{\mathbf{f}}_t\tilde{\mathbf{f}}_t^\intercal(\mathbf{x}_t - \mathbf{u}) \le g_t(\mathbf{u}) \quad \forall \mathbf{u} \in \mathcal{K}$$

and $$g_t(\mathbf{x}_t) \le c_4 n.$$

(E) The construction satisfies $\left[\left\|\tilde{\mathbf{f}}_t - \alpha\tilde{\mathbf{g}}_t\right\|_t^*\right]^2 \le c_5\sqrt{T}$.

(F) On average, the norm is small: $\mathbb{E}_t\left[\left\|\tilde{\mathbf{f}}_t - \alpha\tilde{\mathbf{g}}_t\right\|_t^*\right]^2 \le c_6$.

[3]We thank Ambuj Tewari for very helpful discussions in understanding this.

(G) Conditions for the regret bound in (A) to hold are satisfied (e.g. $\eta \left\| \tilde{\mathbf{f}}_t - \alpha \tilde{\mathbf{g}}_t \right\|_t^* \le \frac{1}{2}$ for log-barrier)

Then, for any fixed $\mathbf{u} \in \mathcal{K}$, with probability at least $1 - (\delta + \delta' + \delta'')$

$$\sum_{t=1}^{T} \mathbf{f}_t^\intercal (\mathbf{y}_t - \mathbf{u}) \le \eta^{-1} \mathcal{R}(\mathbf{u}) + \eta T A_1 + \sqrt{T} A_2,$$

where

$$A_1 = c_1 \left(c_6 + c_5 \sqrt{8 \log(1/\delta'')} \right)$$

and

$$A_2 = \sqrt{8 \log(1/\delta)} + c_2 \sqrt{n} + (2c_3 + c_4 + 2) \sqrt{n} \log(2 \log(T)/\delta').$$

Remark 4.1: As long as $c_1, \dots, c_6$ depend only "weakly" (e.g. logarithmically) on T, we obtain the optimal $\tilde{O}(\sqrt{T})$ dependence by setting $\eta \propto T^{-1/2}$. The growth of the bound in terms of n depends on the problem at hand and the sampling method.

Remark 4.2: To obtain a statement "with probability at least $1 - \delta$, for all $\mathbf{u}$ the guarantee holds", a union bound needs to be taken. For a set $\mathcal{K}$, which can be represented as a convex hull of a number of its vertices, the union bound introduces an extra logarithm of this number of vertices (see the simplex example below). For a set such as sphere, an extra step of discretizing the set into a fine grid and taking a union over this (exponential) discretization is required. This technique can introduce an extra $n \log T$ into the bound (see [10], [5] for details). Since this step depends on the particular $\mathcal{K}$ at hand, we leave it out of the main result.

Remark 4.3: The requirement (B) is a relaxation of $\mathbb{E}_t \mathbf{y}_t = \mathbf{x}_t$. This slack is absolutely crucial for (D) to be even possible. In the simplex case the slack corresponds to mixing in a uniform distribution, which Auer et al [3] interpret as an exploration step. For the sphere case, it corresponds to staying $O(T^{-1/2})$ away from the boundary. From the point of view of the proof, the relaxation allows us to construct g_t, i.e. to control the sum of conditional variances of $\tilde{\mathbf{f}}_t^\intercal (\mathbf{x}_t - \mathbf{u})$. We note that the slack is not necessary for bounding the expected regret only. This points to the large variance of the estimates and the weakness of the "in-expectation" results.

A. A Proof Sketch

Let us sketch the mechanism for proving high-probability bounds, which is applicable to a wide variety of sets and assumptions.

We already mentioned that $R_T(\mathbf{u})$ is concentrated, for a fixed $\mathbf{u} \in \mathcal{K}$ around the sum of conditional expectations $\sum_{t=1}^T \mathbb{E}_t \mathbf{f}_t^\intercal (\mathbf{y}_t - \mathbf{u})$ with typical deviations of $O(\sqrt{T})$. The latter is equal to the sum of conditional expectations $\sum_{t=1}^T \mathbb{E}_t \tilde{\mathbf{f}}_t^\intercal (\mathbf{x}_t - \mathbf{u})$. The tricky part is in proving that $\tilde{R}_T(\mathbf{u})$ is concentrated around this sum. The typical fluctuations of $\tilde{R}(\mathbf{u})$ are more than $\sqrt{T}$, as the magnitude of $\tilde{\mathbf{f}}_t$ depends on T. Thus, the only statement we can make is that, with high probability, $\sum_{t=1}^T \mathbb{E}_t \tilde{\mathbf{f}}_t^\intercal (\mathbf{x}_t - \mathbf{u}) \le \sum_{t=1}^T \tilde{\mathbf{f}}_t^\intercal (\mathbf{x}_t - \mathbf{u}) + c\sqrt{\mathrm{Var}}$, where Var is the sum of conditional variances, growing faster than linear in T. The magic comes from splitting the $\sqrt{\mathrm{Var}}$ term into T terms by the arithmetic-geometric mean inequality and absorbing each of these terms into $\tilde{\mathbf{f}}_t$, thereby biasing the estimates. At a high level, we are adding the standard deviation at each time step to the estimates $\tilde{\mathbf{f}}_t$. Since this confidence interval is a concave function, the black-box optimization over the modified $\tilde{\mathbf{f}}_t$'s will not work; the second magic step (due to this paper) is to find a linear function which uniformly bounds the confidence over the whole set $\mathcal{K}$. If this can be done, the *modified* linear functions are fed to the black-box, which enjoys an upper bound of $\eta \sum_{t=1}^T (\|\tilde{\mathbf{f}}'_t\|_t^*)^2$, with the norms of modified functions. Finally, we show that this quantity is concentrated around the sum of conditional expectations of the terms with the typical deviations of $O(\sqrt{T})$, and the sum of conditional expectations itself is bounded by $O(\sqrt{T})$ if $\tilde{\mathbf{f}}_t$'s have been constructed carefully. The last result critically depends on availability of a regret guarantee with *local* norms, which have been exhibited earlier in the paper.

