JMLR: Workshop and Conference Proceedings vol 49:1–7, 2016

Multi-scale exploration of convex functions and bandit convex optimization

Sébastien Bubeck SEBUBECK@MICROSOFT.COM
Microsoft Research

Ronen Eldan RONENELDAN@GMAIL.COM
Weizmann Institute

Abstract

We construct a new map from a convex function to a distribution on its domain, with the property that this distribution is a multi-scale exploration of the function. We use this map to solve a decade-old open problem in adversarial bandit convex optimization by showing that the minimax regret for this problem is $\widetilde{O}(\mathrm{poly}(n)\sqrt{T})$, where n is the dimension and T the number of rounds. This bound is obtained by studying the dual Bayesian maximin regret via the information ratio analysis of Russo and Van Roy, and then using the multi-scale exploration to construct a new algorithm for the Bayesian convex bandit problem.[1]

1. Introduction

Let $\mathcal{K} \subset \mathbb{R}^n$ be a convex body of diameter at most 1, and $f : \mathcal{K} \to [0, +\infty)$ a non-negative convex function. Suppose we want to test whether some unknown convex function $g : \mathcal{K} \to \mathbb{R}$ is equal to f, with the alternative being that g takes a negative value somewhere on $\mathcal{K}$. In statistical terminology the null hypothesis is

$$\mathrm{H}_0 : g = f,$$

and the alternative is

$$\mathrm{H}_1 : \exists\, \alpha \in \mathcal{K} \text{ such that } g(\alpha) < -\varepsilon,$$

where ε is some fixed positive number. In order to decide between the null hypothesis and the alternative one is allowed to make a single noisy measurement of g. That is one can choose a point $x \in \mathcal{K}$ (possibly at random) and obtain $g(x) + \xi$ where ξ is a zero-mean random variable independent of x (say $\xi \sim \mathcal{N}(0, 1)$). Is there a way to choose x such that the total variation distance between the observed measurement under the null and the alternative is at least (up to logarithmic terms) $\varepsilon/\mathrm{poly}(n)$? Observe that without the convexity assumption on g this distance is always $O(\varepsilon^{n+1})$, and thus a positive answer to this question would crucially rely on convexity. We show that $\varepsilon/\mathrm{poly}(n)$ is indeed attainable by constructing a distribution on $\mathcal{K}$ which guarantees an exploration of the convex function f at *every scale* simultaneously. Precisely we prove the following new result on convex functions. We denote by c a universal constant whose value can change at each occurence.

Theorem 1 *Let $\mathcal{K} \subset \mathbb{R}^n$ be a convex body of diameter at most 1. Let $f : \mathcal{K} \to [0, +\infty)$ be convex and 1-Lipschitz, and let $\varepsilon > 0$. There exists a probability measure μ on $\mathcal{K}$ such that the following*

1. Extended abstract. Full version appears as Bubeck and Eldan (2015).

© 2016 S. Bubeck & R. Eldan.

holds true. For every $\alpha \in \mathcal{K}$ *and for every convex and* 1*-Lipschitz function* $g : \mathcal{K} \to \mathbb{R}$ *satisfying* $g(\alpha) < -\varepsilon$*, one has*

$$\mu\left(\left\{x \in \mathcal{K} : |f(x) - g(x)| > \frac{c}{n^{7.5}\log(1+n/\varepsilon)}\max(\varepsilon, f(x))\right\}\right) > \frac{c}{n^3\log(1+n/\varepsilon)}.$$

Our main application of the above result is to resolve a long-standing gap in bandit convex optimization. We refer the reader to Bubeck and Cesa-Bianchi (2012) for an introduction to bandit problems (and some of their applications). The bandit convex optimization problem can be described as the following sequential game: at each time step $t = 1, \ldots, T$, a player selects an action $x_t \in \mathcal{K}$, and simultaneously an adversary selects a convex (and 1-Lipschitz) loss function $\ell_t : \mathcal{K} \mapsto [0, 1]$. The player's feedback is its suffered loss, $\ell_t(x_t)$. We assume that the adversary is oblivious, that is the sequence of loss functions $\ell_1, \ldots, \ell_T$ is chosen before the game starts. The player has access to external randomness, and can select her action x_t based on the history $H_t = (x_s, \ell_s(x_s))_{s<t}$. The player's perfomance at the end of the game is measured through the *regret*:

