The Sample Complexity of Level Set Approximation: Supplementary Materials

François Bachoc **Tommaso R. Cesari** **Sébastien Gerchinovitz**

A SAMPLE COMPLEXITY: FORMAL DEFINITIONS

We provide formal definitions for the notions of deterministic algorithm, sample complexity, and rate-optimal algorithm.

We first precisely define deterministic algorithms that query values of functions sequentially and rely only on this information to build approximations of their level sets (sketched in Online Protocol 1). The behavior of any such algorithm is completely determined by a pair (φ, ψ), where $\varphi = (\varphi_n)_{n\in\mathbb{N}^*}$ is a sequence of functions $\varphi_n \colon \mathbb{R}^{n-1} \to [0,1]^d$ mapping the $n-1$ previously observed values $f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_{n-1})$ to the next query point $\boldsymbol{x}_n$, and $\psi = (\psi_n)_{n\in\mathbb{N}^*}$ is a sequence of functions $\psi_n \colon \mathbb{R}^n \to \big\{\text{subsets of } [0,1]^d\big\}$ mapping the n currently known values $f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_n)$ to an approximation S_n of the target level set.

We can now define the notion of sample complexity, which corresponds to the smallest number of queries after which the outputs S_n of an algorithm are all ε-approximations of the level set $\{f = a\}$ (recall Definition 1 in the Introduction).

Definition 7 (Sample complexity)**.** For all functions $f \colon [0,1]^d \to \mathbb{R}$, all levels $a \in \mathbb{R}$, any deterministic algorithm A, and any accuracy $\varepsilon > 0$, we denote by $\mathfrak{n}(f, A, \varepsilon, a)$ the smallest number of queries to f that A needs in order for its output sets S_n to be ε-approximations of the level set $\{f = a\}$ for all $n \ge \mathfrak{n}(f, A, \varepsilon, a)$, i.e.,

$$\mathfrak{n}(f, A, \varepsilon, a) := \inf\big\{n' \in \mathbb{N}^* : \forall n \ge n', S_n \text{ is an } \varepsilon\text{-approximation of } \{f = a\}\big\} . \tag{10}$$

We refer to $\mathfrak{n}(f, A, \varepsilon, a)$ as the *sample complexity* of A (for the ε-approximation of $\{f = a\}$).

We can now define rate-optimal algorithms rigorously. At a high-level, they output the tightest (up to constants) approximations of level sets that can possibly be achieved by deterministic algorithms.

Definition 8 (Rate-optimal algorithm)**.** For any level $a \in \mathbb{R}$ and some given family $\mathcal{F}$ of real-valued functions defined on $[0,1]^d$, we say that a deterministic algorithm A is *rate-optimal* (for level a and family $\mathcal{F}$) if, in the worst-case, it needs the same number of queries (up to constants) of the best deterministic algorithm in order to output approximations of level sets within any given accuracy, i.e., if there exists a constant $\kappa = \kappa(a, \mathcal{F}) \ge 1$, depending only on a and $\mathcal{F}$, such that, for all $\varepsilon > 0$,

$$\sup_{f\in\mathcal{F}} \mathfrak{n}(f, A, \varepsilon, a) \le \kappa \inf_{A'\in\mathcal{A}} \sup_{f\in\mathcal{F}} \mathfrak{n}(f, A', \varepsilon, a) , \tag{11}$$

where $\mathcal{A}$ denotes the set of all deterministic algorithms.

B USEFUL INEQUALITIES ABOUT PACKING AND COVERING NUMBERS

For all $r > 0$, the *r-covering number* $\mathcal{M}(E, r)$ of a bounded subset E of $\mathbb{R}^d$ (with respect to the sup-norm $\|\cdot\|_\infty$) is the smallest cardinality of an r-covering of E, i.e.,

$$\mathcal{M}(E, r) := \min\big(k \in \mathbb{N}^* : \exists \boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathbb{R}^d,\, \forall \boldsymbol{x} \in E,\, \exists i \in \{1, \dots, k\},\, \|\boldsymbol{x} - \boldsymbol{x}_i\|_\infty \le r\big)$$

if E is nonempty, zero otherwise.

Covering numbers and packing numbers (2) are closely related. In particular, the following well-known inequalities hold (see, e.g., Wainwright 2019, Lemmas 5.5 and 5.7, with permuted notation of $\mathcal{M}$ and $\mathcal{N}$).[10]

Lemma 2. *For any subset E of $[0,1]^d$ and any real number $r > 0$,*

$$\mathcal{N}(E, 2r) \le \mathcal{M}(E, r) \le \mathcal{N}(E, r) \; . \tag{12}$$

Furthermore, for all $\delta > 0$ and all $r > 0$, if $B(\delta) = \{\boldsymbol{x} \in \mathbb{R} : \|\boldsymbol{x}\|_\infty \le \delta\}$,

$$\mathcal{M}\big(B(\delta), r\big) \le \left(1 + 2\frac{\delta}{r}\mathbb{I}_{r<\delta}\right)^d \; . \tag{13}$$

We now state a known lemma about packing numbers at different scales.

Lemma 3. *For any subset E of $[0,1]^d$ and any real numbers $r_1, r_2 > 0$,*

$$\mathcal{N}(E, r_1) \le \left(1 + 4\frac{r_2}{r_1}\mathbb{I}_{r_2>r_1}\right)^d \times \mathcal{N}(E, r_2) \; .$$

Proof. We can assume without loss of generality that E is nonempty and that $r_1 < r_2$. Then,

$$\begin{aligned}
\mathcal{N}(E, r_1) &\le \mathcal{M}(E, r_1/2) && \text{(by (12))}\\
&\le \mathcal{M}(E, r_2) \times \mathcal{M}\big(B(r_2), r_1/2\big) && \text{(see below)}\\
&\le \mathcal{N}(E, r_2) \times \mathcal{M}\big(B(r_2), r_1/2\big) && \text{(by (12))}\\
&\le \mathcal{N}(E, r_2) \times \left(1 + \frac{4r_2}{r_1}\right)^d \; . && \text{(by (13))}
\end{aligned}$$

The second inequality is obtained by building the $r_1/2$-covering of E in two steps. First, we cover E with balls of radius r_2. Second, we cover each ball of the first cover with balls of radius $r_1/2$. □

The next lemma upper bounds the packing number of the unit hypercube in the sup-norm, at all scales r.

