Math. Stat. Learn. 1 (2018), 73–100
DOI 10.4171/MSL/1-1-3

Mathematical Statistics and Learning

Exploratory distributions for convex functions

Sébastien Bubeck and Ronen Eldan

Abstract. Given a Lipschitz and convex function f on a compact and convex domain in $\mathbb{R}^n$, we construct an *exploratory distribution* μ of f in the following sense. Let g be a Lipschitz and convex function on the same domain, such that either $g = f$, or alternatively the minimum of g is ε smaller than the minimum of f. Then μ is such that poly(n/ε) noisy evaluations of g at i.i.d. points from μ suffices to determine with high probability whether $g = f$ or $g \neq f$. As an example of application for such exploratory distributions we show how to use them to estimate the minimax regret for adversarial bandit convex optimization.

Mathematics Subject Classification (2010). 52A20; 62C05.

Keywords. Learning convex functions, hypothesis testing, exploratory distribution, convex bandit optimization.

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. This paper is concerned with the following question: Can we find a measure μ associated with the function f, such that for every convex function g which is substantially different from f, one has that $\mu(\{f \neq g\})$ is rather large?

Put more precisely, we are interested in finding a measure μ with the following property: For any function g which takes a negative value $-\varepsilon$ at some point, one has $\mu(\{|f - g| > \eta\varepsilon\}) > \delta$ for constants η, δ as large as possible. The constants η, δ are expected to have two features: First, we would like to avoid the curse of dimensionality in the sense that both constants depend polynomially on the dimension n. Second, we would like the dependence on ε to be logarithmic.

We denote by c a universal constant whose value can change at each occurrence. Our main theorem reads:

Theorem 1.1. *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 holds true. For every $\alpha \in \mathcal{K}$ and for every convex and* 1*-Lipschitz function $g\colon \mathcal{K} \to \mathbb{R}$ satisfying $g(\alpha) < -\varepsilon$, one has*

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

Remark 1.2. Observe that without the convexity assumption on g, one can only ensure that $\mu(\{f \neq g\})$ is of the order $\Omega(\varepsilon^{n+1})$. To see this, take $\mathcal{K} = [0,1]^n$, $f \equiv 0$, and for $\alpha \in \mathcal{K}$ define $g_\alpha = \min(0, -\varepsilon + |x-\alpha|)$. In particular f and g_α are distinct only on the ball $\mathrm{B}(\alpha,\varepsilon)$ centered at α and of radius ε. Clearly one has, for any probability measure μ, that $\inf_{\alpha\in\mathcal{K}} \mu(\mathrm{B}(\alpha,\varepsilon)) \lesssim \varepsilon^n$ (since there exist a set of $\alpha \in \mathcal{K}$ of cardinality $\Omega(\varepsilon^n)$ such that such that $\mathrm{B}(\alpha,\varepsilon)$ are mutually disjoint). This implies that for some α the probability under μ to see a point where f and g_α are distinct is $O(\varepsilon^n)$.

As an example of application for the above result we resolve a long-standing gap in bandit convex optimization[1]. We refer the reader to [4] 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 $t = 1, \dots, T$, a player selects an action $x_t \in \mathcal{K}$, and simultaneously an adversary selects a convex (and 1-Lipschitz) loss function $\ell_t\colon \mathcal{K} \mapsto [0,1]$. 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 $\ell_1, \dots, \ell_T$ 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 [8, 11] 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 [5] and our main contribution for this problem is to extend this result to higher dimensions:

Theorem 1.3. *There exists a player's strategy such that for any sequence of convex (and* 1*-Lipschitz) losses one has*

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

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

[1]Since the publication of the conference version of this paper [6], new strategies have been found for bandit convex optimization with improved running time [9] and also improved regret bound [7].

We observe that this result also improves the state of the art regret bound for the easier situation where the losses $\ell_1, \dots, \ell_T$ form an i.i.d. sequence. In this situation the best previous bound was obtained by [2] and is $O(n^{16}\sqrt{T})$, up to logarithmic terms.

Using Theorem 1.1 we prove Theorem 1.3 in Section 4. Theorem 1.1 itself is proven in Section 3.

2. Intuition for Theorem 1.1

2.1. Some examples. Before describing the central ideas of our construction, we begin with a few examples of functions which suggest where some of the difficulties lie, and give an intuition as to how to overcome those difficulties.

In each of these following examples, we will define a function f over a domain $\Omega \subset \mathbb{R}^n$, and for every $\alpha \in \Omega$, the function g_α will be defined as the convex envelope of the function

$$\tilde{g}(x) = \begin{cases} -\varepsilon, & x = \alpha, \\ f(x), & \text{otherwise.} \end{cases}$$

For all $\alpha \in \Omega$, we consider the set

$$S_\alpha = \big\{x \in \Omega : |g_\alpha(x) - f(x)| \geq c\varepsilon\big\}.$$

The exploratory property of the measure μ implies that $\mu(S_\alpha) \geq c$ for all $\alpha \in \Omega$, where the constant c depends inverse-polynomially on n and on $|\log \varepsilon|$.

Our first example is in two dimensions. Consider the function $f\colon [-1,1]^2 \to \mathbb{R}$ defined by $f(x, y) = x^2$. It is not hard to check that for any $\alpha \in \{0\} \times [-1,1]$, we have $S_\alpha \subset D$ for the strip $D = [-2\sqrt{\varepsilon}, 2\sqrt{\varepsilon}\,] \times [0,1]$. Consequently, we get that the $\mu(D) > c$ even though the Lebesgue measure of D is polynomially small with respect to ε.

Next, consider the function $f(x) = |x|^2$ defined on $\Omega = \{x \in \mathbb{R}^n : |x| \leq 1\}$. It is not hard to show that:

$$g(x) = f(x) \text{ iff } |x - \alpha|^2 \geq \varepsilon + |\alpha|^2\,.$$

Another simple calculation then shows that when $|\alpha| = 1$, the volume of S_α is exponentially small as a proportion of Ω. This example shows that in order to attain a polynomial bound in the dimension, some of the "exploration" should occur near the minimum of f. Indeed further scrutiny would reveal that at distance $O(1/\sqrt{n})$ from the origin, the set S_α is effectively a half-space, thus in this case by taking μ to be uniform on a ball of radius $1/\sqrt{n}$ we would be able to differentiate between f and g_α with constant probability.

Finally, consider the function

$$f(x_1, \dots, x_n) = \sum_{i=1}^{n} \varepsilon^{i/n} x_i^2$$

The measure μ assigned to this function should be able to associate a different scale with every coordinate direction. It is clear that by considering a linear transformation this example becomes similar to $f(x) = |x|^2$, however in the general case such a linear transformation does not exist.

2.2. Line of sight argument. One of the central ideas of our proof is based on a "line of sight" argument. First, let us illustrate this idea in one dimension. Suppose that the functions $f, g\colon [0,1] \to \mathbb{R}$ are convex, and fix a point $x_0 \in [0,1]$ and an interval $I \subset [0,1]$. Suppose that $g(x_0) < f(x_0)$. Observe the following fact: if for any $x \in I$ the open segment connecting the points $(x_0, g(x_0))$ and $(x, f(x))$ does not intersect the graph of f, then there must exist a point $x' \in I$ such that $g(x) \neq f(x)$. In this case, we say that the point $(x_0, g(x_0))$ has an open line of sight to the graph of f on the interval I. This can be made quantitative in the following sense: if $f'(x)(x_0 - x) > 0$ for all $x \in I$ and $g(x_0) < -\varepsilon$, then there must exist a point x' such that $|f(x') - g(x')| \geq \varepsilon |I|/2$.

The multi-dimensional version of this argument is slightly more involved: the interval I is replaced by some ball B, and the condition $f'(x)(x_0 - x) > 0$ needs to be replaced by a sort of smoothness condition: one needs that ∇f be contained in a small ball. This inspires the definition of a jolly-good triplet (defined in the beginning of Section 3.4). Roughly speaking, $(z, \theta, t) \in \Omega \times \mathbb{S}^{n-1} \times \mathbb{R}_+$ is a *jolly-good triplet* if the gradient of f is almost constant around the point z, in the sense that for most of the points x in a small ball (with radius $\Omega(1/\text{poly}(n))$ around the point z, one has that $\nabla f(x) \in \mathrm{B}(t\theta, t/\text{poly}(n))$. Given a triplet as above, it is implied that any point in the set $A = \{x : \langle x, \theta \rangle > C\}$ (for some not too small constant $C > 0$) has a line-of sight to most of the points in a ball around the point z, therefore, any function g such that $g(\alpha) < -\varepsilon$ for some $\alpha \in A$ must be quite different from the function f in a large proportion of the points of B.

Thus, provided that one has found a jolly-good triplet (z, θ, t), by asserting that the density of μ in the ball $B(z, \delta)$ is bounded from below, one can make sure that the points of the set A (defined above) will be "explored" by μ. The line-of-sight argument takes place in Section 3.5 and in particular in Lemma 3.6.

2.3. Existence of smooth regions in arbitrary convex functions. The key to apply the line of sight argument in high dimension is to find a jolly-good triplet (in fact many as we will see shortly). That is we want to find a ball where f is essentially smooth (in the sense that the gradient is Lipschitz at most points in that ball). Intuitively this corresponds to a quantitative version of Alexandrov theorem, which states that a convex function is smooth (i.e. it has a second derivative) almost everywhere.

To prove the existence of a jolly-good triplet we use the following contraction argument, see Lemma 3.5 for the details. Fix a ball B of radius δ, then by Poincaré's inequality one can upper bound the "standard deviation" of the gradient of f on that ball by δ times the average value of the Laplacian of f on B. Thus it only remains to control this latter value, which is also equal to the value of the Laplacian of g at the center of B, where g is equal to f convolved with the (normalized) indicator of a ball of radius δ. In other words it only remains to exhibit a single point where the Laplacian of g is small. By Gauss theorem the average value of the Laplacian is controlled by the maximal gradient norm (see Lemma 3.9), and since g is smooth we know that we can find some ball where all the gradients of g are small (namely around the minimum of g). Thus the average of the Laplacian of g on that ball is small, and in particular this means that there exists a point where the Laplacian of g is small.