$$R_T = \sum_{t=1}^{T} \ell_t(x_t) - \min_{x\in\mathcal{K}} \sum_{t=1}^{T} \ell_t(x),$$

which compares her cumulative loss to the best cumulative loss she could have obtained in hindsight with a fixed action, if she had known the sequence of losses played by the adversary. A major open problem since Kleinberg (2004); Flaxman et al. (2005) is to reduce the gap between the $\sqrt{T}$-lower bound and the $T^{3/4}$-upper bound for the minimax regret of bandit convex optimization. In dimension one (i.e., $\mathcal{K} = [0, 1]$) this gap was closed recently in Bubeck et al. (2015) and our main contribution is to extend this result to higher dimensions:

Theorem 2 *There exists a player's strategy such that for any sequence of convex (and* 1*-Lipschitz) losses one has*

$$\mathbb{E}R_T \leq c\, n^{11} \log^4(T)\sqrt{T},$$

where the expectation is with respect to the player's internal randomization.

We observe that this result also improves the state of the art regret bound for the easier situation where the losses $\ell_1, \ldots, \ell_T$ form an i.i.d. sequence. In this situation the best previous bound was obtained by Agarwal et al. (2011) and is $\widetilde{O}(n^{16}\sqrt{T})$.

Using Theorem 1 we prove Theorem 2 in Section 2. Theorem 1 itself is proven in the full version Bubeck and Eldan (2015). In this extended abstract we only provide the proof of Theorem 1 in dimension 1, see Section 3.

2. Proof of Theorem 2

Following Bubeck et al. (2015) we reduce the proof of Theorem 2 to upper bounding the *Bayesian maximin regret* (this reduction is simply an application of Sion's minimax theorem). In other words the sequence $(\ell_1, \ldots, \ell_T)$ is now a random variable with a distribution known to the player. Expectations are now understood with respect to both the latter distribution, and possibly the randomness in the player's strategy. We denote $\mathbb{E}_t$ for the expectation conditionally on the random variable H_t. As in Bubeck et al. (2015) we analyze the Bayesian maximin regret with the information theoretic

approach of Russo and Van Roy (2014a), which we recall in subsection 2.1. A key contribution of our work is then to propose in subsection 2.2 a new strategy for the Bayesian convex bandit problem, which can be viewed as an ε-greedy strategy, where the value of ε is derived from the form of the posterior, and the exploration strategy is derived from Theorem 1.

2.1. The information ratio

Let $\bar{\mathcal{K}} = \{\bar{x}_1, \ldots, \bar{x}_K\}$ be a $1/\sqrt{T}$-net of $\mathcal{K}$. Note that $K \leq (4T)^n$. We define a random variable $\bar{x}^* \in \bar{\mathcal{K}}$ such that $\sum_{t=1}^T \ell_t(\bar{x}^*) = \min_{x \in \bar{\mathcal{K}}} \sum_{t=1}^T \ell_t(x)$. Using that the losses are Lipschitz one has

$$R_T \leq \sqrt{T} + \sum_{t=1}^T (\ell_t(x_t) - \ell_t(\bar{x}^*)). \tag{1}$$

We introduce the following key quantities, for $x \in \mathcal{K}$,

$$r_t(x) = \mathbb{E}_t(\ell_t(x) - \ell_t(\bar{x}^*)), \quad \text{and} \quad v_t(x) = \mathrm{Var}_t(\mathbb{E}_t(\ell_t(x)|\bar{x}^*)). \tag{2}$$