Lemma 4. *For any positive real number $r > 0$, the r-packing number of the unit cube in the sup-norm satisfies*

$$\mathcal{N}\big([0,1]^d, r\big) \le \left(\left\lfloor \frac{1}{r} \right\rfloor + 1\right)^d \; .$$

Proof. Since the diameter (in the sup-norm $\|\cdot\|_\infty$) of the unit hypercube is 1, if $r \ge 1$, then the packing number is $\mathcal{N}\big([0,1]^d, r\big) = 1 \le (\lfloor 1/r \rfloor + 1)^d$. Consider now the case $r < 1$. Let $\rho := 1 - \lfloor 1/r \rfloor r \in [0, r)$ and G be the set of r-equispaced points $\{\rho/2,\ \rho/2 + r,\ \rho/2 + 2r,\ \dots,\ \rho/2 + \lfloor 1/r \rfloor r\}^d$. Note that each point in $[0,1]^d$ is at most $(r/2)$-away from a point in G (in the sup-norm), i.e., G is an $(r/2)$-covering of $[0,1]^d$. We can thus use (12) in Lemma 2 at scale $r/2$ so see that $\mathcal{N}\big([0,1]^d, r\big) \le \mathcal{M}\big([0,1]^d, r/2\big) \le |G| = \big(\lfloor 1/r \rfloor + 1\big)^d$. □

The next lemma upper bounds the r-packing number (in the sup-norm) of an inflated level set at scale r.

Lemma 5. *For any function $f \colon [0,1]^d \to \mathbb{R}$ and all scales $r \in (0,1)$,*

$$\mathcal{N}\Big(\big\{|f - a| \le r\big\},\ r\Big) \le 2^d \left(\frac{1}{r}\right)^d \; .$$

[10] The definition of r-covering number of a subset A of $\mathbb{R}^d$ implied by (Wainwright, 2019, Definition 5.1) is slightly stronger than the one used in our paper, because elements $x_1, \dots, x_k$ of r-covers belong to A rather than just $\mathbb{R}^d$. Even if we do not need it for our analysis, Inequality (13) holds also in this stronger sense.

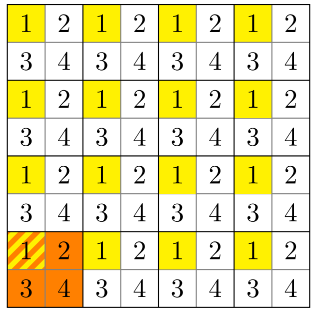

Figure 1: Constructing the partition when $d = 2$. In orange, the original enumeration A_0. In yellow, the family $\mathcal{C}_i(1)$.

Proof. Let $f : [0,1]^d \to \mathbb{R}$ be an arbitrary function and $r \in (0,1)$ any scale. By the monotonicity of the packing number ($E \subseteq F$ implies $\mathcal{N}(E,r) \le \mathcal{N}(E,r)$ by definition of packing number —Definition 2) and the previous lemma (Lemma 4), we get

$$\mathcal{N}\Big(\big\{|f-a| \le r\big\},\ r\Big) \le \mathcal{N}\big([0,1]^d,\, r\big) \le \left(\frac{1}{r}+1\right)^d \le \left(\frac{1}{r}+\frac{1}{r}\right)^d \le 2^d \left(\frac{1}{r}\right)^d .$$

□

C MISSING PROOFS OF SECTION 3

We now provide the missing proof of a claim we made in the proof of Theorem 2.

Claim 1. *Under the assumptions of Theorem 2, let $\delta \in (0,1)$ and $i \in \mathbb{N}^*$. Then, the family of hypercubes $\mathcal{C}_i$ maintained by Algorithm 2 can be partitioned into 2^d subfamilies $\mathcal{C}_i(1), \ldots, \mathcal{C}_i(2^d)$ with the property that for all $k \in \{1, \ldots, 2^d\}$ and all $C, C' \in \mathcal{C}_i(k)$, $C \neq C'$, we have $\inf_{\boldsymbol{x}\in C, \boldsymbol{y}\in C'} \|\boldsymbol{x}-\boldsymbol{y}\|_\infty > \delta\, 2^{-i}$*

Proof. We build our partition by induction. For a two-dimensional picture, see Figure 1. Denote the elements of the standard basis of $\mathbb{R}^d$ by $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d$. For any $\boldsymbol{x} \in [0,1]^d$ and all $E \subseteq [0,1]^d$, we denote by $E + \boldsymbol{x}$ the Minkowski sum $\big\{\boldsymbol{y}+\boldsymbol{x} : \boldsymbol{x} \in E\big\}$. Let E be the collection of all the hypercubes obtained by partitioning $[0,1]^d$ with a standard uniform grid with step size 2^{-i}, i.e., $E := \big\{[0,2^{-i}]^d + \sum_{k=1}^d r_k\, 2^{-i}\, \boldsymbol{e}_k : r_1, \ldots, r_k \in \{0,1,\ldots,2^i-1\}\big\}$.

Consider the family A_0 containing the hypercube $[0, 2^{-i}]$ and all other hypercubes of E adjacent to it; formally, $A_0 := \big\{[0,2^{-i}]^d + \sum_{k=1}^d r_k\, 2^{-i}\, \boldsymbol{e}_k : r_1, \ldots, r_d \in \{0,1\}\big\}$. Assign to each of the 2^d hypercubes in A_0 a distinct number between 1 and 2^d. Fix any $k \in \{0, \ldots, d-1\}$. For each hypercube $C \in A_k$, proceeding in the positive direction of the x_{k+1} axis, assign the same number as C to every other hypercube in E; formally, assign the same number as C to all hypercubes in $\big\{C + 2\,r\,2^{-i}\,\boldsymbol{e}_{k+1} : r \in \{1, \ldots 2^{i-1}-1\}\big\}$. Denote by A_{k+1} the collection of all hypercubes that have been assigned a number so far. By construction, A_d coincides with the whole E and consists of 2^d distinct subfamilies of hypercubes, each containing only hypercubes that have been assigned the same number. For any number $k \in \{1, \ldots, 2^d\}$, we denote by $\mathcal{C}_i(k)$ the subfamily of all hypercubes numbered with k. Fix any $k \in \{1, \ldots, 2^d\}$. By construction, each $C \in \mathcal{C}_i(k)$ contains no adjacent hypercubes. Thus, the smallest distance between two distinct hypercubes $C, C' \in \mathcal{C}_i(k)$ is $\inf_{\boldsymbol{x}\in C, \boldsymbol{y}\in C'} \|\boldsymbol{x}-\boldsymbol{y}\|_\infty \ge 2^{-i} > \delta\, 2^{-i}$ for all $\delta \in (0,1)$. □

D MISSING PROOFS OF SECTION 4

In this section, we prove Theorem 3 of Section 4. The proof is divided into two parts: one for the upper bound, one for the lower bound. Each time, we restate the corresponding result to ease readability.