2.4. The exploratory iteration scheme. Using a jolly-good triplet together with the line of sight argument one can "explore" f efficiently against any g whose minimizer is in some halfspace at a small distance from the origin, see Lemma 3.3. Crucially we show next that in fact one can find many jolly-good triplets, see Section 3.4, so that one can efficiently explore against all points, except possibly in a small strip. This is summarized in Lemma 3.2. Finally we show how to deal with this possibly bad strip in Section 3.3. The idea is simply to repeatedly apply the previous argument in this strip, until one reaches a strip so thin that we have essentially reduced the dimension of the problem by 1, at which point one concludes by induction.

3. An exploratory distribution for convex functions

In this section we will describe the construction of the exploratory distribution μ, and prove Theorem 1.1. We fix a convex function f which satisfies the conditions of the theorem throughout the section.

3.1. The one-dimensional case. Since our proof of Theorem 1.1 will proceed by induction, our first goal is to establish the result in dimension 1. This task will be 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 the sake of completness).

Lemma 3.1. *Let $f, g\colon \mathbb{R} \to \mathbb{R}$ 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*

$$f'(x) \geq 0, \quad \forall x > x_0. \tag{3.1}$$

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

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

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 (3.1) 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

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

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

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

Note that $\tilde{g}(x) \geq g(x)$ for all $x \in [x_0, \alpha]$, which implies that

$$|g(x) - f(x)| \geq |\tilde{g}(x) - f(x)|$$

for all $x \in [x_0, \alpha]$. Since $\tilde{g}(x_0) = f(x_0)$, we may continue the proof as above, replacing the function g by $\tilde{g}$. □

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 of Theorem 1.1. The case $n=1$. Let $x_0 \in \mathcal{K}$ be the point where the function f attains its minimum and set $d = \operatorname{diam}(\mathcal{K})$. Define $N = \lceil \log_2 {}^1/_\varepsilon \rceil + 4$. For all $0 \le k \le N$, consider the interval

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

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

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

where δ_{x_0} is a Dirac measure supported on x_0. Now, let $\alpha \in \mathcal{K}$ and let $g(x)$ be a convex function satisfying $g(\alpha) \le -\varepsilon$. We would like to argue that $\mu(A) \ge 1/(8\log(1 + {}^1/_\varepsilon))$ for

$$A = \left\{x \in \mathcal{K} : |f(x) - g(x)| \ge \tfrac{1}{8}\max(\varepsilon, f(x))\right\}.$$

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

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

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 3.1 for the functions $\widetilde{f}, \widetilde{g}$, the points $x_0, \widetilde{\alpha}$ and the measure $\widetilde{\mu}$ teaches us that

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

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

$$\mu(A) \ge \frac{1}{N+2} \ge \frac{1}{8\log(1 + {}^1/_\varepsilon)}.$$

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 $g(x_0) \le -\frac{\varepsilon}{2}$ which in turn gives

$$|f(x_0) - g(x_0)| \ge \tfrac{1}{8}\max(\varepsilon, f(x_0)).$$

Consequently, $x_0 \in A$ and thus

$$\mu(A) \ge \mu(\{x_0\}) = \frac{1}{N+2} \ge \frac{1}{8\log(1 + {}^1/_\varepsilon)}.$$

The proof is complete. □

3.2. The high-dimensional case. We now consider the case where $n \geq 2$. For a measurable, bounded set $\Omega \subset \mathbb{R}^n$ and a direction $\theta \in \mathbb{R}^n$ we denote

$$S_{\Omega,\theta} = \left\{x \in \Omega : |\langle x, \theta \rangle| \leq \tfrac{1}{4}\right\},$$

and μ_Ω for the uniform measure on Ω. For a distribution μ we write $\mathrm{Cov}(\mu) = \mathbb{E}_{X\sim\mu} XX^\top$. Moreover we denote by $\mathrm{B}(x, r)$ the Euclidean ball of radius r centered at x.

As we explain in Section 3.3 our construction iteratively applies the following lemma:

Lemma 3.2. *Let $\varepsilon > 0$, $L \in [1, 2n]$. Let $\Omega \subset \mathbb{R}^n$ be a convex set with $0 \in \Omega$ and $\mathrm{Cov}(\mu_\Omega) = \mathrm{Id}$. Let $f : \Omega \to [0, \infty]$ be a convex and L-Lipschitz function with $f(0) = 0$. Then there exists a probability measure μ on Ω and a direction $\theta \in \mathbb{S}^{n-1}$ such that for all $\alpha \in \Omega \setminus S_{\Omega,\theta}$ and for every convex function $g : \Omega \to \mathbb{R}$ satisfying $g(\alpha) < -\varepsilon$, one has*

$$\mu\left(\left\{x \in \Omega : |f(x) - g(x)| > \frac{1}{2^{50} n^{7.5} \log(1 + n/\varepsilon)} \max(\varepsilon, f(x))\right\}\right) > \frac{1}{16n}. \tag{3.2}$$

The above lemma is proven in Section 3.4. A central ingredient in its proof is, in turn, the following Lemma, which itself is proven in Section 3.5.

Lemma 3.3. *Let $\varepsilon > 0$, $\Omega \subset \mathbb{R}^n$ a convex set with $\mathrm{diam}(\Omega) \leq M$, and $f : \Omega \to \mathbb{R}_+$ a differentiable convex function. Assume that there exist $\delta \in (0, 1/32n^2)$, $z \in \Omega \cap \mathrm{B}(0, 1/16)$, $\theta \in \mathbb{S}^{n-1}$ and $t > 0$ such that*

$$\mu_{\mathrm{B}(z,\delta)}\left((\nabla f)^{-1}\left(\mathrm{B}\left(t\theta, \frac{t}{16n^2}\right)\right)\right) \geq \frac{1}{2}. \tag{3.3}$$

Then for all $\alpha \in \Omega$ satisfying $\langle \alpha, \theta \rangle \geq \frac{1}{8}$ and $|\alpha| \leq 2n$ and for all convex function $g : \Omega \to \mathbb{R}$ satisfying $g(\alpha) < -\varepsilon$, one has

$$\mu_{\mathrm{B}(z,\delta)}\left(\left\{x \in \Omega : |f(x) - g(x)| > \frac{\delta}{2^{13} M \sqrt{n}} \max(\varepsilon, f(x))\right\}\right) > \frac{1}{8}.$$

3.3. From Lemma 3.2 to Theorem 1.1: a multi-scale exploration. An intermediate lemma in this argument will be the following:

Lemma 3.4. *There exists a universal constant $c > 0$ such that the following holds true. Let $\varepsilon > 0$, $\Omega \subset \mathbb{R}^n$ a convex set with $0 \in \Omega$ and $\mathrm{Cov}(\mu_\Omega) = \mathrm{Id}$. Let $f : \Omega \to [0, \infty)$ be a convex and 1-Lipschitz function. Then there exists a measure μ on Ω, a point $y \in \Omega$ and a direction $\theta \in \mathbb{S}^{n-1}$ such that for all $\alpha \in \Omega$ satisfying*

$$|\langle \alpha - y, \theta \rangle| \geq \frac{c\varepsilon}{16n^{10}}$$

and for every convex function $g\colon \Omega \to \mathbb{R}$ *satisfying* $g(\alpha) < -\varepsilon$*, one has*

$$\mu\Big(\Big\{x \in \Omega : |f(x) - g(x)| > \frac{c}{n^{7.5}\log(1+{}^{n}/_{\varepsilon})}\max(\varepsilon, f(x))\Big\}\Big) > \frac{c}{n^{2}\log(1+{}^{n}/_{\varepsilon})}. \tag{3.4}$$

3.3.1. From Lemma 3.4 to Theorem 1.1. Given Lemma 3.4, the proof of Theorem 1.1 is carried out by induction on the dimension. The case $n = 1$ has already been resolved above. Now, suppose that the theorem is true up to dimension $n-1$, where the constant $c > 0$ is the constant from Lemma 3.4. Let $\mathcal{K} \in \mathbb{R}^n$ and f satisfy the assumptions of the theorem. Denote $Q = \mathrm{Cov}(\mu_{\mathcal{K}})^{-1/2}$ and define

$$\Omega = Q(\mathcal{K}), \quad \tilde{f}(x) = f(Q^{-1}(x))$$

so that $\tilde{f}\colon \Omega \to \mathbb{R}$. Since $\mathrm{diam}(\mathcal{K}) \le 1$, we know that for all $u \in \mathbb{S}^{n-1}$, $\mathbb{V}\mathrm{ar}\,[\mathrm{Proj}_u\,\mu_K] \le 1$, where $\mathrm{Proj}_u\,\mu_K$ denotes the push forward of μ_K by $x \to \langle x, u\rangle$, in other words for a measurable $A \subset \mathbb{R}$, we define

$$\mathrm{Proj}_u\,\mu_K(A) = \mu_K\big(\{x : \langle x, u\rangle \in A\}\big).$$

This implies that $\|Q^{-1}\| \le 1$. Consequently, the function $\tilde{f}$ is 1-Lipschitz. We now invoke Lemma 3.4 on Ω and $\tilde{f}$ which outputs a measure μ_1, a point $y \in \Omega$ and a direction θ. By translating f and $\mathcal{K}$, we can assume without loss of generality that $y = 0$. Fix some linear isometry $T\colon \mathbb{R}^{n-1} \to \theta^{\perp}$. Define

$$\Omega' = T^{-1}\,\mathrm{Proj}_{\theta^{\perp}}\big(\Omega \cap \{x : |\langle x, \theta\rangle| \le \delta\}\big)$$

where $\delta = c\varepsilon/16n^{10}$ and c is the universal constant from Lemma 3.4. Since $\tilde{f}$ is convex, there exists $I \subset \mathbb{R} \times \mathbb{R}^n$ so that

$$\tilde{f}(x) = \sup_{(a,y)\in I} \left(a + \langle x, y\rangle\right), \quad \forall x \in \Omega. \tag{3.5}$$

We may extrapolate $\tilde{f}(x)$ to the domain $\mathbb{R}^n$ by using the above display as a definition. We now define a function $h\colon \Omega' \to \mathbb{R}$ by

$$h(x) := \sup_{w\in[-\delta,\delta]} \tilde{f}(T(x) + w\theta). \tag{3.6}$$

It is clear that $\mathrm{diam}(\Omega') \le 1$. Moreover, h is 1-Lipschitz since it can be written as the supremum of 1-Lipschitz functions. We can therefore use the induction hypothesis with $\Omega', h(x)$ to obtain a measure μ_2 on Ω'. Next, for $y \in \mathbb{R}^{n-1}$, define

$$N(y) := \big\{x \in \Omega : T^{-1}(\mathrm{Proj}_{\theta^{\perp}}\, x) = y\big\}$$

and set

$$\mu(W) = \frac{1}{n}\mu_1(Q(W)) + \frac{n-1}{n}\int_{\Omega'} \frac{\mathrm{Vol}_1(Q(W) \cap N(u))}{\mathrm{Vol}_1(N(u))}\, d\mu_2(u)$$

for all measurable $W \subset \mathbb{R}^n$, where $\mathrm{Vol}_1(\cdot)$ denotes the 1-dimensional Hausdorff–Lebesgue measure.