The above paragraph is an informal description of the proof, which can be found in Section VI. We refer to [2] for the details.

V. Applications: Theorem 4.1 at work

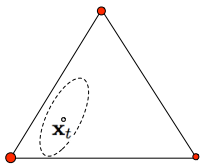
For the sampling schemes below, we show that our construction satisfies conditions of Theorem 4.1, implying a high-probability guarantee of $\tilde{O}(\sqrt{T})$.

For each scheme, we provide a visual depiction of the distribution from which we draw $\mathbf{y}_t$. The size of the dots represents the relative probability mass, while the dotted ellipsoid represents a sphere in the local norm at $\mathbf{x}_t$. Note that in the case of self-concordant $\mathcal{R}$, this ellipsoid (the Dikin ellipsoid) is contained in the set, which allows us to sample from its eigenvectors (see [1]).

A. Example 1: Solution for the simplex

This case corresponds to the non-stochastic multiarmed bandit problem [3]. We assume that $\mathcal{K}$ is the simplex (i.e. $q = 1$) and $0 \le \mathbf{f}_t[i] \le 1$ ($p = \infty$).

- **Regularizer $\mathcal{R}$:** We set our regularization function to be the entropy (6) and use Algorithm 1 or 2 as the black-box.
- **Sampling of $\mathbf{y}_t$:**



Let $\gamma = \sqrt{\frac{n}{T}}$. Given the point $\mathbf{x}_t$ in the simplex, sample $\mathbf{y}_t = \mathbf{e}_i$ with prob. $\mathbf{p}_t[i] := (1 - \gamma)\mathbf{x}_t[i] + \gamma/n$.

- **Construction of $\tilde{\mathbf{f}}_t$:** Given the above sampling scheme, we define our estimates $\tilde{\mathbf{f}}_t$ the usual way:

$$\tilde{\mathbf{f}}_t = \frac{(\mathbf{f}_t^\intercal \mathbf{e}_i)\mathbf{e}_i}{\mathbf{p}_t[i]} = \frac{\mathbf{f}_t[i]\mathbf{e}_i}{\mathbf{p}_t[i]} \quad \text{when } \mathbf{y}_t = \mathbf{e}_i. \tag{10}$$

- **Construction of $\tilde{\mathbf{g}}_t$:** The following g_t is appropriate for this problem:

$$g_t(\mathbf{u}) := 2 + \sum_{i=1}^{n} \frac{\mathbf{e}_i^\intercal \mathbf{u}}{\mathbf{p}_t[i]}.$$

Before we get started, we note a couple of useful facts that we use several times below:

$$\mathbf{x}_t[i] \leq \frac{\mathbf{p}_t[i]}{1-\gamma} \qquad \mathbf{p}_t[i]^{-1} \leq \frac{n}{\gamma}$$

Now we check the conditions of the theorem.

(A) Since we are using entropy as our regularization, we have already shown in Theorem 2.1 how to obtain the necessary bound with $c_1 = 1$.

(B) Notice that $\mathbb{E}\mathbf{y}_t = (1-\gamma)\mathbf{x}_t + \gamma \text{unif}(n)$ and thus $\|\mathbb{E}\mathbf{y}_t - \mathbf{x}_t\|_1 = \gamma\|\mathbf{x}_t - \text{unif}(n)\|_1 \leq 2\sqrt{\frac{n}{T}}$, i.e. $c_2 = 2$.

(C) Since $\mathbf{u}$ is in the simplex, we see that $c_3 = 1$:

$$|\tilde{\mathbf{f}}_t^\intercal \mathbf{u}| \;\leq\; \max_i |\tilde{\mathbf{f}}_t[i]| \;\leq\; \max_i \mathbf{p}_t[i]^{-1} \;\leq\; \frac{n}{\gamma} \;=\; \sqrt{nT}.$$

(D) We check that g_t does indeed bound the variance. We can first compute

$$\mathbb{E}_t \tilde{\mathbf{f}}_t \tilde{\mathbf{f}}_t^\intercal = \sum_{i=1}^{n} \mathbf{p}_t[i] \left(\frac{\mathbf{f}_t[i]}{\mathbf{p}_t[i]}\right)^2 \mathbf{e}_i \mathbf{e}_i^\intercal \preceq \sum_{i=1}^{n} \mathbf{p}_t[i]^{-1} \mathbf{e}_i \mathbf{e}_i^\intercal.$$