D.1 Upper Bound

Proposition 1 (Theorem 3, upper bound)**.** *Consider the BAH algorithm run with input* a, c, γ*. Let* $f\colon [0,1]^d \to \mathbb{R}$ *be an arbitrary* (c,γ)*-Hölder function with level set* $\{f = a\} \neq \varnothing$*. Fix any accuracy* $\varepsilon > 0$*. Then, for all*

$$n > \kappa \, \frac{1}{\varepsilon^{d/\gamma}} \,, \qquad \text{where } \kappa := \left(2^{\gamma/d} 8^\gamma \, 2\, c\right)^{d/\gamma} ,$$

the output S_n *returned after the* n*-th query is an* ε*-approximation of* $\{f = a\}$*.*

Proof. The proof is a simple application of Theorem 2, with $(b, \beta) = (c, \gamma)$. Since we are assuming that the level set $\{f = a\}$ is nonempty, we only need to check that for all iterations i and all hypercubes $C' \in \mathcal{C}'_i$, the constant approximator $g_{C'} \equiv f(\boldsymbol{c}_{C'})$ is a (c,γ)-accurate approximation of f on C'. For any iteration i and all hypercubes $C' \in \mathcal{C}'_i$, we have that

$$\sup_{\boldsymbol{x} \in C'} \bigl| g_{C'}(\boldsymbol{x}) - f(\boldsymbol{x}) \bigr| = \sup_{\boldsymbol{x} \in C'} \bigl| f(\boldsymbol{c}_{C'}) - f(\boldsymbol{x}) \bigr| \le c\, 2^{-\gamma i} \,,$$

by definition of $g_{C'}$, the (c,γ)-Hölderness of f, and the fact that the diameter of all hypercubes $C' \in \mathcal{C}'_i$ (in the sup-norm) is 2^{-i}. Thus, Theorem 2 implies that for all $n > n(\varepsilon)$, the output S_n returned after the n-th query is an ε-approximation of $\{f = a\}$ where $n(\varepsilon)$ is

$$4^d \sum_{i=0}^{i(\varepsilon)-1} \lim_{\delta \to 1^-} \mathcal{N}\Bigl(\bigl\{ |f - a| \le 2\, c\, 2^{-\gamma i} \bigr\},\ \delta\, 2^{-i} \Bigr) \tag{14}$$

and $i(\varepsilon) := \lceil (1/\gamma) \log_2(2c/\varepsilon) \rceil$. If $\varepsilon \ge 2c$, than the sum in (14) ranges from 0 to a *negative* value, thus $n(\varepsilon) = 0$ by definition of sum over an empty set and the result is true with $\kappa = 0$. Assume then that $\varepsilon < 2c$ so that the sum in (14) is not trivially zero. Upper-bounding, for any $\delta \in (0,1)$ and all $i \ge 0$,

$$\mathcal{N}\Bigl(\bigl\{ |f - a| \le 2\, c\, 2^{-\gamma i} \bigr\},\ \delta\, 2^{-i} \Bigr) \le \mathcal{N}\bigl([0,1]^d,\ \delta\, 2^{-i} \bigr) \overset{(\dagger)}{\le} \bigl(2^i/\delta + 1 \bigr)^d \le (2/\delta)^d\, 2^{di}$$

(for completeness, we include a proof of the known upper bound $(\dagger)$ in Section B, Lemma 4) and recognizing the geometric sum below, we can conclude that

$$\begin{aligned} n(\varepsilon) &\le 8^d \sum_{i=0}^{\lceil (1/\gamma) \log_2(2c/\varepsilon) \rceil - 1} \bigl(2^d \bigr)^i \\ &= 8^d \, \frac{2^{d \lceil (1/\gamma) \log_2(2c/\varepsilon) \rceil} - 1}{2^d - 1} \\ &\le 8^d \, \frac{2^{d ((1/\gamma) \log_2(2c/\varepsilon) + 1)}}{2^d - \bigl(2^d/2 \bigr)} = 2\, 8^d \, (2\, c)^{d/\gamma} \, \frac{1}{\varepsilon^{d/\gamma}} \,. \end{aligned}$$

□

D.2 Lower Bound

In this section, we prove our lower bound on the worst-case sample complexity of Hölder functions. We begin by stating a simple known lemma on *bump* functions. Bump functions are a standard tool to build lower bounds in nonparametric regression (see, e.g., (Györfi et al., 2002, Theorem 3.2), whose construction we also adapt for our following result and Proposition 4).

Lemma 6. *Fix any amplitude* $\alpha > 0$*, a step-size* $\eta \in (0, 1/4]$*, let* $Z := \{0, 2\eta, \dots, \lfloor 1/2\eta \rfloor 2\eta\}^d \subseteq [0,1]^d$*, and fix an arbitrary* $\boldsymbol{z} = (z_1, \dots, z_d) \in Z$*. Consider the* bump *functions*

$$\widetilde{f}\colon \mathbb{R} \to [0,1] \qquad\qquad f_{\alpha,\eta,\boldsymbol{z}}\colon \mathbb{R}^d \to \mathbb{R}$$

$$x \mapsto \widetilde{f}(x) := \begin{cases} \exp\left(\dfrac{-x^2}{1 - x^2} \right) & \text{if } x \in (-1,1) \\ 0 & \text{otherwise}\,, \end{cases} \qquad\qquad \boldsymbol{x} \mapsto f_{\alpha,\eta,\boldsymbol{z}}(\boldsymbol{x}) := \alpha \prod_{j=1}^d \widetilde{f}\left(\frac{x_j - z_j}{\eta} \right) .$$

Then $\widetilde{f}$ is 3-Lipschitz and $f_{\alpha,\eta,\boldsymbol{z}}$ satisfies:

1. *$f_{\alpha,\eta,\boldsymbol{z}}$ is infinitely differentiable;*
2. *$f_{\alpha,\eta,\boldsymbol{z}}(\boldsymbol{x}) \in [0,\alpha)$ for all $\boldsymbol{x} \in \mathbb{R}^d \setminus \{\boldsymbol{z}\}$, and $f_{\alpha,\eta,\boldsymbol{z}}(\boldsymbol{z}) = \alpha$;*
3. *$\{f_{\alpha,\eta,\boldsymbol{u}_1} > 0\} \cap \{f_{\alpha,\eta,\boldsymbol{u}_2} > 0\} = \varnothing$ for any two distinct $\boldsymbol{u}_1, \boldsymbol{u}_2 \in Z$;*
4. *$\|\boldsymbol{x}-\boldsymbol{y}\|_\infty \le 2\eta$ for all $\boldsymbol{x},\boldsymbol{y}$ in the closure $\overline{\{f_{\alpha,\eta,\boldsymbol{z}} > 0\}}$ of $\{f_{\alpha,\eta,\boldsymbol{z}} > 0\}$ and all $\boldsymbol{z} \in Z$.*

The proof is a straightforward verification and it is therefore omitted. We now prove our worst-case lower bound for Hölder functions.