In words, conditionally on the history, $r_t(x)$ is the (approximate) expected regret of playing x at time t, and $v_t(x)$ is a proxy for the information about $\bar{x}^*$ revealed by playing x at time t. It will be convenient to rewrite these functions slightly more explicitly. Let $i^* \in [K]$ be the random variable such that $\bar{x}^* = \bar{x}_{i^*}$. We denote by α^* its distribution, which we view as a point in the $K-1$ dimensional simplex. Let $\alpha_t = \mathbb{E}_t \alpha^*$. In words $\alpha_t = (\alpha_{1,t}, \ldots, \alpha_{K,t})$ is the posterior distribution of x^* at time t. Let $f_{i,t}, f_t : \mathcal{K} \to [0,1], i \in [K], t \in [T]$, be defined by, for $x \in \mathcal{K}$,

$$f_t(x) = \mathbb{E}_t \ell_t(x), \quad f_{i,t}(x) = \mathbb{E}_t(\ell_t(x)|\bar{x}^* = \bar{x}_i).$$

Then one can easily see that

$$r_t(x) = f_t(x) - \sum_{i=1}^K \alpha_{i,t} f_{i,t}(\bar{x}_i), \quad \text{and} \quad v_t(x) = \sum_{i=1}^K \alpha_{i,t}(f_t(x) - f_{i,t}(x))^2. \tag{3}$$

The main observation in Russo and Van Roy (2014a) is the following lemma, which gives a bound on the accumulation of information (see also [Appendix B, Bubeck et al. (2015)] for a short proof).

Lemma 3 *One always has* $\mathbb{E} \sum_{t=1}^T v_t(x_t) \leq \frac{1}{2} \log(K)$.

An important consequence of Lemma 3 is the following result which follows from an application of Cauchy-Schwarz (and (1)):

$$\mathbb{E} \sum_{t=1}^T r_t(x_t) \leq \sqrt{T} + C \sum_{t=1}^T \sqrt{\mathbb{E} v_t(x_t)} \;\Rightarrow\; \mathbb{E} R_T \leq 2\sqrt{T} + C\sqrt{\frac{T}{2} \log(K)}. \tag{4}$$

In particular a strategy which obtains at each time step an information proportional to its instantaneous regret has a controlled cumulative regret:

$$\mathbb{E}_t r_t(x_t) \leq \frac{1}{\sqrt{T}} + C\sqrt{\mathbb{E}_t v_t(x_t)}, \; \forall t \in [T] \;\Rightarrow\; \mathbb{E} R_T \leq 2\sqrt{T} + C\sqrt{\frac{T}{2} \log(K)}. \tag{5}$$

Russo and Van Roy (2014a) refers to the quantity $\mathbb{E}_t r_t(x_t)/\sqrt{\mathbb{E}_t v_t(x_t)}$ as the *information ratio*. They show that Thompson Sampling (which plays x_t at random, drawn from the distribution α_t) satisfies $\mathbb{E}_t r_t(x_t)/\sqrt{\mathbb{E}_t v_t(x_t)} \leq K$ (without any assumptions on the loss functions $\ell_t : \mathcal{K} \to [0,1]$). In Bubeck et al. (2015) it is shown that in dimension one (i.e., $n = 1$), the latter bound can be improved using the convexity of the losses by replacing K with a polylogarithmic term in K (Thompson Sampling is also slightly modified). In the present paper we propose a completely different strategy, which is loosely related to the Information Directed Sampling of Russo and Van Roy (2014b). We describe and analyze our new strategy in the next subsection.

2.2. A two-point strategy

We describe here a new strategy to select x_t, conditionally on H_t, and show that it satisfies a bound of the form given in (5). To lighten notation we drop all time subscripts, e.g. one has $r(x) = f(x) - \sum_{i=1}^K \alpha_i f_i(\bar{x}_i)$, and $v(x) = \sum_{i=1}^K \alpha_i \left(f_i(x) - f(x)\right)^2$. Our objective is to describe a random variable $X \in \mathcal{K}$ which satisfies

$$\mathbb{E} r(X) \leq \frac{1}{\sqrt{T}} + C\sqrt{\mathbb{E} v(X)}, \tag{6}$$

where C is polylogarithmic in K (recall that $K \leq (4T)^n$). We now describe the construction of our proposed random variable X (or to put it differently we describe a new algorithm for the Bayesian convex bandit problem), and we prove that it satisfies (6).