We can now upper bound the variance of the estimated losses, but we need to do this on the entire simplex. Fortunately, since we are upper bounding a quadratic (a convex function) it suffices to check the corners $\mathbf{u} = \mathbf{e}_i$:

$$\begin{aligned}
&(\mathbf{x}_t - \mathbf{e}_i)^\intercal \mathbb{E}_t \tilde{\mathbf{f}}_t \tilde{\mathbf{f}}_t^\intercal (\mathbf{x}_t - \mathbf{e}_i) \leq \sum_{j=1}^{n} \frac{(\mathbf{x}_t[j] - \mathbf{1}[i=j])^2}{\mathbf{p}_t[j]} \\
&< \frac{1}{\mathbf{p}_t[i]} + \sum_{j \neq i} \frac{\mathbf{x}_t[j]^2}{\mathbf{p}_t[j]} \leq \frac{1}{\mathbf{p}_t[i]} + \sum_{j \neq i} \frac{(\mathbf{p}_t[j])^2}{(1-\gamma)^2 \mathbf{p}_t[j]} \\
&\leq \quad 2 + \frac{1}{\mathbf{p}_t[i]} \quad = \quad g_t(\mathbf{e}_i).
\end{aligned}$$

where we use the fact that $(1-\gamma)^2 \geq 1/2$ when $T \geq 16n$. Additionally, we see that $c_4 = 3$:

$$g_t(\mathbf{x}_t) = 2 + \sum_{i=1}^{n} \frac{\mathbf{x}_t[i]}{\mathbf{p}_t[i]} \leq 2 + \sum_{i=1}^{n} \frac{\frac{\mathbf{p}_t[i]}{1-\gamma}}{\mathbf{p}_t[i]} \leq 3n.$$

(E) We now check that, in the $\mathbf{x}_t$-norm, the biased estimate is not too big. It is easy to check that

$$\nabla^2_{\mathbf{x}_t} \mathcal{R} = \sum_{i=1}^{n} \mathbf{x}_t[i]^{-1} \mathbf{e}_i \mathbf{e}_i^\intercal \Rightarrow \nabla^2_{\mathbf{x}_t} \mathcal{R}^{-1} = \sum_{i=1}^{n} \mathbf{x}_t[i] \mathbf{e}_i \mathbf{e}_i^\intercal.$$

Now assume $\mathbf{y}_t = \mathbf{e}_j$, we can bound:

$$\begin{aligned}
\|\tilde{\mathbf{f}}_t - \alpha \tilde{\mathbf{g}}_t\|^{*2}_{\mathbf{x}_t} &= \sum_{i=1}^{n} \left(\frac{\mathbf{1}[i=j] - \alpha}{\mathbf{p}_t[i]}\right)^2 \mathbf{x}_t[i] \\
&\leq \sum_{i=1}^{n} \left(\frac{\mathbf{1}[i=j] - \alpha}{\mathbf{p}_t[i]}\right)^2 \left(\frac{\mathbf{p}_t[i]}{1-\gamma}\right) \\
&\leq 2\frac{\alpha^2 n^2}{\gamma} + 2\frac{n(1-\alpha)^2}{\gamma}
\end{aligned}$$

Substituting $\gamma = \sqrt{\frac{n}{T}}$, we obtain $c_5 = 2\alpha^2 n^{3/2} + 2\sqrt{n}(1-\alpha)^2$.

(F) We also must check that, in expectation, the biased estimate is of constant order in the $\mathbf{x}_t$-norm:

$$\begin{aligned}
\mathbb{E}_t \left[\left\|\tilde{\mathbf{f}}_t - \alpha \tilde{\mathbf{g}}_t\right\|^*\right]^2 &\leq \mathbb{E}_t \left[2\|\tilde{\mathbf{f}}_t\|^{*2}_{\mathbf{x}_t} + 2\alpha^2 \|\tilde{\mathbf{g}}_t\|^{*2}_{\mathbf{x}_t}\right] \\
&\leq 2\left(\sum_{i=1}^{n} \mathbf{p}_t[i] \frac{1}{\mathbf{p}_t[i]^2} \mathbf{x}_t[i] + \alpha^2 \sum_{i=1}^{n} \frac{1}{\mathbf{p}_t[i]^2} \mathbf{x}_t[i]\right) \\
&\leq \frac{2}{1-\gamma}\left(n + \alpha^2 \sum_{i=1}^{n} \frac{1}{\mathbf{p}_t[i]}\right) \quad \leq \quad 4\left(n + \frac{\alpha^2 n^2}{\gamma}\right) \\
&= 4(n + (\alpha^2\sqrt{T})n^{3/2}) =: c_6 \,.
\end{aligned}$$

We conclude that

$$\begin{aligned}
A_1 &= 4(n + (\alpha^2\sqrt{T})n^{3/2}) \\
&\quad + \left(2\alpha^2 n^{3/2} + 2\sqrt{n}(1-\alpha)^2\right)\sqrt{8\log(1/\delta'')}
\end{aligned}$$

and

$$A_2 = \sqrt{8\log(1/\delta)} + 2\sqrt{n} + 7\sqrt{n}\log(2\log(T)/\delta').$$

Now we switch to the Big-O notation to elucidate the dependence on T and n. Recalling that $\alpha^2 = O(\log\log T/(nT))$, we observe that $A_1 = O(n)$ and $A_2 = O(\sqrt{n}\log\log T)$. Theorem 4.1 now states that with probability at least $1 - (\delta + \delta' + \delta'')$,

$$\sum_{t=1}^{T} \mathbf{f}_t^\intercal (\mathbf{y}_t - \mathbf{u}) \leq \eta^{-1}\mathcal{R}(\mathbf{u}) + \eta T A_1 + \sqrt{T} A_2$$

for any fixed $\mathbf{u}$. Since the regret is a linear functional, it attains its maximum at one of the vertices of the simplex. Hence, unlike in the next section, we only need to take a union bound over these vertices to arrive at a statement for all $\mathbf{u} \in \mathcal{K}$. We thus set $\delta = \delta' = \delta'' = \delta^*/n$. Observe that A_1's asymptotic dependence on n does not change, while A_2 now becomes $O(\sqrt{n}\log(n\log T))$.