Proposition 2 (Theorem 3, lower bound)**.** *Fix any level $a \in \mathbb{R}$, any two Hölder constants $c > 0$, $\gamma \in (0,1]$, and an arbitrary accuracy $\varepsilon \in \big(0,\, c/(3d2^\gamma)\big)$. Let $n < \kappa/\varepsilon^{d/\gamma}$ be a positive integer, where $\kappa := (c/12d)^{d/\gamma}$. For each deterministic algorithm A there is a (c,γ)-Hölder function f such that, if A queries n values of f, then its output set S_n is not an ε-approximation of $\{f = a\}$. This implies in particular that (recall Definition 7),*

$$\inf_A \sup_f \mathfrak{n}(f, A, \varepsilon, a) \ge \kappa \frac{1}{\varepsilon^{d/\gamma}} \ ,$$

where the inf *is over all deterministic algorithms A and the* sup *is over all (c,γ)-Hölder functions f.*

Note that the leading constant $\kappa = (c/12d)^{d/\gamma}$ in our lower bound decreases quickly with the dimension d. Though we keep our focus on sample complexity rates, there are ways to improve the multiplicative constants appearing in our lower bounds. For instance, in the proof below, a larger constant $\kappa := (1/4\,(c/2)^{1/\gamma})^d$ can be obtained by replacing bump functions with *spike* functions $\boldsymbol{x} \mapsto \big[2\,\varepsilon - c\,\|\boldsymbol{x}-\boldsymbol{z}\|_\infty^\gamma\big]^+$, where $x \mapsto [x]^+ := \max\{x, 0\}$ denotes the positive part of x. We choose to use bump functions instead because they are well-suited for any smoothness (e.g., in Section E.2, we will apply the same argument to gradient-Hölder functions).

Proof. The following construction is a standard way to prove lower bounds on sample complexity (for a similar example, see Györfi et al. 2002, Theorem 3.2). Consider the set of bump functions $\{f_{\boldsymbol{z}}\}_{\boldsymbol{z}\in Z}$, where Z and $f_{\boldsymbol{z}} := f_{\alpha,\eta,\boldsymbol{z}}$ are defined as in Lemma 6,[11] for $\alpha := 2\varepsilon$ and some $\eta \in (0, 1/4]$ to be selected later. Fix an arbitrary $\boldsymbol{u} = (u_1,\dots,u_d) \in Z$. We show now that $f_{\boldsymbol{u}}$ is (c,γ)-Hölder, for a suitable choice of η. For all $\boldsymbol{x},\boldsymbol{y}$ in the closure $\overline{\{f_{\boldsymbol{u}} > 0\}}$ of $\{f_{\boldsymbol{u}} > 0\}$, Lemma 6 gives

$$\begin{aligned}\big|f_{\boldsymbol{u}}(\boldsymbol{x}) - f_{\boldsymbol{u}}(\boldsymbol{y})\big| &\le 2\varepsilon \sum_{j=1}^d \left| \widetilde{f}\left(\frac{x_j-u_j}{\eta}\right) - \widetilde{f}\left(\frac{y_j-u_j}{\eta}\right)\right| \le 2\varepsilon \sum_{j=1}^d 3\left|\frac{x_j-u_j}{\eta} - \frac{y_j-u_j}{\eta}\right| \le \frac{6\varepsilon d}{\eta}\|\boldsymbol{x}-\boldsymbol{y}\|_\infty \\ &= \frac{6\varepsilon d}{\eta}\|\boldsymbol{x}-\boldsymbol{y}\|_\infty^{1-\gamma}\|\boldsymbol{x}-\boldsymbol{y}\|_\infty^{\gamma} \le \frac{6\varepsilon d}{\eta}(2\eta)^{1-\gamma}\|\boldsymbol{x}-\boldsymbol{y}\|_\infty^{\gamma} = \frac{6\varepsilon d 2^{1-\gamma}}{\eta^\gamma}\|\boldsymbol{x}-\boldsymbol{y}\|_\infty^{\gamma} \ ,\end{aligned}$$

where the first inequality follows by applying d times the elementary consequence of the triangular inequality $\big|g_1(\boldsymbol{x}_1)g_2(\boldsymbol{x}_2) - g_1(\boldsymbol{y}_1)g_2(\boldsymbol{y}_2)\big| \le \max\big\{\|g_1\|_\infty, \|g_2\|_\infty\big\}\big(|g_1(\boldsymbol{x}_1) - g_1(\boldsymbol{y}_1)| + |g_2(\boldsymbol{x}_2) - g_2(\boldsymbol{y}_2)|\big)$, which holds for any two bounded functions $g_i \colon E_i \subseteq \mathbb{R}^{d_i} \to \mathbb{R}$ ($d_i \in \mathbb{N}^*, i \in \{1,2\}$). If $\boldsymbol{x}', \boldsymbol{y}' \notin \overline{\{f_{\boldsymbol{u}} > 0\}}$, then $f_{\boldsymbol{u}}(\boldsymbol{x}') = 0 = f_{\boldsymbol{u}}(\boldsymbol{y}')$, hence $\big|f_{\boldsymbol{u}}(\boldsymbol{x}') - f_{\boldsymbol{u}}(\boldsymbol{y}')\big| = 0$. Finally, if $\boldsymbol{x} \in \{f_{\boldsymbol{u}} > 0\}$ but $\boldsymbol{y}' \notin \overline{\{f_{\boldsymbol{u}} > 0\}}$, let $\boldsymbol{y}$ be the unique[12] point in the intersection of the segment $[\boldsymbol{x}, \boldsymbol{y}']$ and the boundary $\partial\{f_{\boldsymbol{u}} > 0\}$ of $\{f_{\boldsymbol{u}} > 0\}$; since $f_{\boldsymbol{u}}$ vanishes at the boundary of $\{f_{\boldsymbol{u}} > 0\}$, then $f_{\boldsymbol{u}}(\boldsymbol{y}) = f_{\boldsymbol{u}}(\boldsymbol{y}')$, therefore $\big|f_{\boldsymbol{u}}(\boldsymbol{x}) - f_{\boldsymbol{u}}(\boldsymbol{y}')\big| = \big|f_{\boldsymbol{u}}(\boldsymbol{x}) - f_{\boldsymbol{u}}(\boldsymbol{y})\big|$ and we can reapply the argument above for $\boldsymbol{x},\boldsymbol{y}$ now both in $\overline{\{f_{\boldsymbol{u}} > 0\}}$, obtaining