Fix $\alpha \in \mathcal{K}$, let $g\colon \mathcal{K} \to \mathbb{R}$ be a convex and 1-Lipschitz function satisfying $g(\alpha) \le -\varepsilon$. Recall that c denotes the universal constant from Lemma 3.4. Define

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

The proof will be concluded by showing that $\mu(A) \ge c/(n^3 \log(1 + n/\varepsilon))$.

Define $\tilde{g}(x) = g(Q^{-1}(x))$ and remark that $\tilde{g}$ is 1-Lipschitz. First consider the case that $|\langle Q\alpha, \theta\rangle| \ge \delta$, then by construction, we have

$$\begin{aligned} \mu(A) &\ge \frac{1}{n}\mu_1(Q(A)) \\ &= \frac{1}{n}\mu_1\Big(\Big\{x \in \Omega;\ |\tilde{f}(x) - \tilde{g}(x)| > \frac{c}{n^{7.5}\log(1+n/\varepsilon)} \max(\varepsilon, f(x))\Big\}\Big) \\ &\overset{(3.4)}{\ge} \frac{c}{n^3 \log(1 + n/\varepsilon)}, \end{aligned}$$

and we're done.

Otherwise, we need to deal with the case that $|\langle Q\alpha, \theta\rangle| < \delta$. Define $q(x)$ to be the function obtained by replacing $\tilde{f}(x)$ with $\tilde{g}(x)$ in equation (3.6) and consider the set

$$A' = \Big\{ x \in \Omega';\ |h(x) - q(x)| > \frac{c}{(n-1)^{7.5}\log(1 + n/\varepsilon)} \max(\varepsilon, h(x)) \Big\}.$$

By construction of the measure μ_2 we have $\mu_2(A') \ge c/((n-1)^3 \log(1 + n/\varepsilon))$. We claim that $N(A') \subset Q(A)$, which implies that

$$\mu(A) \ge \frac{n-1}{n} \int_{A'} \frac{\mathrm{Vol}_1(Q(A) \cap N(u))}{\mathrm{Vol}_1(N(u))}\, d\mu_2(u) = \frac{n-1}{n}\mu_2(A') \ge \frac{c}{n^3 \log(1 + n/\varepsilon)}$$

which will complete the proof. Indeed, let $y \in N(A')$. Define $z = T^{-1}(\mathrm{Proj}_{\theta^\perp} y)$, so that $z \in A'$. Let $w_1, w_2 \in N(z)$ be points such that

$$h(z) = \tilde{f}(w_1), \quad q(z) = \tilde{g}(w_2).$$

Such points exist since, by continuity, the maximum in equation (3.6) is attained. Now, since $z \in A'$, we have by definition that

$$|\tilde{f}(w_1) - \tilde{g}(w_2)| > \frac{c}{(n-1)^{7.5}\log(1 + n/\varepsilon)} \max(\varepsilon, \tilde{f}(w_1)).$$

Finally, since the functions $\tilde{f}, \tilde{g}$ are 1-Lipschitz, we have that

$$\begin{aligned}
|\tilde{f}(y) - \tilde{g}(y)| &\geq |\tilde{f}(w_1) - \tilde{g}(w_2)| - |\tilde{f}(y) - \tilde{f}(w_1)| - |\tilde{g}(y) - \tilde{g}(w_2)| \\
&\geq \frac{c}{(n-1)^{7.5}\log(1+n/\varepsilon)} \max(\varepsilon, \tilde{f}(w_1)) - |y - w_1| - |y - w_2| \\
&\geq \frac{c}{(n-1)^{7.5}\log(1+n/\varepsilon)} \max(\varepsilon, \tilde{f}(y)) - 4\delta \\
&= \frac{c}{(n-1)^{7.5}\log(1+n/\varepsilon)} \max(\varepsilon, \tilde{f}(y)) - \frac{c\varepsilon}{4n^{10}} \\
&\geq \frac{c}{n^{7.5}\log(1+n/\varepsilon)} \max(\varepsilon, \tilde{f}(y))
\end{aligned}$$

which implies, by definition, that $y \in Q(A)$. The proof is complete.

3.3.2. From Lemma 3.2 to Lemma 3.4. We construct below a decreasing sequence of domains $\Omega_0 \supset \Omega_1 \supset \cdots \supset \Omega_N$. Let $x_0 \in \Omega$ be a point where $f(x)$ attains its minimum on Ω. Set $\Omega_0 = \Omega - x_0$. Given $i \geq 0$, we define the domain Ω_{i+1}, given the domain Ω_i, by induction as follows. Define $Q_i = \mathrm{Cov}(\mu_{\Omega_i})^{-1/2}$ and $f_i(x) = f(Q_i^{-1}(x + x_0)) - f(x_0)$. We have

$$|\nabla f_i(x)| = \left|Q_i^{-1}\nabla f(Q_i^{-1}(x))\right| \leq \|Q_i^{-1}\|.$$

Now, by Lemma 3.7 we know that

$$\mathrm{diam}(\Omega_i) \leq \mathrm{diam}(\Omega) \leq n+1$$

which implies that $\left\|Q_i^{-1}\right\| \leq n+1$. We conclude that f_i is $(n+1)$-Lipschitz. We may therefore invoke Lemma 3.2 for the function f_i defined by on the set $Q_i\Omega_i$, with $L = n+1$. This lemma outputs a direction θ and a measure μ which we denote by θ_i and μ_i respectively. We define

$$\Omega_{i+1} = Q_i^{-1} S_{Q_i\Omega_i, \theta_i}.$$

Equation (3.2) yields that for a universal constant $c > 0$,

$$\mu_i\left(\left\{x - x_0 : |f(x) - g(x)| > \frac{c}{n^{7.5}\log(1+n/\varepsilon)} \max(\varepsilon, f(x))\right\}\right) > \frac{c}{n} \tag{3.7}$$

for all functions $g(x)$ such that $g(\alpha) < -\varepsilon$, whenever $\alpha \in \Omega_i \setminus \Omega_{i+1}$.

Fix a constant $c' > 0$ whose value will be assigned later on. Define $\delta = c'\varepsilon/16n^{10}$ and let

$$N = \min\left\{i : \exists\theta \in \mathbb{S}^{n-1} \text{ such that } |\langle x, \theta\rangle| < \delta, \forall x \in \Omega_i\right\}.$$

In other words, N is the smallest value of i such that Ω_i is contained in a slab of width 2δ. Our next goal is to give an upper bound for the value of N. To this end, we claim that

$$\mathrm{Vol}(\Omega_{i+1}) \leq \tfrac{1}{2}\mathrm{Vol}(\Omega_i), \tag{3.8}$$

which equivalently says

$$\mathrm{Vol}(S_{Q_i\Omega_i,\theta_i}) \le \tfrac{1}{2}\,\mathrm{Vol}(Q_i\Omega_i).$$

Let $X \sim \mu_{Q_i\Omega_i}$ and observe that $\mathbb{P}(|\langle X,\theta_i\rangle| \le \frac{1}{4}) = \mathrm{Vol}(S_{\Omega_i,\theta_i})/\,\mathrm{Vol}(Q_i\Omega_i)$. Clearly $\langle X,\theta_i\rangle$ is a log-concave random variable which, by definition of Q_i, has variance 1. Using that the density of a log-concave distribution of unit variance is bounded by 1 (see e.g. [12, Lemma 5.5]) one gets $\mathbb{P}(|\langle X,\theta_i\rangle| \le \frac{1}{4}) \le \frac{1}{2}$, which proves (3.8). It is now a simple application of Lemma 3.8 to see that for all i there exists a direction $v_i \in \mathbb{S}^{n-1}$ such that

$$\langle v_i, \mathrm{Cov}(\Omega_i) v_i\rangle \le c_1\sqrt{n}\,2^{-2i/n}.$$

where $c_1 > 0$ is a universal constant. Together with Lemma 3.7, this yields

$$\mathrm{diam}(\mathrm{Proj}_{v_i}\,\Omega_i) \le 2\sqrt{c_1}\,n^{5/4}2^{-i/n}.$$

By definition of N, this gives

$$N \le n\log_{1/2}\frac{n^{5/4}}{\sqrt{c_1}} + n\log_{1/2}\delta \le n(12 + 2c_1 + 40\log(1 + {}^{n}\!/\!_{\varepsilon}) - \log c').$$

Take $c' = \min(c,1)^2/(2^8(1+c_1))$. A straightforward calculation gives

$$\frac{c}{N} > \frac{c'}{n\log(1 + {}^{n}\!/\!_{\varepsilon})}. \tag{3.9}$$

Finally, we define

$$\mu(W) = \frac{1}{N}\sum_{i=1}^{N}\mu_i(W - x_0)$$

for all measurable $W \subset \mathbb{R}^n$.