For any vertex $\mathbf{u}$, the (shifted) entropy is $\mathcal{R}(\mathbf{u}) = \log n$. Setting $\eta = \sqrt{\frac{\log n}{nT}}$, we conclude that, with high probability,

$$\forall \mathbf{u} \in \mathcal{K}, \quad \sum_{t=1}^{T} \mathbf{f}_t^\intercal (\mathbf{y}_t - \mathbf{u}) = O(\sqrt{Tn}\log(n\log T)).$$

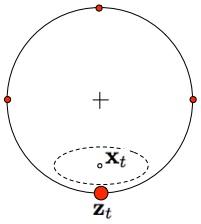
This bound improves upon the result of Auer et al [3], who obtained an $O(\sqrt{nT\log(nT)})$ bound for the problem (Algorithm EXP3.P). Our bound replaces the $\log T$ term with $\log\log T$, closing the gap to the lower bound of $\Omega(\sqrt{nT})$. We conjecture that $\sqrt{T\log\log T}$ growth in terms of T is the most sharp bound possible, due to the Law of the Iterated Logarithm. In the full version of the paper, we will use sharper concentration inequalities to keep $\log\log T$ under the square root.

B. Example 2: Solution for the Euclidean sphere

Suppose that $\mathcal{K} = B_2 \subset \mathbb{R}^n$ and that the choices of the adversary are also ℓ_2-bounded by 1, i.e. $p = q = 2$.

We point out that with the sampling scheme of [1] it is impossible to construct g_t to satisfy the requirements of Theorem 4.1 (see also Section V-C). The modified sampling procedure below is key to reducing the variance of the estimates.

- **Regularizer $\mathcal{R}$:** We set our regularization function to be the standard log-barrier $\mathcal{R}(\mathbf{x}) = -\log(1 - \|\mathbf{x}\|^2)$ for the sphere and use Algorithm 1 as the black-box.
- **Sampling of $\mathbf{y}_t$:** We can assume without loss of generality that $\mathbf{x}_t \neq \mathbf{0}$, so define $\mathbf{z}_t := \mathbf{x}_t/\|\mathbf{x}_t\|_2$. Towards the goal of keeping our sampled point $\mathbf{y}_t$ away from the boundary, define $\gamma := \max(1 - \|\mathbf{x}_t\|_2, \sqrt{\frac{n}{T}})$. Now construct some $n-1$ orthonormal basis of the subspace perpendicular to $\mathbf{z}_t$, which we will call $\mathrm{Perp}(\mathbf{z}_t)$.



Sample our prediction $\mathbf{y}_t$ as follows:

$$\mathbf{y}_t = \begin{cases} \mathbf{z}_t & \text{w.p. } 1 - \frac{3\gamma}{4} \\ -\mathbf{z}_t & \text{w.p. } \frac{\gamma}{4} \\ \pm\mathbf{w} \in \mathrm{Perp}(\mathbf{z}_t) & \text{w.p. } \frac{\gamma}{4(n-1)} \end{cases} \tag{11}$$

- **Construction of $\tilde{\mathbf{f}}_t$:** Given the above sampling scheme, we define our estimates $\tilde{\mathbf{f}}_t$ as follows,

$$\tilde{\mathbf{f}}_t = \frac{(\mathbf{f}_t^T \mathbf{y}_t)\mathbf{y}_t}{2\Pr(\mathbf{y}_t)} \tag{12}$$

  where the probabilities $\Pr(\cdot)$ are defined in equation (11). It is straightforward to check that $\mathbb{E}_t \tilde{\mathbf{f}}_t = \mathbf{f}_t$.
- **Construction of $\tilde{\mathbf{g}}_t$:** The following choice of g_t will be shown to satisfy the requirements:

$$g_t(\mathbf{u}) := 4n\left(3 + \frac{2 - 2\mathbf{z}_t^\mathsf{T}\mathbf{u}}{\gamma}\right).$$

We now check the conditions of the theorem to verify that this construction leads to a high-probability bound.

(A) Since we are using a self-concordant regularizer, we already showed in Theorem 2.3 how to obtain the necessary bound with $c_1 = 2$.

(B) Notice that $\mathbb{E}\mathbf{y}_t = (1-\gamma)\mathbf{z}_t = \frac{(1-\gamma)\mathbf{x}_t}{\|\mathbf{x}_t\|_2}$ and because $|(1-\gamma) - \|\mathbf{x}_t\|_2| \le \sqrt{\frac{n}{T}}$ it follows that $\|\mathbb{E}\mathbf{y}_t - \mathbf{x}_t\| \le \sqrt{\frac{n}{T}}$. Hence, $c_2 = 1$.

(C) Since we assume that $\|\mathbf{u}\|_2 \le 1$, we see that $c_3 = 2$:

$$\begin{aligned} |\tilde{\mathbf{f}}_t^\mathsf{T}\mathbf{u}| \;&\le\; \|\tilde{\mathbf{f}}_t\|_2\|\mathbf{u}\|_2 \;\le\; \|\tilde{\mathbf{f}}_t\|_2 \\ &\le\; (2\Pr(\mathbf{y}_t))^{-1} \;\le\; \frac{2n}{\gamma} \;\le\; 2\sqrt{nT}. \end{aligned}$$