Let $x^* \in \operatorname{argmin}_{x \in \mathcal{K}} f(x)$. We translate the functions so that $f(x^*) = 0$ and denote $L = \sum_{i=1}^K \alpha_i f_i(\bar{x}_i)$. If $L \geq -1/\sqrt{T}$ then $X := x^*$ satisfies (6), and thus in the following we assume that $L \leq -1/\sqrt{T}$.

Step 1: We claim that there exists $\varepsilon \in [|L|/2, 1]$ such that

$$\alpha\left(\{i \in [K] : f_i(\bar{x}_i) \leq -\varepsilon\}\right) \geq \frac{|L|}{2\log(2/|L|)\varepsilon}. \tag{7}$$

Indeed assume that (7) is false for all $\varepsilon \in [|L|/2, 1]$, and let Y be a random variable such that $\mathbb{P}(Y = -f_i(\bar{x}_i)) = \alpha_i$, then

$$|L| = \mathbb{E} Y \leq |L|/2 + \int_{|L|/2}^1 \mathbb{P}(Y \geq x) dx < |L|/2 + \int_{|L|/2}^1 \frac{|L|}{2\log(2/|L|)x} dx = |L|,$$

thus leading to a contradiction. We denote $I = \{i \in [K] : f_i(\bar{x}_i) \leq -\varepsilon\}$ with ε satisfying (7).

Step 2: We show here the existence of a point $\bar{x} \in \mathcal{K}$ and a set $J \subset I$ such that $\alpha(J) \geq \frac{c}{n^3 \log(1+n/\varepsilon)} \alpha(I)$ and for any $i \in J$,

$$|f(\bar{x}) - f_i(\bar{x})| \geq \frac{c}{n^{7.5}\log(1+n/\varepsilon)} \max(\varepsilon, f(\bar{x})). \tag{8}$$

We say that a point is *good* for f_i if it satisfies (8), and thus we want to prove the existence of a point $\bar{x}$ which is good for a large fraction (with respect to the posterior) of the f_i's. Denote

$$A_i = \left\{x \in \mathcal{K} : |f(x) - f_i(x)| \geq \frac{c}{n^{7.5}\log(1+n/\varepsilon)} \max(\varepsilon, f(x))\right\},$$

and let μ be the distribution given by Theorem 1. Then one obtains:

$$\sup_{x\in\mathcal{K}}\sum_{i\in I}\alpha_i\mathbb{1}\{x\in A_i\}\geq\int_{x\in\mathcal{K}}\sum_{i\in I}\alpha_i\mathbb{1}\{x\in A_i\}d\mu(x)=\sum_{i\in I}\alpha_i\mu(A_i)\geq\frac{c}{n^3\log(1+n/\varepsilon)}\alpha(I),$$

which clearly implies the existence of J and $\bar{x}$.

Step 3: Let X be such that $\mathbb{P}(X=\bar{x})=\alpha(J)$ and $\mathbb{P}(X=x^*)=1-\alpha(J)$. Then

$$\mathbb{E}r(X)=|L|+\alpha(J)f(\bar{x}),$$

and using the definition of $\bar{x}$ one easily see that:

$$\sqrt{\mathbb{E}v(X)}\geq\sqrt{\alpha(J)v(\bar{x})}\geq\sqrt{\alpha(J)\sum_{i\in J}\alpha_i(f_i(\bar{x})-f(\bar{x}))^2}\geq\frac{c}{n^{7.5}\log(1+n/\varepsilon)}\alpha(J)\max(\varepsilon,f(\bar{x})).$$

Finally, since $\alpha(J)\geq\frac{c|L|}{\varepsilon n^3\log^2(1+n/\varepsilon)}$, the two above displays clearly implies (6).

3. An exploratory distribution for convex functions

In this section we construct an exploratory distribution μ of a convex function f which satisfies the conditions of Theorem 1. Our construction proceeds by induction on the dimension, and in this extended abstract we only provide the proof ot the base case (see the full version Bubeck and Eldan (2015) for the complete proof). The base case is much simpler than the proof for a general dimension, but already contains some of the central ideas used in the general case. In particular, a (much simpler) multi-scale argument is used.