For $\alpha \in \Omega \setminus \{x : |\langle x - x_0, v_N\rangle| \le \delta\}$ consider a convex function $g(x)$ satisfying $g(\alpha) < -\varepsilon$. Define $\widetilde{\alpha} = \alpha - x_0$ and $\widetilde{g}(x) = g(x + x_0) - f(x_0)$ and remark that $\widetilde{g}(\widetilde{\alpha}) < -\varepsilon$. By definition of N, there exists $1 \le i \le N$ such that $\widetilde{\alpha} \in \Omega_i \setminus \Omega_{i+1}$. Thus, equation (3.7) gives

$$\mu\Big(\Big\{x \in \Omega : |f(x) - g(x)| > \frac{c'}{2n^{7.5}\log(1 + {}^{n}\!/\!_{\varepsilon})}\max(\varepsilon, f(x))\Big\}\Big) > \frac{c}{nN} \overset{(3.9)}{>} \frac{c'}{n^2\log(1 + {}^{n}\!/\!_{\varepsilon})}.$$

The proof is complete. □

3.4. From Lemma 3.3 to Lemma 3.2: covering the space via regions with stable gradients. We say that a (z, θ, t) is a *jolly-good triplet* if $|z| \le \frac{1}{16}$ and (3.3) is satisfied for some appropriate δ, namely $\delta = 1/(Cn^6 |\log(1 + {}^{Ln}/_{\varepsilon})|)$ with $C > 0$ a universal constant whose value will be decided upon later on. Intuitively given Lemma 3.3 it is enough to find a polynomial (in n) number of jolly-good triplets for which the corresponding set of θ-directions partially covers the sphere $\mathbb{S}^{n-1}$. The notion of covering we use is the following: For a subset $H \subset \mathbb{S}^{n-1}$ and for $\gamma > 0$, we say that H is a γ*-cover* if for all $x \in \mathbb{S}^{n-1}$, there exists $\theta \in H$ such that $\langle \theta, x \rangle \ge -\gamma$.

Next we explain how to find jolly-good triplets in Section 3.4.1, and then how to find a γ-cover with such triplets in Section 3.4.2.

3.4.1. A contraction lemma. The following result shows that jolly-good triplets always exist, or in other words that a convex function always has a relatively big set on which the gradient map is approximately constant. Quite naturally the proof is based on a smoothing argument together with a Poincaré inequality.

Lemma 3.5. *Let $r, \eta, L > 0$ and $0 < \xi < 1$ such that $L > 2\eta r$. Let $\Omega \subset \mathbb{R}^n$ be a convex set, and $f\colon \Omega \to \mathbb{R}$ be L-Lipschitz and η-strongly convex, that is*

$$\nabla^2 f(x) \succeq \eta \, \mathrm{Id}, \quad \forall x \in \Omega.$$

Let $x_0 \in \Omega$ such that $\mathrm{B}(x_0, r) \subset \Omega$*. Then there exist a triplet*

$$(z, \theta, t) \in \mathrm{B}(x_0, r) \times \mathbb{S}^{n-1} \times [{}^{\eta r}/_2, +\infty)$$

such that

$$\mu_{\mathrm{B}(z,\delta)}\big((\nabla f)^{-1}(\mathrm{B}(t\theta, \xi t))\big) \ge \frac{1}{2} \tag{3.10}$$

for $\delta = \xi r/(16 n^2 \log {}^{L}/_{\eta r})$.

Proof. We consider the convolution $g = f \star h$, where h is defined by

$$h(x) = \frac{\mathbf{1}_{\{x \in \mathrm{B}(0,\delta)\}}}{\mathrm{Vol}(\mathrm{B}(0,\delta))}.$$

We clearly have that g is also η-strongly convex. Let $x_{\min}$ be the point where g attains its minimum in Ω. We claim that

$$|\nabla g(x)| \ge \frac{\eta r}{2}, \quad \forall x \in \Omega \setminus \mathrm{B}(x_{\min}, {}^{r}/_2). \tag{3.11}$$

Indeed by strong-convexity of g we have for all $y \in \Omega$,

$$|\nabla g(y)| \ge \frac{1}{|y - x_{\min}|} \langle \nabla g(y), y - x_{\min} \rangle \ge |y - x_{\min}| \eta.$$

which proves (3.11).

Next, define $B_0 = \mathrm{B}(x_0, r)$ and $D = B_0 \setminus \mathrm{B}(x_{\min}, {}^{r}\!/\!{}_{2})$. It is clear that

$$\frac{\mathrm{Vol}(D)}{\mathrm{Vol}(B_0)} \geq \frac{1}{2}.$$

Let ν be the push forward of μ_D under $x \mapsto |\nabla g(x)|$. According to (3.11) and by the assumption that f is L-Lipschitz, we know that ν is supported on $[{}^{\eta r}\!/\!{}_{2}, L]$. Thus, there exists some $t \in [{}^{\eta r}\!/\!{}_{2}, L]$ such that $\nu([t, 2t]) \geq (2 \log {}^{L}\!/\!{}_{\eta r})^{-1}$. Define

$$A = \left\{ x \in B_0 : |\nabla g(x)| \in [t, 2t] \right\},$$

so we know that

$$\frac{\mathrm{Vol}(A)}{\mathrm{Vol}(B_0)} \geq \frac{\mathrm{Vol}(A)}{\mathrm{Vol}(D)} \frac{\mathrm{Vol}(D)}{\mathrm{Vol}(B_0)} \geq \frac{1}{4 \log {}^{L}\!/\!{}_{\eta r}}.$$

Recall that

$$\frac{\mathrm{Vol}_{n-1}(\partial B(0, r))}{\mathrm{Vol}_n(B(0, r))} = \frac{n+1}{r}.$$

Using Lemma 3.9, we now have that

$$\begin{aligned}\frac{1}{\mathrm{Vol}(A)} \int_A \Delta g(x)\, dx &\leq t \frac{\mathrm{Vol}_{n-1}(\partial B_0)}{\mathrm{Vol}(A)} \\ &= t \frac{\mathrm{Vol}_{n-1}(\partial B_0)}{\mathrm{Vol}(B_0)} \frac{\mathrm{Vol}(B_0)}{\mathrm{Vol}(A)} \leq 8ntr^{-1} \log {}^{L}\!/\!{}_{\eta r}.\end{aligned}$$

Consequently, there exists a point $z \in A$ for which $|\nabla g(z)| \geq t$ and $\Delta g(z) \leq 8ntr^{-1} \log {}^{L}\!/\!{}_{\eta r}$. In other words, by the definition of g, we have that

$$\frac{1}{\mathrm{Vol}(\mathrm{B}(z, \delta))} \int_{\mathrm{B}(z,\delta)} \Delta f(x)\, dx \leq 8ntr^{-1} \log {}^{L}\!/\!{}_{\eta r}.$$

Fix $1 \leq i \leq n$, and define $w(x) = \langle \nabla f(x) - \nabla g(z), e_i \rangle$, where e_i is the ith vector of the standard basis. Note that

$$|\nabla w(x)| = |\nabla^2 f(x) e_i| \leq \Delta f(x).$$

By definition of $g(x)$ and $w(x)$, we have $\int_{\mathrm{B}(z,\delta)} w(x)\, dx = 0$, thus applying the Poincaré inequality for a ball (see e.g. [1]) yields that

$$\int_{\mathrm{B}(z,\delta)} |w(x)|\, dx \leq \delta \int_{\mathrm{B}(z,\delta)} |\nabla w(x)|\, dx.$$

Thus combining the last three displays, and using that $\delta = \xi r / (16 n^2 \log {}^{L}\!/\!{}_{\eta r})$, one obtains

$$\frac{1}{\mathrm{Vol}(\mathrm{B}(z, \delta))} \int_{\mathrm{B}(z,\delta)} |w(x)|\, dx \leq 8\delta ntr^{-1} \log {}^{L}\!/\!{}_{\eta r} \leq \frac{\xi t}{2n}.$$

By using the fact that $|\nabla f(x) - \nabla g(z)| \leq \sum_{i=1}^{n} |\langle \nabla f(x) - \nabla g(z), e_i \rangle|$, this yields

$$\frac{1}{\mathrm{Vol}(\mathrm{B}(z,\delta))} \int_{\mathrm{B}(z,\delta)} |\nabla f(x) - \nabla g(z)|\, dx \leq \frac{\xi t}{4} \leq \frac{\xi |\nabla g(z)|}{2}.$$

Finally applying Markov's inequality one obtains (3.10) for the triplet

$$\left(z, \frac{\nabla g(z)}{|\nabla g(z)|}, |\nabla g(z)| \right). \qquad \square$$

3.4.2. Concluding the proof with the contraction lemma. We first fix some $\eta > 0$ and, at this point, suppose that $\nabla^2 f(x) \succeq \eta \mathrm{Id}$ for all $x \in \Omega$. Later on we will argue that this assumption can be removed. Define $h_\Omega(x) = \sup_{y \in \Omega} \langle x, y \rangle$, the support function of Ω. Consider the set

$$\Theta = \left\{ \theta \in \mathbb{S}^{n-1} : \ h_\Omega(\theta) \leq \tfrac{1}{8} \right\}$$

and let H be set of directions obtained from jolly-good triplets, more precisely,

$$H = \Big\{ \theta \in \mathbb{S}^{n-1} : \exists z \in \mathbb{R}^n, t \in (0,1) \text{ such that (3.3) is true with } \delta = \frac{1}{2^{28} n^6 \log(1 + {}^{Ln}/_{\eta})} \Big\}.$$

Define $\gamma = 1/16n$. Next, we show that $H \cup \Theta$ is a γ-cover. Let $\varphi \in \mathbb{S}^{n-1}$. Our objective is to find $\theta \in H \cup \Theta$ such that $\langle \theta, \varphi \rangle \geq -\gamma$.