$$\big|f_{\boldsymbol{u}}(\boldsymbol{x}) - f_{\boldsymbol{u}}(\boldsymbol{y}')\big| = \big|f_{\boldsymbol{u}}(\boldsymbol{x}) - f_{\boldsymbol{u}}(\boldsymbol{y})\big| \le \frac{6\varepsilon d 2^{1-\gamma}}{\eta^\gamma}\|\boldsymbol{x}-\boldsymbol{y}\|_\infty^{\gamma} \le \frac{6\varepsilon d 2^{1-\gamma}}{\eta^\gamma}\|\boldsymbol{x}-\boldsymbol{y}'\|_\infty^{\gamma} \ ,$$

where the last inequality follows by $\|\boldsymbol{x}-\boldsymbol{y}\|_\infty \le \|\boldsymbol{x}-\boldsymbol{y}'\|_\infty$ and the monotonocity of $x \mapsto x^\gamma$ on $[0,\infty)$. Thus, selecting $\eta = (6\varepsilon d 2^{1-\gamma}/c)^{1/\gamma}$ so that $6\varepsilon d 2^{1-\gamma}/\eta^\gamma = c$, we obtain that $f_{\boldsymbol{z}}$ is (c,γ)-Hölder for all $\boldsymbol{z} \in Z$.

[11] More precisely, $f_{\boldsymbol{z}}$ is the restriction of $f_{\alpha,\eta,\boldsymbol{z}}$ to $[0,1]^d$.

[12] This follows from two simple observations. First, since $f_{\boldsymbol{u}}$ is continuous, the set $\{f_{\boldsymbol{u}} > 0\}$ is open, hence $\boldsymbol{x}$ belongs to its interior. Second, $\{f_{\boldsymbol{u}} > 0\}$ is (the interior of) a hypercube, therefore it is convex.

Moreover, by definition of Z (Lemma 6) and κ, we have that

$$|Z| = \left\lfloor \frac{1}{2\eta} + 1 \right\rfloor^d \geq \left(\frac{1}{2\eta}\right)^d = \left(\frac{1}{2(6\varepsilon d 2^{1-\gamma}/c)^{1/\gamma}}\right)^d = \left(\frac{c}{12d}\right)^{d/\gamma} \frac{1}{\varepsilon^{d/\gamma}} = \kappa \frac{1}{\varepsilon^{d/\gamma}} \ .$$

Recall that the sets $\{f_{\boldsymbol{z}_1} > 0\}$ and $\{f_{\boldsymbol{z}_2} > 0\}$ are disjoint for distinct $\boldsymbol{z}_1, \boldsymbol{z}_2 \in Z$ (Lemma 6). Thus, consider an arbitrary deterministic algorithm and assume that only $n < \kappa/\varepsilon^{d/\gamma}$ values are queried. By construction, there exists at least a $\boldsymbol{z} \in \mathcal{P}$ such that, if the algorithm is run for the level set $\{f = 0\}$ of the constant function $f \equiv 0$, no points are queried inside $\{f_{\boldsymbol{z}} > 0\}$ (and being f constant, the algorithm always observes 0 as feedback for the n evaluations). Being deterministic, if the algorithm is run for the level set $\{f_{\boldsymbol{z}} = 0\}$ of $f_{\boldsymbol{z}}$ it will also query no points inside $\{f_{\boldsymbol{z}} > 0\}$, observing only zeros for all the n evaluations. Since either way, only zeros are observed, using again the fact that the algorithm is deterministic, it returns the same output set S_n in both cases. This set cannot be simultaneously an ε-approximation of both $\{f = 0\}$ and $\{f_{\boldsymbol{z}} = 0\}$. Indeed, for the first set we have that $\{f = 0\} = [0,1]^d = \{f \leq \varepsilon\}$. Thus, if S_n is an ε-approximation of $\{f = 0\}$ it has to satisfy $\{f = 0\} \subseteq S_n \subseteq \{f \leq \varepsilon\}$, which in turn gives $S_n = [0,1]^d$. On the other hand, $\max_{\boldsymbol{x} \in [0,1]^d} f_{\boldsymbol{z}}(\boldsymbol{x}) = 2\,\varepsilon$, which implies that $\{f_{\boldsymbol{z}} \leq \varepsilon\}$ is *properly* included in $[0,1]^d$. Hence, if $S_n = [0,1]^d$ were also an ε-approximation of $\{f_{\boldsymbol{z}} = 0\}$, we would have that $[0,1]^d = S_n \subseteq \{f_{\boldsymbol{z}} \leq \varepsilon\} \neq [0,1]^d$, which yields a contradiction. This concludes the proof of the first claim. The second claim follows directly from the first part and Definition 7. □

E MISSING PROOFS OF SECTION 5

In this section, we present all missing proofs of our results in Section 4. We restate them to ease readability.

E.1 Upper Bound

Lemma (Lemma 1)**.** *Let $f \colon C' \to \mathbb{R}$ be a (c_1, γ_1)-gradient-Hölder function, for some $c_1 > 0$ and $\gamma_1 \in (0,1]$. Let $C' \subseteq [0,1]^d$ be a hypercube with diameter $\ell \in (0,1]$ and set of vertices V', i.e., $C' = \prod_{j=1}^d [u_j, u_j + \ell]$, for some $\boldsymbol{u} := (u_1, \dots, u_d) \in [0, 1-\ell]^d$, and $V' = \prod_{j=1}^d \{u_j, u_j + \ell\}$. The function*

$$\begin{aligned} h_{C'} \colon C' &\to \mathbb{R} \\ \boldsymbol{x} &\mapsto \sum_{\boldsymbol{v} \in V'} f(\boldsymbol{v}) \prod_{j=1}^d p_{v_j}(x_j) \ , \end{aligned}$$

where

$$p_{v_j}(x_j) := \left(1 - \frac{x_j - u_j}{\ell}\right) \mathbb{I}_{v_j = u_j} + \frac{x_j - u_j}{\ell} \mathbb{I}_{v_j = u_j + \ell},$$

interpolates the 2^d pairs $\big\{(\boldsymbol{v}, f(\boldsymbol{v}))\big\}_{\boldsymbol{v} \in V'}$ and it satisfies

$$\sup_{\boldsymbol{x} \in C'} \big| h_{C'}(\boldsymbol{x}) - f(\boldsymbol{x}) \big| \leq c_1 d\, \ell^{1+\gamma_1} \ .$$