(D) We check that g_t does indeed bound the variance. We first upper bound the matrix $\mathbb{E}_t\tilde{\mathbf{f}}_t\tilde{\mathbf{f}}_t^\mathsf{T}$:

$$\begin{aligned} \mathbb{E}_t\tilde{\mathbf{f}}_t\tilde{\mathbf{f}}_t^\mathsf{T} &= \sum_{\mathbf{y}_t} \Pr(\mathbf{y}_t)\left(\frac{\mathbf{f}_t^\mathsf{T}\mathbf{y}_t}{2\Pr(\mathbf{y}_t)}\right)^2 \mathbf{y}_t\mathbf{y}_t^\mathsf{T} \\ &\preceq \quad \frac{1}{2}(\max_{\mathbf{y}_t}\Pr(\mathbf{y}_t)^{-1})I_n, \end{aligned}$$

since the range of $\mathbf{y}_t$ is over $\pm$ vectors from an orthonormal basis. By construction each of these probabilities is $\ge \gamma/(4n)$. Now we can bound:

$$\begin{aligned} (\mathbf{x}_t - \mathbf{u})^\mathsf{T}\mathbb{E}_t\tilde{\mathbf{f}}_t\tilde{\mathbf{f}}_t^\mathsf{T}(\mathbf{x}_t - \mathbf{u}) &\le \frac{2n}{\gamma}\|\mathbf{x}_t - \mathbf{u}\|_2^2 \\ &\le \frac{2n}{\gamma}\left(2\|\mathbb{E}_t\mathbf{y}_t - \mathbf{u}\|_2^2 + 2\|\mathbf{x}_t - \mathbb{E}_t\mathbf{y}_t\|_2^2\right) \\ &\le \frac{4n}{\gamma}\|(1-\gamma)\mathbf{z}_t - \mathbf{u}\|_2^2 + 4n \\ &\le \frac{4n}{\gamma}\left(2 - 2\mathbf{u}^\mathsf{T}\mathbf{z}_t + 2\gamma\mathbf{u}^\mathsf{T}\mathbf{z}_t\right) + 4n \le g_t(\mathbf{u}). \end{aligned}$$

Additionally, we check that the bias is not large at $\mathbf{x}_t$. Recalling that $\mathbf{z}_t = \mathbf{x}_t/\|\mathbf{x}_t\|$ and since $\gamma \ge 1 - \|\mathbf{x}_t\|$ by construction,

$$g_t(\mathbf{x}_t) = 4n\left(3 + 2\frac{1 - \|\mathbf{x}_t\|}{\gamma}\right) \le 20n, \quad \text{i.e. } c_4 = 20.$$

(E) We now check that, in the $\mathbf{x}_t$-norm, the biased estimate is not too big. We can roughly lower bound

$$\begin{aligned} \nabla^2_{\mathbf{x}_t} R &= \frac{2}{1 - \|\mathbf{x}_t\|^2}I + \frac{4}{(1 - \|\mathbf{x}_t\|^2)^2}\mathbf{x}_t\mathbf{x}_t^\mathsf{T} \\ &\succeq \frac{1}{1 - \|\mathbf{x}_t\|}I + \frac{\|\mathbf{x}_t\|^2}{(1 - \|\mathbf{x}_t\|)^2}\mathbf{z}_t\mathbf{z}_t^\mathsf{T} \end{aligned}$$

where we used that $\frac{1}{1-\|\mathbf{x}_t\|^2} = \frac{1}{(1-\|\mathbf{x}_t\|)(1+\|\mathbf{x}_t\|)} \ge \frac{1}{2(1-\|\mathbf{x}_t\|)}$ whenever $\|\mathbf{x}_t\| \in [0,1]$. This tells us that the eigenvalues of $\nabla^2_{\mathbf{x}_t} R$ are bounded from below $(1 - \|\mathbf{x}_t\|)^{-1}$ in all directions orthogonal to $\mathbf{x}_t$, and by $\frac{1}{1-\|\mathbf{x}_t\|}\left(1 + \frac{\|\mathbf{x}_t\|^2}{1-\|\mathbf{x}_t\|}\right)$ in the direction of $\mathbf{x}_t$. Thus,

$$\begin{aligned} \nabla^2_{\mathbf{x}_t} R^{-1} &\preceq (1 - \|\mathbf{x}_t\|)(I - \mathbf{z}_t\mathbf{z}_t^\mathsf{T}) + \frac{(1 - \|\mathbf{x}_t\|)^2}{1 - \|\mathbf{x}_t\| + \|\mathbf{x}_t\|^2}\mathbf{z}_t\mathbf{z}_t^\mathsf{T} \\ &\preceq (1 - \|\mathbf{x}_t\|)(I - \mathbf{z}_t\mathbf{z}_t^\mathsf{T}) + 2(1 - \|\mathbf{x}_t\|)^2\mathbf{z}_t\mathbf{z}_t^\mathsf{T} \end{aligned}$$

where the last inequality holds since $1 - x + x^2 > 1/2$ when $x \in [0,1]$. Now that we have control of the norm $\nabla^2_{\mathbf{x}_t} R^{-1}$, we can bound

$$\begin{aligned} \|\tilde{\mathbf{g}}_t\|^{*2}_{\mathbf{x}_t} \le \tilde{\mathbf{g}}_t^\mathsf{T}\nabla^2_{\mathbf{x}_t} R^{-1}\tilde{\mathbf{g}}_t &= \left(\frac{8n}{\gamma}\right)^2 \mathbf{z}_t^\mathsf{T}\nabla^2_{\mathbf{x}_t} R^{-1}\mathbf{z}_t \\ &\le \frac{64n^2 \cdot 2(1 - \|\mathbf{x}_t\|)^2}{\gamma^2} \;\le\; 128n^2 \end{aligned}$$

If $\mathbf{y}_t = \mathbf{z}_t$ or $-\mathbf{z}_t$,

$$\|\tilde{\mathbf{f}}_t\|^{*2}_{\mathbf{x}_t} \le (\frac{4}{\gamma})^2\mathbf{z}_t^\mathsf{T}\nabla^2_{\mathbf{x}_t} R^{-1}\mathbf{z}_t \le \frac{16}{\gamma^2}2(1 - \|\mathbf{x}_t\|)^2 \le 32,$$