First suppose that $\varphi \notin 8\Omega$. In that case, by the Hahn–Banach theorem and since $0 \in \Omega$, there exists $w \in \mathbb{R}^n$ such that $\langle \varphi, w \rangle = 1$ and $\langle w, y \rangle \leq \frac{1}{8}$ for all $y \in \Omega$. In other words, we have for $\theta = w/|w|$ that

$$h_\Omega(\theta) \leq \frac{1}{8|w|} \leq \frac{1}{8},$$

which implies that $\theta \in \Theta$. Since $\langle \varphi, {}^{w}/_{|w|} \rangle \geq 0$, we are done.

We may therefore assume that $\varphi/8 \in \Omega$. Since $\mathrm{Cov}(\mu_\Omega) = \mathrm{Id}$, then by Lemma 3.7 there exists a point $w \in \mathbb{R}^n$ such that $|w| \leq n+1$ and $\mathrm{B}(w,1) \subset \Omega$. Define $r = 1/2^{13} n^2$ and take

$$B_0 = \mathrm{B}({}^{\varphi}/_{32} + rw, r).$$

Note that by convexity and by the fact that $0 \in \Omega$, we have that $B_0 \subset \Omega$. We now use Lemma 3.5 for the ball B_0 with $\xi = 1/2^{11} n^2$, and $\delta = 1/(2^{28} n^6 \log(1 + {}^{Ln}/_{\eta}))$ to obtain a jolly-good triplet $(z(\theta), \theta, t)$. Denote $z = z(\theta)$. We want to show that $\langle \theta, \varphi \rangle \geq -\gamma$. Observe that by convexity of f and since f attains its minimum at $x = 0$, one has $\langle \nabla f(x), x \rangle \geq 0$ for any x. Thus, by definition of a jolly-good triplet one can easily see that $\langle \theta, z\rangle \geq -(\xi+\delta)$. Also by definition z is in B_0 and thus $|32z-\varphi-32rw| \leq 32r$. This implies:

$$\begin{aligned}\langle \theta, \varphi\rangle &= \langle \theta, \varphi - 32z + 32rw\rangle + 32\langle \theta, z\rangle - 32r\langle \theta, w\rangle \\ &\geq -|\varphi - 32z + 32rw| - 32r|w| - 32\xi - 32\delta \geq -\frac{1}{16n}.\end{aligned}$$

This concludes the proof that $H \cup \Theta$ is a γ-cover.

Next we use Lemma 3.10 to extract a subset $H' \subset H$ such that $|H'| \leq n+1$ and $H' \cup \Theta$ is also a γ-cover for $\mathbb{S}^{n-1}$. An application of Lemma 3.11 with $M = 2n$ now gives that there exists $v \in \mathbb{S}^{n-1}$ such that

$$\Omega \cap \Big(\bigcap_{\theta \in H' \cup \Theta} \{x : \langle x, \theta\rangle \leq \tfrac{1}{8}\}\Big) = \Omega \cap \Big(\bigcap_{\theta \in H'} \{x : \langle x, \theta\rangle \leq \tfrac{1}{8}\}\Big) \subset S_{\Omega, v}.$$

Finally, an application of Lemma 3.3 gives us that for all $\alpha \in \Omega \setminus S_{\Omega,v}$ and every function g such that $g(\alpha) < -\varepsilon$ one has for some $\theta \in H'$,

$$\mu_{\mathrm{B}(z(\theta),\delta)}\Big(\Big\{x \in \Omega : |f(x) - g(x)| > \frac{\delta}{2^{13} M\sqrt{n}} \max(\varepsilon, f(x))\Big\}\Big) > \frac{1}{8}.$$

Defining $\mu = \frac{1}{|H'|}\sum_{\theta \in H'} \mu_{\mathrm{B}(z(\theta),\delta)}$, we get

$$\mu\Big(\Big\{x \in \Omega : |f(x) - g(x)| > \frac{1}{2^{42} n^{7.5} \log(1 + Ln/\eta)} \max(\varepsilon, f(x))\Big\}\Big) > \frac{1}{16n}. \tag{3.12}$$

It remains to remove the uniform convexity assumption. This is done by considering the function

$$x \mapsto f(x) + \eta|x|^2$$

in place of f in the above argument. Since $|x| \leq M \leq 2n$ for all $x \in \Omega$, the equation (3.12) becomes

$$\mu\Big(\Big\{x \in \Omega : |f(x) - g(x)| > \frac{c}{2^{42} n^{7.5} \log(1 + Ln/\eta)} \max(\varepsilon, f(x)) - 4n^2\eta\Big\}\Big) > \frac{1}{16n}.$$

Finally choosing $\eta = (\varepsilon/2^{20} n^{10})^2$ one easily obtains

$$\mu\Big(\Big\{x \in \Omega : |f(x) - g(x)| > \frac{1}{2^{50} n^{7.5} \log(1 + n/\varepsilon)} \max(\varepsilon, f(x))\Big\}\Big) > \frac{1}{16n},$$

which concludes the proof. □

3.5. Proof of Lemma 3.3. The main ingredient of the proof is the following technical result.

Lemma 3.6. *Let* $\Omega \subset \mathbb{R}^n$ *be a domain satisfying* $\mathrm{Diam}(\Omega) \leq M$. *Let* $f : \Omega \to [0, \infty)$ *be a non-negative convex function let* $g : \Omega \to \mathbb{R}$ *be a convex function satisfying* $g(\alpha) < -\varepsilon$, *for some* $\alpha \in \Omega$. *Let* $z \in \mathbb{R}^n$ *and consider the ball* $B = \mathrm{B}(z, \delta)$. *Let* $D \subset B$ *be a set satisfying*

$$\langle \nabla f(x), \alpha - x \rangle \geq 0, \quad \forall x \in D. \tag{3.13}$$

Assume also that $\mu_B(D) \geq \frac{1}{2}$ *and that* $|z - \alpha| \geq n\delta$. *Define*

$$A = \left\{ x : |f(x) - g(x)| > \frac{\delta}{2^{13} M \sqrt{n}} \max(\varepsilon, f(x)) \right\}.$$

Then one has $\mu_D(A) \geq \frac{1}{4}$.

Proof. For $x \in \Omega$, define $\Theta_\alpha(x) = (x - \alpha)/(|x - \alpha|)$ and for $\theta \in \mathbb{S}^{n-1}$ write $N(\theta) = \Theta_\alpha^{-1}(\theta)$. Denote by λ_θ the one-dimensional Lebesgue measure on the needle $N(\theta)$. Let σ_B, σ_D be the push-forward of μ_B, μ_D under Θ_α. Moreover, for every $\theta \in \mathbb{S}^{n-1}$, the disintegration theorem ensures the existence of a probability measure $\mu_{D,\theta}$ on $N(\theta)$, defined so that for every measurable test function h one has

$$\int h(x) \, d\mu_D(x) = \int_{\mathbb{S}^{n-1}} \int_{N(\theta)} h(x) \, d\mu_{D,\theta}(x) \, d\sigma_D(\theta) \tag{3.14}$$

(in other words, $\mu_{D,\theta}$ is the normalized restriction of μ_D to $N(\theta)$). Define the measures $(\mu_{B,\theta})_\theta$ in the same manner.

It is easy to verify that σ_D is absolutely continuous with respect the the uniform measure on $\mathbb{S}^{n-1}$, which we denote by σ. Denote

$$q(\theta) := \frac{d\sigma_D}{d\sigma}(\theta) \quad \text{and} \quad w(\theta) := \frac{d\sigma_B}{d\sigma}(\theta).$$

Using Lemma 3.13 we obtain that

$$\frac{d\mu_{D,\theta}}{d\lambda_\theta}(x) = \frac{\zeta_n}{\mathrm{Vol}(D) q(\theta)} |x - \alpha|^{n-1} \mathbf{1}_{\{x \in D\}}, \tag{3.15}$$

and

$$\frac{d\mu_{B,\theta}}{d\lambda_\theta}(x) = \frac{\zeta_n}{\mathrm{Vol}(B) w(\theta)} |x - \alpha|^{n-1} \mathbf{1}_{\{x \in B\}}, \tag{3.16}$$

where ζ_n is a constant depending only on n.

For every $\theta \in \mathbb{S}^{n-1}$, define $L(\theta)$ to be the length of the interval $N(\theta) \cap B$. Consider the set

$$\mathcal{L} = \left\{ \theta : \; L(\theta) > \frac{\delta}{32\sqrt{n}} \right\}.$$

According to Lemma 3.12 we have that

$$\int_{\mathbb{S}^{n-1}\setminus\mathcal{L}} w(\theta)\,d\sigma(\theta) \le \frac{1}{8}.$$

Now, since $D \subset B$ and $\mu_B(D) \ge \frac{1}{2}$, we have that $q(\theta) \le 2w(\theta)$ for all $\theta \in \mathbb{S}^{n-1}$, which gives

$$\sigma_D(\mathcal{L}) = \int_{\mathcal{L}} q(\theta)\,d\sigma(\theta) \ge \frac{3}{4}.$$

Next, consider the set

$$\mathcal{S} = \left\{\theta \in \mathbb{S}^{n-1};\ q(\theta) \ge \frac{w(\theta)}{4}\right\}.$$

Since $\int_{\mathbb{S}^{n-1}} \frac{q(\theta)}{w(\theta)} d\sigma_B(\theta) = 1$ we have

$$\sigma_D(\mathcal{S}) = \int_{\mathcal{S}} \frac{q(\theta)}{w(\theta)}\,d\sigma_B(\theta) = 1 - \int_{\mathbb{S}^{n-1}\setminus\mathcal{S}} \frac{q(\theta)}{w(\theta)}\,d\sigma_B(\theta) \ge \frac{3}{4}.$$

Using a union bound, we have that $\sigma_D(\mathcal{L} \cap \mathcal{S}) \ge \frac{1}{2}$.