Proof. Up to applying the translation $\boldsymbol{x} \mapsto \boldsymbol{x} + \boldsymbol{u}$, we can (and do) assume without loss of generality that $\boldsymbol{u} = \boldsymbol{0}$. The hypercube and its set of vertices then become $C' = [0,\ell]^d$ and $V' = \{0,\ell\}^d$ respectively. To verify that $h_{C'}$ interpolates the 2^d pairs $\big\{(\boldsymbol{v}, f(\boldsymbol{v}))\big\}_{\boldsymbol{v} \in V'}$, note that by definition of $h_{C'}$, for any vertex $\boldsymbol{w} \in V' = \{0,\ell\}^d$, we have

$$h_{C'}(\boldsymbol{w}) = \sum_{\boldsymbol{v} \in V'} f(\boldsymbol{v}) \prod_{j=1}^d p_{v_j}(w_j) = \sum_{\boldsymbol{v} \in V'} f(\boldsymbol{v}) \prod_{j=1}^d \mathbb{I}_{w_j = v_j} = f(\boldsymbol{w}) \ .$$

To prove the inequality, for all $k \in \{0, \dots, d\}$, let (P_k) be the property: if an $\boldsymbol{x} \in C'$ has at most k components which are not in $\{0,\ell\}$, then it holds that $\big| h_{C'}(\boldsymbol{x}) - f(\boldsymbol{x}) \big| \leq c_1 k\, \ell^{1+\gamma_1}$. To show that $\big| h_{C'}(\boldsymbol{x}) - f(\boldsymbol{x}) \big| \leq c_1 d\, \ell^{1+\gamma_1}$ for all $\boldsymbol{x} \in C'$ (therefore concluding the proof) we then only need to check that the property (P_d) is true. We do so by induction. If $k=0$, then (P_0) follows by $h_{C'}$ being an approximator for $\big\{(\boldsymbol{v}, f(\boldsymbol{v}))\big\}_{\boldsymbol{v}\in V'}$. Assume now that (P_k) holds for $k \in \{0,\dots,d-1\}$. To prove (P_{k+1}), fix an arbitrary $\boldsymbol{x} := (x_1,\dots,x_d) \in C'$, assume that $k+1$ components of $\boldsymbol{x}$ are not in $\{0,\ell\}$ and let $i \in \{1,\dots,d\}$ be any one of them (i.e., $x_i \in (0,\ell)$). Consider the two univariate functions

$$\begin{aligned} h_i \colon [0,\ell] &\to \mathbb{R} \\ t &\mapsto h_{C'}(x_1,\dots,x_{i-1},t,x_{i+1},\dots,x_d) \ , \\ f_i \colon [0,\ell] &\to \mathbb{R} \\ t &\mapsto f(x_1,\dots,x_{i-1},t,x_{i+1},\dots,x_d) \ . \end{aligned}$$

Being h_i linear (by definition of $h_{C'}$), we get

$$h_i(x_i) = \frac{\ell - x_i}{\ell} h_i(0) + \frac{x_i}{\ell} h_i(\ell) \ . \tag{15}$$

Being f_i continuous on $[0,\ell]$ and derivable on $(0,\ell)$ (by our assumptions on f), the mean value theorem applied to f_i on $[0,x_i]$ and $[0,\ell]$ respectively yields the existence of $\xi_1 \in (0,x_i)$ and $\xi_2 \in (0,\ell)$ such that

$$f_i(x_i) = f_i(0) + f_i'(\xi_1)\, x_i \ , \tag{16}$$

$$f_i(\ell) = f_i(0) + f_i'(\xi_2)\, \ell \ . \tag{17}$$

Putting everything together, we get that

$$\big|h_{C'}(\boldsymbol{x}) - f(\boldsymbol{x})\big| = \big|h_i(x_i) - f_i(x_i)\big|$$

(by definition of h_i and f_i). By (15) and (16), the right-hand side is equal to

$$\left| \frac{\ell - x_i}{\ell} h_i(0) + \frac{x_i}{\ell} h_i(\ell) - f_i(0) - f_i'(\xi_1)\, x_i \right| ,$$

which by the triangular inequality is at most

$$\left| \frac{\ell - x_i}{\ell} h_i(0) + \frac{x_i}{\ell} h_i(\ell) - f_i(0) - x_i \frac{f_i(\ell) - f_i(0)}{\ell} \right| + \left| x_i \frac{f_i(\ell) - f_i(0)}{\ell} - f_i'(\xi_1)\, x_i \right| .$$

By (17), this is equal to

$$\left| \frac{\ell - x_i}{\ell} \big(h_i(0) - f_i(0)\big) + \frac{x_i}{\ell} \big(h_i(\ell) - f_i(\ell)\big) \right| + \big| x_i\, f_i'(\xi_2) - f_i'(\xi_1)\, x_i \big| \ .$$

Finally, using again the triangular inequality, we can further upper bound with

$$\frac{\ell - x_i}{\ell} \underbrace{\Big|\big(h_i(0) - f_i(0)\big)\Big|}_{\le c_1 k \ell^{1+\gamma_1}} + \frac{x_i}{\ell} \underbrace{\Big|\big(h_i(\ell) - f_i(\ell)\big)\Big|}_{\le c_1 k \ell^{1+\gamma_1}} + \underbrace{x_i}_{\le \ell} \underbrace{\big|f_i'(\xi_2) - f_i'(\xi_1)\big|}_{\le c_1 \ell^{\gamma_1}} \le c_1(k+1)\ell^{1+\gamma_1} \ ,$$

where on the last line, we applied property (P_k) to the first two terms and we upper bounded the last one leveraging the (c_1,γ_1)-Hölderness of the gradients of f. This proves (P_{k+1}) and concludes the proof. □

Proposition 3 (Theorem 4, upper bound)**.** *Consider the BAG algorithm (Algorithm 4) run with input a, c_1, γ_1. Let $f \colon [0,1]^d \to \mathbb{R}$ be an arbitrary (c_1,γ_1)-gradient-Hölder function with level set $\{f=a\} \neq \varnothing$. Fix any accuracy $\varepsilon > 0$. Then, for all*

$$n > \kappa \frac{1}{\varepsilon^{d/(1+\gamma_1)}} \ , \ \text{if } \kappa := \big(2^{5+4\gamma_1+(1+\gamma_1)/d}\, c_1 d\big)^{d/\gamma} \ ,$$

the output S_n returned after the n-th query is an ε-approximation of $\{f=a\}$.