If $\pm\mathbf{y}_t \in \mathrm{Perp}(\mathbf{z}_t)$,

$$\|\tilde{\mathbf{f}}_t\|^{*2}_{\mathbf{x}_t} \le \left(\frac{2(n-1)}{\gamma}\right)^2 \mathbf{y}_t^\intercal \nabla^2_{\mathbf{x}_t} R^{-1} \mathbf{y}_t \le \frac{4(n-1)^2}{\gamma^2}(1-\|\mathbf{x}_t\|) \le 4n^{3/2}T^{1/2}.$$

These last two bounds give us, for large enough T:

$$\|\tilde{\mathbf{f}}_t - \alpha\tilde{\mathbf{g}}_t\|^{*2}_{\mathbf{x}_t} \le\ 8n^{3/2}T^{1/2} + 128n^2\alpha^2$$

i.e. $c_5 = 8n^{3/2} + \frac{128n^2\alpha^2}{\sqrt{T}}$.

(F) We also must check that, in expectation, the biased estimate is of constant order in the $\mathbf{x}_t$-norm:

$$\mathbb{E}_t\left[\left\|\tilde{\mathbf{f}}_t - \alpha\tilde{\mathbf{g}}_t\right\|_t^{*}\right]^2 \le \mathbb{E}_t\left[2\|\tilde{\mathbf{f}}_t\|^{*2}_{\mathbf{x}_t} + 2\alpha^2\|\tilde{\mathbf{g}}_t\|^{*2}_{\mathbf{x}_t}\right]$$
$$\le 2\sum_{\mathbf{y}\in\{\pm\mathbf{z}_t,\mathrm{Perp}(\mathbf{z}_t)\}} \Pr(\mathbf{y})\left[\|\tilde{\mathbf{f}}_t\|^{*2}_{\mathbf{x}_t} \mid \mathbf{y}_t = \mathbf{y}\right] + 128n^2\alpha^2$$
$$< 64 + 4n^2 + 128n^2\alpha^2 =: c_6\ .$$

(G) Theorem 2.3 comes with the requirement that $\eta\|\tilde{\mathbf{f}}_t - \alpha\tilde{\mathbf{g}}_t\|^*_t \le 1/2$. From (E), $\|\tilde{\mathbf{f}}_t - \alpha\tilde{\mathbf{g}}_t\|^*_{\mathbf{x}_t} = O(T^{1/4})$. By taking $\eta = O(T^{-1/2})$, the requirement is satisfied for large enough T.

We conclude that

$$A_1 = 2\left(64 + 4n^2 + 128n^2\alpha^2 + \left(8n^{3/2} + \frac{128n^2\alpha^2}{\sqrt{T}}\right)\sqrt{8\log(1/\delta'')}\right)$$

and

$$A_2 = \sqrt{8\log(1/\delta)} + \sqrt{n} + 26\sqrt{n}\log(2\log(T)/\delta').$$

Recalling that $\alpha = \sqrt{\frac{\log(2\log(T)/\delta')}{nT}}$, we observe that

$$A_1 = O(n^2) \quad \text{and} \quad A_2 = O(\sqrt{n}\log\log T).$$

Theorem 4.1 then gives us, with $\eta = \frac{\sqrt{\log T}}{n\sqrt{T}}$,

$$\sum_{t=1}^{T}\mathbf{f}_t^\intercal(\mathbf{y}_t - \mathbf{u}) \le \eta^{-1}\mathcal{R}(\mathbf{u}) + \eta T A_1 + \sqrt{T}A_2 = O(n\sqrt{T\log T})$$

with high probability for any $\mathbf{u} \in \mathcal{K}$ which is $T^{-1/2}$ away from the boundary. The asymptotic behavior in terms of n and T exactly matches the "in-expectation" result of [1], as the self-concordance parameter $\vartheta = 1$ for the sphere. Now, to make the result uniform for any $\mathbf{u}$, we discretize the set $\mathcal{K}$ into a grid of size $T^{n/2}$ and take a union bound for all $\mathbf{u}$ in this set (see [9], [5] for details). Setting $\delta = \delta' = \delta'' = \frac{\delta^*}{T^{n/2}}$ leads to replacing all three "$\log 1/\delta$" terms by $n\log T + \log 1/\delta^*$. Inspecting A_1, we observe that this substitution introduces $\sqrt{n\log T}$ in front of $8n^{3/2}$, which, when balanced with η, exhibits $\eta T A_1 = O(n\sqrt{T}\log T)$ behavior. However, $A_2 = O(n^{3/2}\sqrt{T}\log T)$ now becomes the dominating term, as the $\log\log T/\delta'$ is not under the square root. We conclude that, with high probability,

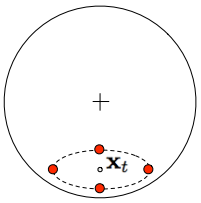
$$\forall \mathbf{u} \in \mathcal{K}, \quad \sum_{t=1}^{T}\mathbf{f}_t^\intercal(\mathbf{y}_t - \mathbf{u}) = O(n^{3/2}\sqrt{T}\log T).$$

A more careful analysis, involving a sharper inequality in one of the steps of the proof of Theorem 4.1 (see [2]), should reduce the dependence on n to linear. This will be carried out in the full version of this paper.