The main ingredient is the following lemma which is easy to verify by picture (we provide a formal proof for sake of completness).

Lemma 4 *Let $f,g:\mathbb{R}\to\mathbb{R}$ be two convex functions. Suppose that $f(x)\geq 0$. Let $x_0,\alpha\in\mathbb{R}$ be two points satisfying $\alpha-1<x_0<\alpha$, and suppose that $g(\alpha)<-\varepsilon$ for some $\varepsilon>0$ and that*

$$f'(x)\geq 0,\ \forall x>x_0. \tag{9}$$

Let μ be a probability measure supported on $[x_0,\alpha]$ whose density with respect to the Lebesgue measure is bounded from above by some $\beta>1$. Then we have

$$\mu\left(\left\{x:|f(x)-g(x)|>\tfrac{1}{4}\beta^{-1}\max(\varepsilon,f(x))\right\}\right)\geq\frac{1}{2}.$$

Proof We first argue that, without loss of generality, one may assume that f attains its minimum at x_0. Indeed, we may clearly change f as we please on the interval $(-\infty,x_0)$ without affecting the assumptions or the result of the Lemma. Using the condition (9) we may therefore make this assumption legitimate.

Assume, for now, that there exists $x_1\in[x_0,\alpha]$ for which $f(x_1)=g(x_1)$. By convexity, and since $f(x_0)\geq 0$ and $g(\alpha)<0$, if such point exists then it is unique. Let $h(x)$ be the linear function passing through $(\alpha,g(\alpha))$ and $(x_1,f(x_1))$. By convexity of g, we have that $|g(x)-f(x)|\geq|h(x)-f(x)|$ for all $x\in[x_0,\alpha]$. Now, since $h(\alpha)<-\varepsilon$ and since $\alpha<x_1+1$, we have

$h'(x_0) < -(\varepsilon + f(x_0))$. Moreover, since we know that $f(x)$ is non-decreasing in $[x_0, \alpha]$, we conclude that

$$\begin{aligned} |g(x) - f(x)| &\geq |h(x) - f(x)| \\ &= |h(x) - f(x_1)| + |f(x) - f(x_1)| \\ &= (\varepsilon + f(x_1))|x - x_1| + |f(x) - f(x_1)| \\ &\geq \max(\varepsilon, f(x))|x - x_1|, \quad \forall x \in [x_0, \alpha]. \end{aligned}$$

It follows that

$$\left\{x;\ |f(x) - g(x)| < \tfrac{1}{4}\beta^{-1}\max(\varepsilon, f(x))\right\} \subset I := \left[x_1 - \tfrac{1}{4}\beta^{-1}, x_1 + \tfrac{1}{4}\beta^{-1}\right]$$

but since the density of μ is bounded by β, we have $\mu(I) \leq \frac{1}{2}$ and we're done.

It remains to consider the case that $g(x) < f(x)$ for all $x \in [x_0, \alpha]$. In this case, we may define

$$\widetilde{g}(x) = g(x) + \frac{f(x_0) - g(x_0)}{\alpha - x_0}(\alpha - x).$$

Note that $\widetilde{g}(x) \geq g(x)$ for all $x \in [x_0, \alpha]$, which implies that $|g(x) - f(x)| \geq |\widetilde{g}(x) - f(x)|$ for all $x \in [x_0, \alpha]$. Since $\widetilde{g}(x_0) = f(x_0)$, we may continue the proof as above, replacing the function g by $\widetilde{g}$. ■

We are now ready to prove the one dimensional case. The proof essentially invokes the above lemma on every scale between ε and 1.