Fix $\theta \in \mathcal{L} \cap \mathcal{S}$, we would like to give a lower bound on $\mu_{D,\theta}(A)$. In view of Lemma 3.1, we thus need an upper bound on the density of $\mu_{D,\theta}$. Recall that $\theta \in \mathcal{S}$, implies $\frac{q(\theta)}{w(\theta)} \ge \frac{1}{4}$ and that by (3.15) and (3.16), we have for all $x \in N(\theta) \cap B$,

$$\frac{d\mu_{D,\theta}}{d\mu_{B,\theta}}(x) = \frac{\operatorname{Vol}(B)w(\theta)}{\operatorname{Vol}(D)q(\theta)}\mathbf{1}_{x\in D} \le 8. \tag{3.17}$$

Denote $[a,b] = B \cap N(\theta)$ for $a,b \in \mathbb{R}^n$. Assume that a is the interior of the interval $[\alpha, b]$ (if this is not the case, we simply interchange between a and b). By the assumption $\theta \in \mathcal{L}$, we know that $|b-a| \ge \delta/32\sqrt{n}$. Writing $Z = \zeta_n / \operatorname{Vol}(B)w(\theta)$ so that, according to (3.16),

$$\frac{d\mu_{B,\theta}}{d\lambda_\theta}(x) = Z|x-\alpha|^{n-1}\mathbf{1}_{\{x\in B\}},$$

and since $\mu_{B,\theta}$ is a probability measure,

$$Z^{-1} = \int_a^b |x-\alpha|^{n-1}\,dx$$

where, by slight abuse of notation we assume that $a, b, \alpha \in \mathbb{R}$. Thus,

$$Z \le \frac{32\sqrt{n}}{\delta|a-\alpha|^{n-1}}.$$

Combined with (3.17), this finally gives

$$\frac{d\mu_{D,\theta}}{d\lambda_\theta}(x) \le 2^8 \frac{\sqrt{n}}{\delta} \frac{|x-\alpha|^{n-1}}{|a-\alpha|^{n-1}} \le 2^8 \frac{\sqrt{n}}{\delta}\left(\frac{|b-\alpha|}{|a-\alpha|}\right)^{n-1}$$
$$= 2^8 \frac{\sqrt{n}}{\delta}\left(1+\frac{|b-a|}{|a-\alpha|}\right)^{n-1} \le 2^8 \frac{\sqrt{n}}{\delta}\left(1+\frac{2\delta}{n\delta-\delta}\right)^{n-1} \le 2^8 e^2 \frac{\sqrt{n}}{\delta},$$

where in the second to last inequality we used the assumption that $|z-\alpha| \ge n\delta$.

Define the map $U:\mathbb{R} \to N(\theta)$ by

$$U(x) = \alpha + M(|\alpha| - x)\theta$$

and consider the functions $\widetilde{f}(x) = f(U(x))$ and $\widetilde{g}(x) = g(U(x))$. Denote $x_0 = \min U^{-1}(D \cap N(\theta))$ and remark that $x_0 \in [|\alpha|-1, |\alpha|]$. Note that, thanks to equation (3.13), the assumption (3.1) holds for the functions $\widetilde{f}, \widetilde{g}$ and the points $x_0, |\alpha|$. We can now invoke Lemma 3.1 for these functions with μ being the pullback of $\mu_{D,\theta}$ by $U(x)$. According to the above inequality one may take $\beta = 2^8 e^2 (M\sqrt{n}/\delta)$ and obtain

$$\mu_{D,\theta}(A) \ge \frac{1}{2}.$$

Integrating over $\theta \in \mathcal{L} \cap \mathcal{S}$ concludes the proof:

$$\mu_D(A) \ge \int_{\mathcal{S}\cap\mathcal{L}} \mu_{D,\theta}(A)\, d\sigma_D(\theta) \ge \frac{1}{2}\sigma_D(\mathcal{L}\cap\mathcal{S}) \ge \frac{1}{4}.$$ □

Proof of Lemma 3.3. Suppose that (z, θ, t) satisfy equation (3.10). Fix $\alpha \in \Omega$ satisfying $\langle \alpha, \theta \rangle \ge \frac{1}{8}$ and a function $g(x)$ satisfying $g(\alpha) < -\varepsilon$. Define $B = B(z,\delta)$ and $D = \left\{x \in B; |\nabla f(x) - \theta t| < \frac{1}{16} n^{-2} t\right\}$. Let μ_B be the uniform measure on B. According to (3.10), we know that $\mu_B(D) \ge \frac{1}{2}$. Now, for all $x \in D$ we have that $\nabla f(x) = t(\theta + y)$ with $|y| < \frac{1}{16} n^{-2}$ so we get

$$\left\langle \nabla f(x), \frac{\alpha - x}{|\alpha - x|} \right\rangle = \frac{t}{|\alpha - x|}\big(\langle \alpha, \theta\rangle + \langle \alpha - x, y\rangle - \langle x, \theta\rangle\big)$$
$$> \frac{t}{|\alpha - x|}\left(\tfrac{1}{8} - \tfrac{1}{16}\big(|\alpha| + |x|\big)n^{-2} - |x|\right) \ge 0, \quad \forall x \in D$$

where we used the fact that $D \subset B$ and so $|x| < |z| + \delta \le \frac{1}{16}$ and the fact that $|\alpha| \le 2n$. Note that the above implies the assumption (3.13). Moreover remark that

$$|z - \alpha| \ge \tfrac{1}{4} - \tfrac{1}{8} \ge \tfrac{1}{8} \ge n\delta.$$

We can thus now invoke Lemma 3.6 to get $\mu_B(A) \ge \frac{1}{8}$ where

$$A = \left\{x \in \Omega : |f(x) - g(x)| > \tfrac{\delta}{2^{13} M \sqrt{n}} \max(\varepsilon, f(x))\right\}.$$

This completes proof. □

3.6. Technical lemmas. We gather here various technical lemmas.

Lemma 3.7. *Let C be a convex body in $\mathbb{R}^n$. Then*

$$\operatorname{diam}(C) \le (n+1)\|\operatorname{Cov}(\mu_C)\|^{1/2}. \tag{3.18}$$

On the other hand, if $\operatorname{Cov}(\mu_C) \succeq \mathrm{Id}$ *then C contains a ball of radius* 1.

Furthermore, for all $v \in \mathbb{S}^{n-1}$ one has

$$\sup_{x\in C}\langle v, x\rangle - \inf_{x\in C}\langle v, x\rangle \le (n+1)\langle v, \operatorname{Cov}(\mu_C), v\rangle^{1/2}.$$

Proof. The first and second parts of the lemma are found in [3, Section 3.2.1]. For the third part, we write $C' = \operatorname{Cov}(C)^{-1/2}C$ and $u = \operatorname{Cov}(C)^{1/2}v/|\operatorname{Cov}(C)^{1/2}v|$. We have

$$\begin{aligned}\sup_{x\in C}\langle v, x\rangle - \inf_{x\in C}\langle v, x\rangle &= \sup_{x\in C'}\langle v, \operatorname{Cov}(C)^{1/2}x\rangle - \inf_{x\in C'}\langle v, \operatorname{Cov}(C)^{1/2}x\rangle \\ &= \sup_{x\in C'}\langle \operatorname{Cov}(C)^{1/2}v, x\rangle - \inf_{x\in C'}\langle \operatorname{Cov}(C)^{1/2}v, x\rangle \\ &= |\operatorname{Cov}(C)^{1/2}v|\Big(\sup_{x\in C'}\langle u, x\rangle - \inf_{x\in C'}\langle u, x\rangle\Big) \\ &\overset{(3.18)}{\le} (n+1)|\operatorname{Cov}(C)^{1/2}v|. \qquad \square\end{aligned}$$

Lemma 3.8. *Let $C \subset D \subset \mathbb{R}^n$ be two convex bodies with $0 \in C$. Suppose that* $\operatorname{Vol}(C)/\operatorname{Vol}(D) \le \delta$*, then there exists $u \in \mathbb{S}^{n-1}$ such that*

$$\langle u, \operatorname{Cov}(\mu_C)u\rangle \le c\sqrt{n}\delta^{2/n}\langle u, \operatorname{Cov}(\mu_D)u\rangle. \tag{3.19}$$

where $c > 0$ is a universal constant.

Proof. Define $\mu = \mu_D$ and $\nu = \mu_C$. By applying a linear transformation to both μ and ν, we can clearly assume that $\operatorname{Cov}(\mu) = \mathrm{Id}$. Let $f(x)$ be a log-concave probability density in $\mathbb{R}^n$. According to [10, Corollary 1.2 and Lemma 2.7], we have that

$$c_1 \le \Big(\sup_{x\in\mathbb{R}^n} f(x)\Big)^{1/n}\Big(\det\operatorname{Cov}(f)\Big)^{1/2n} \le c_2 n^{1/4} \tag{3.20}$$

where $c_1, c_2 > 0$ are universal constants. Denote by $f(x)$ and $g(x)$ the densities of μ and ν, respectively. Since the densities of μ, ν are binary-valued, we have that

$$\sup_{x\in\mathbb{R}^n} f(x) = f(0) \le \delta g(0) = \delta \sup_{x\in\mathbb{R}^n} g(x).$$

We finally get

$$\begin{aligned}\big(\det\operatorname{Cov}(\nu)\big)^{1/n} &\overset{(3.20)}{\le} c_2^2\sqrt{n}\,g(0)^{-2/n} \\ &= c_2^2\delta^{2/n}\sqrt{n}\,f(0)^{-2/n} \\ &\overset{(3.20)}{\le} (c_2/c_1)^2\sqrt{n}\,\big(\det\operatorname{Cov}(\mu)\big)^{1/n}\delta^{2/n} = (c_2/c_1)^2\sqrt{n}\,\delta^{2/n}.\end{aligned}$$

The lemma follows by taking u to be the eigenvector corresponding to the smallest eigenvalue of $\mathrm{Cov}(\nu)$. □