C. Example 3: Recovering the result of [1]

While it does not require Theorem 4.1, for the sake of completeness we show that the in-expectation result of [1] immediately follows from Theorem 2.3. For any convex set, the sampling procedure proposed in that paper is

- **Regularizer $\mathcal{R}$:** The regularization function is a ϑ-self-concordant barrier for $\mathcal{K}$, whose existence is guaranteed (see [15], [14]). We use Algorithm 1 as the black-box.
- **Sampling of $\mathbf{y}_t$:**



Let $\{\mathbf{e}_1,\dots,\mathbf{e}_n\}$ and $\{\lambda_1,\dots,\lambda_n\}$ be the set of eigenvectors and eigenvalues of $\nabla^2\mathcal{R}(\mathbf{x}_t)$. Choose i_t uniformly at random from $\{1,\dots,n\}$ and $\varepsilon_t = \pm 1$ with probability $1/2$. Sample $\mathbf{y}_t = \mathbf{x}_t + \varepsilon_t\lambda_{i_t}^{-1/2}\mathbf{e}_{i_t}$.

- **Construction of $\tilde{\mathbf{f}}_t$:** Define $\tilde{\mathbf{f}}_t := n\,(\mathbf{f}_t^\intercal\mathbf{y}_t)\,\varepsilon_t\lambda_{i_t}^{1/2}\cdot\mathbf{e}_{i_t}$.

Since here we are not interested in high-probability bounds, we do not need to construct $\tilde{\mathbf{g}}_t$. Appealing to (9) and Theorem 2.3, it only remains to bound $\|\tilde{\mathbf{f}}_t\|^*_t$. By construction, $(\|\tilde{\mathbf{f}}_t\|^*_t)^2 = \tilde{\mathbf{f}}_t^\intercal\nabla^{-1}R(\mathbf{x}_t)\tilde{\mathbf{f}}_t \le n^2$. For any $\mathbf{u}$ which is $T^{-1/2}$ away from the boundary, $\mathcal{R}(\mathbf{u}) \le 2\vartheta\log T$ (see [1]). Thus, with $\eta = \frac{\sqrt{\vartheta\log T}}{n\sqrt{T}}$, we obtain

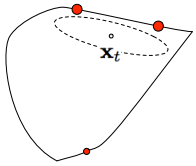
$$\mathbb{E}R_T(\mathbf{u}) \le 4n\sqrt{\vartheta T\log T},$$

which recovers the in-expectation result with a slightly better constant.

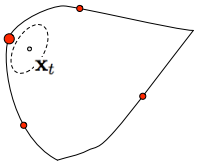
The sampling scheme presented here does not satisfy conditions of Theorem 4.1. Indeed, following the discussion in Remark 4.3, it is easy to prove (even for $\mathcal{K} = [0,1]$) that it is impossible to construct g_t with the desired properties. In other words, the variance of the estimates is larger than the desired regime. This realization was indeed the main motivation for this paper.

D. Example 4: Sampling schemes for general bodies

We remark that, while $\mathcal{R}$ has to be fixed throughout the game, the sampling scheme does not. As long as the requirements of Theorem 4.1 are satisfied at each step, the high probability bound holds true. The main difficulty in obtaining a result for general convex bodies $\mathcal{K}$ is in construction of $g_t(\mathbf{u})$, an upper-bound on the variance. Such a function heavily depends on the geometry and must be constructed on per-case basis. We conjecture that the following two sampling schemes, one for the curved boundary (similar to the spherical case) and one for the flat boundary (similar to the simplex case), should be enough to deal with most "nice" sets $\mathcal{K}$.



Put large mass (e.g. $O(\frac{1}{n})$ on $n-1$ points along the flat boundary, and put a small probability mass on a far away point.



As in the spherical case, put large mass (close to 1) on a single point close to $\mathbf{x}_t$ and small mass on $2n-1$ other points far away.

VI. Proofs

We state four lemmas whose proof can be found in the technical report [2].

Lemma 6.1: With probability at least $1-\delta$,

$$\sum_{t=1}^{T}\mathbf{f}_t^{\mathsf{T}}(\mathbf{y}_t-\mathbf{u})\le\sum_{t=1}^{T}\mathbf{f}_t^{\mathsf{T}}(\mathbf{x}_t-\mathbf{u})+\sqrt{8T\log(1/\delta)}+c_2\sqrt{nT}.$$

The following lemma is based on a result proved in [5].

Lemma 6.2: For any $\delta<e^{-1}$ and $T\ge 4$, with probability at least $1-2\log(T)\delta$,

$$\begin{aligned}\sum_{t=1}^{T}\mathbf{f}_t^{\mathsf{T}}(\mathbf{x}_t-\mathbf{u})&\le\tilde{R}_T(\mathbf{u})\\&+2\max\left\{2\sqrt{\sum_{t=1}^{T}(\mathbf{x}_t-\mathbf{u})^{\intercal}\mathbb{E}_t\tilde{\mathbf{f}}_t\tilde{\mathbf{f}}_t^{\intercal}(\mathbf{x}_t-\mathbf{u})},\right.\\&\left.(1+2c_3\sqrt{nT})\sqrt{\log(1/\delta)}\right\}\sqrt{\log(1/\delta)}.\end{aligned}$$

Lemma 6.3: For any $\delta<e^{-1}$ and $T\ge 4$, with probability at least $1-\delta'$,

$$\begin{aligned}\sum_{t=1}^{T}\mathbf{f}_t^{\mathsf{T}}(\mathbf{x}_t-\mathbf{u})&\le\sum_{t=1}^{T}\left(\tilde{\mathbf{f}}_t-\alpha\tilde{\mathbf{g}}_t\right)^{\mathsf{T}}(\mathbf{x}_t-\mathbf{u})\\&+\left[(2c_3+c_4+2)\sqrt{nT}\right]\log(2\log(T)/\delta').\end{aligned}$$

The final ingredient is the following concentration result.