Proof. We proceed as in the proof of Theorem 3. Theorem 2 implies that for all $n > n(\varepsilon)$, where $i(\varepsilon) := \lceil (1/(1+\gamma_1)) \log_2(2c_1 d/\varepsilon) \rceil$ and $n(\varepsilon)$ is

$$8^d \sum_{i=0}^{i(\varepsilon)-1} \lim_{\delta \to 1^-} \mathcal{N}\Big(\big\{ |f-a| \le 2c_1 d 2^{-(1+\gamma_1)i} \big\}, \delta 2^{-i} \Big) ,$$

the output S_n returned after the n-th query is an ε-approximation of $\{f = a\}$. If $\varepsilon \ge 2c_1 d$, than the sum in the definition of $n(\varepsilon)$ ranges from 0 to a *negative* value, thus $n(\varepsilon) = 0$ by definition of sum over an empty set and the result is true with $\kappa = 0$. Assume then that $\varepsilon < 2c_1 d$ so that such sum is not trivially zero. Upper-bounding, for any $\delta \in (0,1)$ and all $i \ge 0$,

$$\mathcal{N}\Big(\big\{ |f-a| \le 2c_1 d\, 2^{-(1+\gamma_1)i} \big\},\ \delta\, 2^{-i} \Big) \le \mathcal{N}\big([0,1]^d,\ \delta\, 2^{-i} \big) \overset{(\dagger)}{\le} \big(2^i/\delta + 1 \big)^d \le (2/\delta)^d\, 2^{di}$$

(for completeness, we include a proof of the known upper bound (†) in Section B, Lemma 4) and recognizing the geometric sum below, we can conclude that

$$\begin{aligned} n(\varepsilon) &\le 16^d \sum_{i=0}^{\lceil (1/(1+\gamma_1)) \log_2(2c_1 d/\varepsilon) \rceil - 1} \big(2^d\big)^i \\ &\le 16^d\, \frac{2^{d((1/(1+\gamma_1)) \log_2(2c_1 d/\varepsilon)+1)}}{2^d - \big(2^d/2\big)} \\ &= 2 \cdot 16^d\, (2c_1 d)^{d/(1+\gamma_1)}\, \frac{1}{\varepsilon^{d/(1+\gamma_1)}} \ . \end{aligned}$$

□

E.2 Lower Bound

We conclude the section by proving a matching lower bound. Similarly to Proposition 2, we adapt some already known techniques from nonparametric regression (see, e.g., Györfi et al. 2002, Theorem 3.2).

Proposition 4 (Theorem 4, lower bound)**.** *Fix any level $a \in \mathbb{R}$, any two Hölder constants $c_1 > 0$, $\gamma_1 \in (0,1]$, and any accuracy $\varepsilon \in \big(0,\, c_1/(132 d 2^{3+\gamma_1})\big)$. Let $n < \kappa/\varepsilon^{d/(1+\gamma_1)}$ be a positive integer, where $\kappa := \big(c_1/(528d)\big)^{d/(1+\gamma_1)}$. For each deterministic algorithm A there is a (c_1,γ_1)-gradient-Hölder function f such that, if A queries n values of f, then its output set S_n is not an ε-approximation of $\{f = a\}$. This implies in particular that (recall Definition 7),*

$$\inf_A \sup_f \mathfrak{n}(f, A, \varepsilon, a) \ge \kappa\, \frac{1}{\varepsilon^{d/(1+\gamma_1)}} \ ,$$

where the inf *is over all deterministic algorithms A and the* sup *is over all (c_1,γ_1)-gradient-Hölder functions f.*

As we pointed out after Proposition 2, the leading constant $\kappa = \big(c_1/(528d)\big)^{d/(1+\gamma_1)}$ in our lower bound is small, and could likely be improved using smoothness-specific perturbations of the zero function, instead of the more universal bump functions.

Proof. The following construction is a standard way to prove lower bounds on sample complexity (for a similar example, see Györfi et al. 2002, Theorem 3.2). Consider the set of bump functions $\{f_{\boldsymbol{z}}\}_{\boldsymbol{z}\in Z}$, where Z and $f_{\boldsymbol{z}} := f_{\alpha,\eta,\boldsymbol{z}}$ are defined as in Lemma 6,[13] for $\alpha := 2\varepsilon$ and some $\eta \in (0, 1/4]$ to be selected later. Fix an arbitrary $\boldsymbol{u} = (u_1, \dots, u_d) \in Z$. We show now that $f_{\boldsymbol{u}}$ is (c_1,γ_1)-gradient-Hölder, for a suitable choice of η. This is sufficient to prove the result, following the same argument as in the proof of Proposition 2. Note

[13] More precisely, $f_{\boldsymbol{z}}$ is the restriction of $f_{\alpha,\eta,\boldsymbol{z}}$ to $[0,1]^d$.

first that for all $i \in \{1,\dots,d\}$ and any $\boldsymbol{x} \in [0,1]^d$, denoting by ∂_i the partial derivative with respect to the i-th variable,

$$\partial_i f_{\boldsymbol{u}}(\boldsymbol{x}) = \frac{2\varepsilon}{\eta} \widetilde{f}'\left(\frac{x_i - u_i}{\eta}\right) \prod_{\substack{j=1 \\ j \neq i}}^{d} \widetilde{f}\left(\frac{x_j - u_j}{\eta}\right) .$$

Hence, using the fact that $\widetilde{f}$ is 3-Lipschitz (Lemma 6) and 22-gradient-Lipschitz (the latter can be done by checking that $\left\|\widetilde{f}''\right\|_\infty \le 22$), for all $i \in \{1,\dots,d\}$ and any $\boldsymbol{x},\boldsymbol{y} \in [0,1]^d$, we get

$$\left|\partial_i f_{\boldsymbol{u}}(\boldsymbol{x}) - \partial_i f_{\boldsymbol{u}}(\boldsymbol{y})\right| \le \frac{2\varepsilon}{\eta} 3 \left(22 \left| \frac{x_i - u_i}{\eta} - \frac{y_i - u_i}{\eta} \right| + 3 \sum_{\substack{j=1 \\ j \neq i}}^{d} \left| \frac{x_j - u_j}{\eta} - \frac{y_j - u_j}{\eta} \right| \right) \le \frac{132\varepsilon d}{\eta^2} \|\boldsymbol{x} - \boldsymbol{y}\|_\infty , \quad (18)$$