Lemma 3.9. *Let g be a convex function defined on a Euclidean ball $B \subset \mathbb{R}^n$. Let $A \subset B$ be a closed set such that $\forall x \in A$, $|\nabla g(x)| \le t$. Then*

$$\int_A \Delta g(x)\, dx \le t \ \mathrm{Vol}_{n-1}(\partial B).$$

Proof. Since g is convex, we can write

$$g(x) = \sup_{y \in B} w_y(x),$$

where $w_y(x) = \langle x - y, \nabla g(y) \rangle + g(y)$. Define

$$\widetilde{g}(x) = \sup_{y \in A} w_y(x).$$

Clearly $\widetilde{g}$ is convex and $\widetilde{g}(x) = g(x)$ for all $x \in A$. Moreover $|\nabla \widetilde{g}(x)| \le t$ for all $x \in \mathbb{R}^n$. Using Gauss's theorem, we have

$$\int_A \Delta g(x)\, dx \le \int_B \Delta \widetilde{g}(x)\, dx = \int_{\partial B} \langle \nabla \widetilde{g}(x), n(x) \rangle \, d\mathcal{H}_{n-1}(x) \le t \ \mathrm{Vol}_{n-1}(\partial B),$$

which concludes the proof. □

Let $\gamma > 0$. Recall that we say that $H \subset \mathbb{S}^{n-1}$ is a *γ-cover* if for all $x \in \mathbb{S}^{n-1}$, there exists $\theta \in H$ satisfying

$$\langle \theta, x \rangle \ge -\gamma. \tag{3.21}$$

Lemma 3.10. *Let $H \subset \mathbb{S}^{n-1}$ be a γ-cover. Then there exists a subset $I \subset H$ with $|I| \le n+1$ such that I is a γ-cover.*

Proof. We first claim that there is a point $y \in \mathrm{Conv}(H)$ with $|y| \le \gamma$. Indeed, if we assume otherwise then by the Hahn–Banach theorem there exists $\widetilde{\theta} \in \mathbb{S}^{n-1}$ such that $\langle \theta, \widetilde{\theta} \rangle > \gamma$ for all $\theta \in H$, which means the vector $-\widetilde{\theta}$ violates the assumption (3.21). By Caratheodory's theorem, there exists $I \subset H$ with $|I| \le n+1$ such that $y \in \mathrm{Conv}(I)$. Write $I = (\theta_1, \dots, \theta_{n+1})$. Now let $x \in \mathbb{R}^n$ with $|x| \le 1$. Then since $\langle x, y \rangle \ge -\gamma$, we have

$$\sum_{i=1}^{n+1} \alpha_i \langle x, \theta_i \rangle \ge -\gamma$$

for some non-negative coefficients $\{\alpha_i\}_{i=1}^{n+1}$ satisfying $\sum_{i=1}^{n+1} \alpha_i = 1$. Thus there exists $\theta \in I$ for which (3.21) holds. □

Lemma 3.11. *Let $\Omega \subset \mathbb{R}^n$ be a convex set with* $\mathrm{diam}(\Omega) \le M$ *and such that* $0 \in \Omega$. *Let H be a γ-cover. Then there exists $\widetilde{\theta} \in \mathbb{S}^{n-1}$ such that*

$$\big\{\alpha \in \Omega : \forall \theta \in H, \langle \alpha, \theta \rangle < M\gamma\big\} \subset \big\{\alpha \in \Omega : |\langle \alpha, \widetilde{\theta} \rangle| \le 2M\gamma\big\}.$$

Proof. Since $\{\alpha \in \Omega : \forall \theta \in H, \langle \alpha, \theta \rangle < M\gamma\}$ is a convex set which contains 0, showing that it does not contain a ball of radius $2M\gamma$ is enough to show that it is included in some slab $\{\alpha \in \Omega : |\langle \alpha, \widetilde{\theta} \rangle| \le 2M\gamma\}$. Now suppose that our set of interest $\{\alpha : \forall \theta \in H, \langle \alpha, \theta \rangle < M\gamma\}$ actually contains a ball $\mathrm{B}(x, 2M\gamma)$ with $|x| \in (0, M)$. Let $\theta \in H$ be such that $\langle x/|x|, \theta \rangle \ge -\gamma$, and thus in particular $\langle x, \theta \rangle \ge -M\gamma$. Then one has by the inclusion assumption that $\langle \theta, x + 2M\gamma\theta \rangle < M\gamma$, but on the other hand one also has $\langle \theta, x + 2\gamma M\theta \rangle \ge \gamma M$ which yields a contradiction, thus concluding the proof. □

Lemma 3.12. *Let $\delta > 0$, $x_0 \in \mathbb{R}^n$, $B = \mathrm{B}(x_0, \delta)$ and $\alpha \in \mathbb{R}^n \setminus B$. For $x \in \mathbb{R}^n$, define $\Theta_\alpha(x) = (x - \alpha)/(|x - \alpha|)$, and let σ_B be the push-forward of μ_B under Θ_α. For every $\theta \in \mathbb{S}^{n-1}$, define $L(\theta)$ to be the length of the interval $\Theta_\alpha^{-1}(\theta) \cap B$. Then one has*

$$\sigma_B\left(\theta : L(\theta) > \frac{\delta}{32\sqrt{n}}\right) \ge \frac{7}{8}.$$

Proof. Note that, by definition,

$$x \in B \text{ and } x + \frac{\delta}{32\sqrt{n}} \frac{\alpha - x}{|\alpha - x|} \in B \Rightarrow L\left(\Theta_\alpha(x)\right) > \frac{\delta}{32\sqrt{n}}.$$

Furthermore it is easy to show that for all $y \in B$,

$$y + \frac{\delta}{32\sqrt{n}} \frac{\alpha - x_0}{|\alpha - x_0|} \in B \Rightarrow y + \frac{\delta}{32\sqrt{n}} \frac{\alpha - y}{|\alpha - y|} \in B.$$

Thus letting $X \sim \mu_B$ we see that the lemma will be concluded by showing that

$$\mathbb{P}\left(X + \frac{\delta}{32\sqrt{n}} \frac{\alpha - x_0}{|\alpha - x_0|} \in B\right) \ge \frac{7}{8}.$$

Defining

$$\widetilde{B} = \mathrm{B}\left(x_0 - \frac{\delta}{32\sqrt{n}} \frac{\alpha - x_0}{|\alpha - x_0|}, \delta\right),$$

the statement boils down to proving that $\mathbb{P}(X \in \widetilde{B}) \ge \frac{7}{8}$. By applying an affine linear transformation to both B and $\widetilde{B}$, this is equivalent to

$$\frac{\mathrm{Vol}\Big(\mathrm{B}\left(-\frac{c}{2\sqrt{n}} e_1, 1\right) \cap B\left(\frac{c}{2\sqrt{n}} e_1, 1\right)\Big)}{\mathrm{Vol}(\mathrm{B}(0, 1))} \ge \frac{7}{8},$$

where e_1 is the first vector of the standard basis. Next, by symmetry around the hyperplane $e_1^{\perp}$, we have

$$\frac{\mathrm{Vol}\Big(\mathrm{B}\Big(-\frac{1}{64\sqrt{n}}\,e_1,1\Big)\cap\mathrm{B}\Big(\frac{1}{64\sqrt{n}}\,e_1,1\Big)\Big)}{\mathrm{Vol}(\mathrm{B}(0,1))} = \frac{2\,\mathrm{Vol}\Big(\mathrm{B}\Big(-\frac{1}{64\sqrt{n}}\,e_1,1\Big)\cap\{x;\langle x,e_1\rangle\geq 0\}\Big)}{\mathrm{Vol}(\mathrm{B}(0,1))}.$$

Thus, it is enough to show that $\mathbb{P}(|Z|>1/64\sqrt{n})\geq\frac{7}{8}$, where $Z=\langle X',e_1\rangle$ and $X'\sim\mu_{\mathrm{B}(0,1)}$. Observe that $\mathbb{V}\mathrm{ar}\,[Z]\geq\frac{1}{8n}$ and that Z is log-concave (in particular the density of $Z/\mathbb{V}\mathrm{ar}\,[Z]$ is bounded by 1). This implies that for any $t>0$

$$\mathbb{P}\big(|Z|<t\sqrt{\mathbb{V}\mathrm{ar}\,[Z]}\big)<2t,$$

and thus the lemma follows by taking $t=\frac{1}{16}$. □

Lemma 3.13. *Let $A\subset\mathbb{R}^n$. For $x\in\mathbb{R}^n$, define $\Theta_\alpha(x)=(x-\alpha)/(|x-\alpha|)$, and let σ_A be the push-forward of μ_A under Θ_α. Assume that σ_A is absolutely continuous with respect the the uniform measure σ on $\mathbb{S}^{n-1}$ and denote $q(\theta):=\frac{d\sigma_A}{d\sigma}(\theta)$. Finally let $\mu_{A,\theta}$ be the normalized restriction of μ_A on $N(\theta)=\Theta_\alpha^{-1}(\theta)$, defined so that for every measurable test function h one has*

$$\int h(x)\,d\mu_D(x)=\int_{\mathbb{S}^{n-1}}\int_{N(\theta)}h(x)\,d\mu_{A,\theta}(x)\,d\sigma_A(\theta). \tag{3.22}$$

Denoting ζ_n for the $(n-1)$-dimensional Hausdorff measure of $\mathbb{S}^{n-1}$ one then obtains

$$\frac{d\mu_{A,\theta}}{d\lambda_\theta}(x)=\frac{\zeta_n}{\mathrm{Vol}(A)q(\theta)}|x-\alpha|^{n-1}\mathbf{1}_{\{x\in B\}}. \tag{3.23}$$

Proof. First observe that the existence of $\mu_{A,\theta}$ is ensured by the disintegration theorem. Now remark that using the integration by polar coordinates formula we have for every measurable test function φ,

$$\int_{\mathbb{R}^n}\varphi(x)\,dx=\zeta_n\int_{\mathbb{S}^{n-1}}\int_0^\infty r^{n-1}\varphi(\alpha+r\theta)\,dr\,d\sigma(\theta).$$