Lemma 6.4: With probability at least $1-\delta''$,

$$\begin{aligned}\eta\sum_{t=1}^{T}\left[\left\|\tilde{\mathbf{f}}_t-\alpha\tilde{\mathbf{g}}_t\right\|^{*}\right]^2\le&\eta\sum_{t=1}^{T}\mathbb{E}_t\left[\left\|\tilde{\mathbf{f}}_t-\alpha\tilde{\mathbf{g}}_t\right\|^{*}\right]^2\\&+\eta Tc_5\sqrt{8\log(1/\delta'')}.\end{aligned}$$

Combining the above lemmas, we now prove the Theorem.

Proof: [Proof of Theorem 4.1] Combining Lemma 6.1 and Lemma 6.3, we obtain that

$$\begin{aligned}\sum_{t=1}^{T}\mathbf{f}_t^{\mathsf{T}}(\mathbf{y}_t-\mathbf{u})\le&\sum_{t=1}^{T}\left(\tilde{\mathbf{f}}_t-\alpha\tilde{\mathbf{g}}_t\right)^{\mathsf{T}}(\mathbf{x}_t-\mathbf{u})+\sqrt{8T\log(1/\delta)}\\&+c_2\sqrt{nT}+(2c_3+c_4+2)\sqrt{nT}\log(2\log T/\delta')\end{aligned}$$

with probability at least $1-(\delta+\delta')$. By the black-box guarantee applied to functions $\left(\tilde{\mathbf{f}}_t-\alpha\tilde{\mathbf{g}}_t\right)$, for any fixed $\mathbf{u}\in\mathcal{K}$,

$$\sum_{t=1}^{T}\left(\tilde{\mathbf{f}}_t-\alpha\tilde{\mathbf{g}}_t\right)^{\mathsf{T}}(\mathbf{x}_t-\mathbf{u})\le\eta^{-1}\mathcal{R}(\mathbf{u})+c_1\eta\sum_{t=1}^{T}\left\|\tilde{\mathbf{f}}_t-\alpha\tilde{\mathbf{g}}_t\right\|_t^{*2}.$$

Combining the results, with probability at least $1-(\delta+\delta')$,

$$\begin{aligned}&\sum_{t=1}^{T}\mathbf{f}_t^{\mathsf{T}}(\mathbf{y}_t-\mathbf{u})\le\eta^{-1}\mathcal{R}(\mathbf{u})+c_1\eta\sum_{t=1}^{T}\left\|\tilde{\mathbf{f}}_t-\alpha\tilde{\mathbf{g}}_t\right\|_t^{*2}+c_2\sqrt{nT}\\&+\sqrt{8T\log(1/\delta)}+(2c_3+c_4+2)\sqrt{nT}\log(2\log T/\delta').\end{aligned}$$

Finally, by Lemma 6.4 and our assumption (F), with a bit of algebra we arrive at the statement of Theorem 4.1. ∎

References

[1] J. Abernethy, E. Hazan, and A. Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Proceedings of The Twenty First Annual Conference on Learning Theory*, 2008.

[2] J. Abernethy and A. Rakhlin. Beating the adaptive bandit with high probability. Technical Report UCB/EECS-2009-10, EECS Department, University of California, Berkeley, Jan 2009.

[3] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, 32(1):48–77, 2003.

[4] Baruch Awerbuch and Robert D. Kleinberg. Adaptive routing with end-to-end feedback: distributed learning and geometric approaches. In *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 45–53, New York, NY, USA, 2004. ACM.

[5] P. L. Bartlett, V. Dani, T. Hayes, S. Kakade, A. Rakhlin, and A. Tewari. High probability regret bounds for online optimization. In *Proceedings of The Twenty First Annual Conference on Learning Theory*, 2008.

[6] Amir Beck and Marc Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Oper. Res. Lett.*, 31(3):167–175, 2003.

[7] Y. Censor and S. A. Zenios. *Parallel Optimization: Theory, Algorithms, and Applications*. Oxford University Press, 1997.

[8] Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[9] Varsha Dani, Thomas Hayes, and Sham Kakade. The price of bandit information for online optimization. In J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*. MIT Press, Cambridge, MA, 2008.

[10] Varsha Dani and Thomas P. Hayes. Robbing the bandit: less regret in online geometric optimization against an adaptive adversary. In *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 937–943, New York, NY, USA, 2006. ACM.

[11] Abraham D. Flaxman, Adam Tauman Kalai, and H. Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 385–394, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.

[12] A. György, T. Linder, G. Lugosi, and G. Ottucsák. The on-line shortest path problem under partial monitoring. *Journal of Machine Learning Research*, 8:2369–2403, 2007.

[13] H. Brendan McMahan and Avrim Blum. Online geometric optimization in the bandit setting against an adaptive adversary. In *COLT*, pages 109–123, 2004.

[14] A. Nemirovski and M. Todd. Interior-point methods for optimization. *Acta Numerica*, pages 191—234, 2008.

[15] Y. E. Nesterov and A. S. Nemirovskii. *Interior Point Polynomial Algorithms in Convex Programming*. SIAM, Philadelphia, 1994.

[16] A. Rakhlin and A. Tewari. Lecture notes on online learning, 2008. Available at http://www-stat.wharton.upenn.edu/~rakhlin/papers/online_learning.pdf.

[17] Shai Shalev-Shwartz. *Online Learning: Theory, Algorithms, and Applications*. PhD thesis, Hebrew University, 2007.