Proof [Proof of Theorem 1, the case $n = 1$] Let $x_0 \in \mathcal{K}$ be the point where the function f attains its minimum and set $d = \mathrm{diam}(\mathcal{K})$. Define $N = \lceil \log_2 \frac{1}{\varepsilon} \rceil + 4$. For all $0 \leq k \leq N$, consider the interval

$$I_k = [x_0 - d2^{-k}, x_0 + d2^{-k}] \cap \mathcal{K}$$

and define the measure μ_k to be the uniform measure over the interval I_k. Finally, we set

$$\mu = \frac{1}{N+2}\sum_{k=0}^{N} \mu_k + \frac{1}{N+2}\delta_{x_0}.$$

Now, let $\alpha \in \mathcal{K}$ and let $g(x)$ be a convex function satisfying $g(\alpha) \leq -\varepsilon$. We would like to argue that $\mu(A) \geq \frac{1}{8\log(1+1/\varepsilon)}$ for $A = \left\{x \in \mathcal{K} :\ |f(x) - g(x)| \geq \frac{1}{8}\max(\varepsilon, f(x))\right\}$.

Set $k = \lceil \log_{1/2}(|\alpha - x_0|/d) \rceil$. Define $Q(x) = x_0 + d2^{-k}(x - x_0)$ and set $\widetilde{f}(x) = f(Q(x))$, $\widetilde{g}(x) = g(Q(x))$, $\widetilde{\alpha} = Q^{-1}(\alpha)$ and consider the interval

$$I = Q^{-1}(I_k) \cap \{x : (x - x_0)(\alpha - x_0) \geq 0\}$$

It is easy to check that, by definition I is an interval of length 1, contained in the interval $[x_0, \widetilde{\alpha}]$. Defining $\widetilde{\mu} = \mu_I$, we have that the density of $\widetilde{\mu}$ with respect to the Lebesgue measure is equal to 1. An application of Lemma 4 for the functions $\widetilde{f}, \widetilde{g}$, the points $x_0, \widetilde{\alpha}$ and the measure $\widetilde{\mu}$ teaches us that

$$\begin{aligned} \mu_k(A) &= \mu_{Q^{-1}(I_k)}\left(\left\{x :\ \left|\widetilde{f}(x) - \widetilde{g}(x)\right| \geq \frac{1}{8}\max(\varepsilon, \widetilde{f}(x))\right\}\right) \\ &\geq \frac{1}{2}\widetilde{\mu}\left(\left\{x :\ \left|\widetilde{f}(x) - \widetilde{g}(x)\right| \geq \frac{1}{8}\max(\varepsilon, \widetilde{f}(x))\right\}\right) \geq \frac{1}{4}. \end{aligned}$$

By definition of the measure μ, we have that whenever $k \leq N$, one has

$$\mu\left(A\right) \geq \frac{1}{N+2} \geq \frac{1}{8\log(1+1/\varepsilon)}.$$

Finally, if $k > N$, it means that $|\alpha - x_0| < 2^{-N} < \frac{\varepsilon}{4}$. Since the function g is 1-Lipschitz, this implies that $g(x_0) \leq -\varepsilon/2$ which in turn gives $|f(x_0) - g(x_0)| \geq \frac{1}{8}\max(\varepsilon, f(x_0))$. Consequently, $x_0 \in A$ and thus $\mu(A) \geq \mu(\{x_0\}) = \frac{1}{N+2} \geq \frac{1}{8\log(1+1/\varepsilon)}$. The proof is complete. ■

References

A. Agarwal, D.P. Foster, D. Hsu, S.M. Kakade, and A. Rakhlin. Stochastic convex optimization with bandit feedback. In *Advances in Neural Information Processing Systems (NIPS)*, 2011.

S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

S. Bubeck and R. Eldan. Multi-scale exploration of convex functions and bandit convex optimization. *Arxiv preprint arXiv:1507.06580*, 2015.

S. Bubeck, O. Dekel, T. Koren, and Y. Peres. Bandit convex optimization: $\sqrt{T}$ regret in one dimension. In *Proceedings of the 28th Annual Conference on Learning Theory (COLT)*, 2015.

A. Flaxman, A. Kalai, and B. McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *In Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2005.

R. Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. *Advances in Neural Information Processing Systems (NIPS)*, 2004.

D. Russo and B. Van Roy. An information-theoretic analysis of thompson sampling. *arXiv preprint arXiv:1403.5341*, 2014a.

D. Russo and B. Van Roy. Learning to optimize via information directed sampling. *arXiv preprint arXiv:1403.5556*, 2014b.