Now, by definition of $q(\cdot)$, we have for every test function φ,

$$\int_{\mathbb{S}^{n-1}}\int_0^\infty r^{n-1}\varphi(\alpha+r\theta)\,dr\,d\sigma(\theta)=\int_{\mathbb{S}^{n-1}}\int_0^\infty r^{n-1}q(\theta)^{-1}\varphi(\alpha+r\theta)\,dr\,d\sigma_A(\theta).$$

Taking $\varphi(x)=h(x)\mathbf{1}_{x\in A}$, we finally get

$$\begin{aligned}\int h(x)\,d\mu_A(x)&=\frac{1}{\mathrm{Vol}(A)}\int_A h(x)\,dx\\&=\frac{\zeta_n}{\mathrm{Vol}(A)}\int_{\mathbb{S}^{n-1}}\int_0^\infty r^{n-1}q(\theta)^{-1}h(\alpha+r\theta)\mathbf{1}_{\{\alpha+r\theta\in A\}}\,dr\,d\sigma_A(\theta).\end{aligned}$$

Since the above is true for every measurable function h, together with equation (3.22) we get that for every function h and every $\theta \in \mathbb{S}^{n-1}$, one must have

$$\int_{N(\theta)} h(x)\, d\mu_\theta(x) = \frac{\zeta_n}{\mathrm{Vol}(D) q(\theta)} \int_0^\infty r^{n-1} h(\alpha + r\theta) \mathbf{1}_{\{\alpha + r\theta \in A\}}\, dr$$
$$= \frac{\zeta_n}{\mathrm{Vol}(A) q(\theta)} \int_{N(\theta)} |x - \alpha|^{n-1} h(x) \mathbf{1}_{\{x \in A\}}\, d\lambda_\theta(x)$$

and the claimed identity (3.23) follows. □

4. Proof of Theorem 1.3

Following [5] we reduce the proof of Theorem 1.3 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 [5] we analyze the Bayesian maximin regret with the information theoretic approach of [13], which we recall in Subsection 4.1. A key contribution of our work is then to propose in Subsection 4.2 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 1.1.

4.1. The information ratio. Let $\bar{\mathcal{K}} = \{\bar{x}_1, \ldots, \bar{x}_K\}$ be a $1/\sqrt{T}$-net of $\mathcal{K}$, formally taken to be $K_{1/\sqrt{T}} \cap (1/\sqrt{T})\mathbb{Z}^n$ where K_ε is the euclidean ε-extension of K. Note that

$$|\bar{\mathcal{K}}| \le \left| \tfrac{1}{\sqrt{T}} \mathbb{Z}^n \cap [-1,1]^n \right| \le (4T)^{n/2}.$$

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

$$R_T \le \sqrt{T} + \sum_{t=1}^T \big(\ell_t(x_t) - \ell_t(\bar{x}^*)\big). \tag{4.1}$$

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

$$r_t(x) = \mathbb{E}_t\big(\ell_t(x) - \ell_t(\bar{x}^*)\big) \quad \text{and} \quad v_t(x) = \mathrm{Var}_t\big(\mathbb{E}_t(\ell_t(x)|\bar{x}^*)\big), \tag{4.2}$$

where Var_t denotes the variance conditional on H_t. 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 $f_{i,t}, f_t \colon \mathcal{K} \to [0,1]$, $i \in [K]$, $t \in [T]$, be defined by, for $x \in \mathcal{K}$,

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

Then one can easily see that

$$r_t(x) = f_t(x) - \sum_{i=1}^{K} \alpha_{i,t} f_{i,t}(\bar{x}_i) \quad \text{and} \quad v_t(x) = \sum_{i=1}^{K} \alpha_{i,t}\big(f_t(x) - f_{i,t}(x)\big)^2. \tag{4.3}$$

The main observation in [13] is the following lemma, which gives a bound on the accumulation of information (see also [5, Appendix B] for a short proof).

Lemma 4.1. *One always has* $\mathbb{E} \sum_{t=1}^{T} v_t(x_t) \le \frac{1}{2}\log(K)$.

An important consequence of Lemma 4.1 is the following result which follows from an application of the Cauchy–Schwarz inequality (and (4.1)):

$$\mathbb{E}\sum_{t=1}^{T} r_t(x_t) \le \sqrt{T} + C \sum_{t=1}^{T} \sqrt{\mathbb{E} v_t(x_t)} \;\Rightarrow\; \mathbb{E} R_T \le 2\sqrt{T} + C\sqrt{\frac{T}{2}\log(K)}. \tag{4.4}$$

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

$$\mathbb{E}_t r_t(x_t) \le \frac{1}{\sqrt{T}} + C\sqrt{\mathbb{E}_t v_t(x_t)},\ \forall t \in [T] \;\Rightarrow\; \mathbb{E} R_T \le 2\sqrt{T} + C\sqrt{\frac{T}{2}\log(K)}. \tag{4.5}$$

[13] 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 $\mathbb{E}_t r_t(x_t)/\sqrt{\mathbb{E}_t v_t(x_t)} \le K$ (without any assumptions on the loss functions $\ell_t \colon \mathcal{K} \to [0,1]$). In [5] 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 [14]. We describe and analyze our new strategy in the next subsection.

4.2. 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 (4.5). To lighten notation we drop all time subscripts, e.g. one has

$$r(x) = f(x) - \sum_{i=1}^{K} \alpha_i f_i(\bar{x}_i) \quad \text{and} \quad v(x) = \sum_{i=1}^{K} \alpha_i \big(f_i(x) - f(x)\big)^2.$$

Our objective is to describe a random variable $X \in \mathcal{K}$ which satisfies

$$\mathbb{E} r(X) \leq \frac{1}{\sqrt{T}} + C\sqrt{\mathbb{E} v(X)}, \tag{4.6}$$

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 (4.6).

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 (4.6), 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

$$\alpha\big(\{i \in [K] : f_i(\bar{x}_i) \leq -\varepsilon\}\big) \geq \frac{|L|}{2\log(2/|L|)\varepsilon}. \tag{4.7}$$

Indeed assume that (4.7) 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

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

thus leading to a contradiction. We denote $I = \{i \in [K] : f_i(\bar{x}_i) \leq -\varepsilon\}$ with ε satisfying (4.7).

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

$$|f(\bar{x}) - f_i(\bar{x})| \geq \frac{c}{n^{7.5}\log(1 + n/\varepsilon)} \max(\varepsilon, f(\bar{x})). \tag{4.8}$$

We say that a point is *good* for f_i if it satisfies (4.8), 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 1.1. 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 $\alpha(J) \geq c|L|/(\varepsilon n^3 \log^2(1 + {n}/{\varepsilon}))$, the two above displays clearly implies (4.6).

References

[1] G. Acosta and R. Durán, An optimal poincaré inequality in L^1 for convex domains, *Proc. Amer. Math. Soc.*, **132** (2004), no. 1, 195–202. Zbl 1057.26010 MR 2021262

[2] A. Agarwal, D. P. Foster, D. Hsu, S. M. Kakade, and A. Rakhlin, Stochastic convex optimization with bandit feedback, *SIAM J. Optim.*, **23** (2013), no. 1, 213–240. Zbl 1270.90107 MR 3033105

[3] S. Brazitikos, A. Giannopoulos, P. Valettas, and B.-H. Vritsiou, *Geometry of isotropic convex bodies*, Mathematical Surveys and Monographs, 196, American Mathematical Society, Providence, RI, 2014. Zbl 1304.52001 MR 3185453

[4] S. Bubeck and N. Cesa-Bianchi, Regret analysis of stochastic and nonstochastic multi-armed bandit problems, *Found. Trends Mach. Learn.*, **5** (2012), no. 1, 1–122. Zbl 1281.91051

[5] 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)*, 266–278, JMLR: Workshop and Conference Proceedings, 40, 2015.

[6] S. Bubeck and R. Eldan, Multi-scale exploration of convex functions and bandit convex optimization, in *Proceedings of the 29th Annual Conference on Learning Theory (COLT)*, 583–589, JMLR: Workshop and Conference Proceedings, 49, 2016.

[7] S. Bubeck, R. Eldan, and Y. Lee, Kernel-based methods for bandit convex optimization, in *Proceedings of 49th Annual ACM SIGACT Symposium on the Theory of Computing, (STOC '17)*, 72–85, ACM, New York, 2017. Zbl 1370.90175 MR 3678172

[8] A. Flaxman, A. Kalai, and B. McMahan, Online convex optimization in the bandit setting: Gradient descent without a gradient, in *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 385–394, ACM, New York, 2005. Zbl 1297.90117 MR 2298287

[9] E. Hazan and Y. Li, An optimal algorithm for bandit convex optimization, 2016. arXiv:1603.04350

[10] B. Klartag, On convex perturbations with a bounded isotropic constant, *Geom. Funct. Anal.*, **16** (2006), no. 6, 1274–1290. Zbl 1113.52014 MR 2276540

[11] R. Kleinberg, Nearly tight bounds for the continuum-armed bandit problem, in *Advances in Neural Information Processing Systems (NIPS, 2004)*, 17, 697–704, MIT Press, Cambridge, MA, 2005.

[12] L. Lovász and S. Vempala, The geometry of logconcave functions and sampling algorithms, *Random Struct. Algorithms*, **30** (2007), no. 3, 307–358. Zbl 1122.65012 MR 2309621

[13] D. Russo and B. Van Roy, An information-theoretic analysis of Thompson sampling, *J. Mach. Learn. Res.*, **17** (2016), Paper No. 68, 30pp. Zbl 1360.62030 MR 3517091

[14] D. Russo and B. Van Roy, Learning to optimize via information directed sampling, 2017. arXiv:1403.5556

Received 20 June, 2017; revised 23 February, 2018

S. Bubeck, Microsoft Research, 1 Microsoft Way, Building 99/2955,
Redmond, WA 98052, USA

E-mail: sebubeck@microsoft.com

R. Eldan, Department of Mathematics, Weizmann Institute of Science,
234 Herzl Street, 76100 Rehovot, Israel

E-mail: ronen.eldan@gmail.com