where the first inequality follows by applying d times the elementary consequence of the triangular inequality $\left|g_1(\boldsymbol{x}_1)g_2(\boldsymbol{x}_2) - g_1(\boldsymbol{y}_1)g_2(\boldsymbol{y}_2)\right| \le \max\left\{\|g_1\|_\infty, \|g_2\|_\infty\right\}\left(|g_1(\boldsymbol{x}_1) - g_1(\boldsymbol{y}_1)| + |g_2(\boldsymbol{x}_2) - g_2(\boldsymbol{y}_2)|\right)$, which holds for any two bounded functions $g_i \colon E_i \subseteq \mathbb{R}^{d_i} \to \mathbb{R}$ ($d_i \in \mathbb{N}^*, i \in \{1,2\}$), and then using the Lipschitzness of $\widetilde{f}$ and $\widetilde{f}'$. Similarly to Proposition 2, to prove that $f_{\boldsymbol{u}}$ is (c_1,γ_1)-gradient-Hölder, we only need to check that the gradient of $f_{\boldsymbol{u}}$ is (c_1,γ_1)-Hölder on the closure $\overline{\{f_{\boldsymbol{u}} > 0\}}$ of $\{f_{\boldsymbol{u}} > 0\}$. For all $\boldsymbol{x},\boldsymbol{y} \in \overline{\{f_{\boldsymbol{u}} > 0\}}$, Equation (18) and Lemma 6 yield

$$\begin{aligned}
\left\|\nabla f_{\boldsymbol{u}}(\boldsymbol{x}) - \nabla f_{\boldsymbol{u}}(\boldsymbol{y})\right\|_\infty &\le \frac{132\varepsilon d}{\eta^2}\|\boldsymbol{x} - \boldsymbol{y}\|_\infty = \frac{132\varepsilon d}{\eta^2}\|\boldsymbol{x} - \boldsymbol{y}\|_\infty^{1-\gamma_1}\|\boldsymbol{x} - \boldsymbol{y}\|_\infty^{\gamma_1} \\
&\le \frac{132\varepsilon d}{\eta^2}(2\eta)^{1-\gamma_1}\|\boldsymbol{x} - \boldsymbol{y}\|_\infty^{\gamma_1} = \frac{132\varepsilon d 2^{1-\gamma_1}}{\eta^{1+\gamma_1}}\|\boldsymbol{x} - \boldsymbol{y}\|_\infty^{\gamma_1} .
\end{aligned}$$

Therefore, selecting $\eta = (132\varepsilon d 2^{1-\gamma_1}/c_1)^{1/(1+\gamma_1)}$ so that $132\varepsilon d 2^{1-\gamma_1}/\eta^{1+\gamma_1} = c_1$, we obtain that $f_{\boldsymbol{z}}$ is (c_1,γ_1)-Hölder for all $\boldsymbol{z} \in Z$. Moreover, by definition of Z and κ, we have that

$$|Z| \ge \left(\frac{1}{2\eta}\right)^d = \left(\frac{1}{2(132\varepsilon d 2^{1-\gamma_1}/c_1)^{1/(1+\gamma_1)}}\right)^d = \left(\frac{c_1}{528d}\right)^{d/(1+\gamma_1)} \frac{1}{\varepsilon^{d/(1+\gamma_1)}} = \kappa \frac{1}{\varepsilon^{d/(1+\gamma_1)}} .$$

Thus, proceeding as in the proof of Proposition 2, no deterministic algorithm can output a set that is an ε-approximation of the level set $\{f = 0\} = [0,1]^d$ of the constant function $f \equiv 0$ and simultaneously an ε-approximation of the level set $\{f_{\boldsymbol{z}} = 0\}$ of all bump functions $f_{\boldsymbol{z}}$ ($\boldsymbol{z} \in Z$), without querying at least one value in each one of the $|Z| \ge \kappa/\varepsilon^{d/(1+\gamma_1)}$ disjoint sets $\{f_{\boldsymbol{z}} > 0\}$ ($\boldsymbol{z} \in Z$) when applied to $f \equiv 0$. □

F THE BENEFITS OF ADDITIONAL STRUCTURAL ASSUMPTIONS

In this section, we present some examples showing how our general results can be applied to yield (slightly) improved sample complexity bounds when f satisfies additional structural assumptions, such as convexity.

F.1 NLS Dimension

In order to derive more readable bounds on the number of queries needed to return approximations of level sets, we now introduce a quantity that measures the difficulty of finding such approximations.

Definition 9 (NLS dimension)**.** Fix any level $a \in \mathbb{R}$ and a function $f \colon [0,1]^d \to \mathbb{R}$. We say that $d^\star \in [0,d]$ is a *NLS (or Near-Level-Set) dimension* of the level set $\{f = a\}$ if there exists $C^\star > 0$ such that (recalling Definition 2 —packing number)

$$\forall r \in (0,1),\ \mathcal{N}\Big(\big\{|f - a| \le r\big\},\ r\Big) \le C^\star \left(\frac{1}{r}\right)^{d^\star} . \quad (19)$$

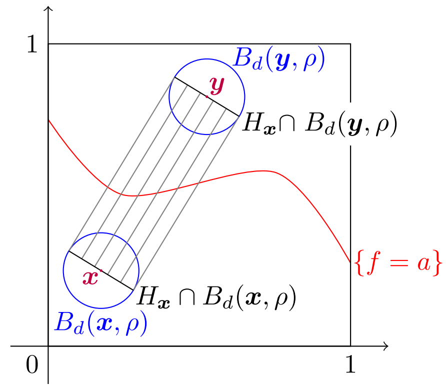

Figure 2: The "dimension" of the level set is at least the same as that of hyperplanes $H_{\boldsymbol{x}}$ and $H_{\boldsymbol{y}}$.

NLS dimensions are a natural generalization of the well-known concept of near-optimality dimension, from the field of non-convex optimization (see, e.g., (Bouttier et al., 2020, Section 2.3 and following discussion in Appendix B)). The idea behind Inequality (19) is that inflated level sets at, say, scale $r \in (0,1)$, are hard to pinpoint if their complement $\big\{|f-a|>r\big\}$ is large. Since for any increasing sequence $r := r_0 < r_1 < r_2 < \dots$, the set $\big\{|f-a|>r\big\}$ of points at which f is more than r-away from a can be decomposed into a union of "layers" $\big\{r_0 < |f-a| \le r_1\big\}, \big\{r_1 < |f-a| \le r_2\big\}, \big\{r_2 < |f-a| \le r_3\big\}, \dots$, and each of these layers $\big\{r_{s-1} < |f-a| \le r_s\big\}$ is by definition included in $\big\{|f-a| \le r_s\big\}$, by controlling the size of each of these $\big\{|f-a| \le r_s\big\}$ we can control the size of $\big\{|f-a|>r\big\}$. Therefore, by controlling how large the inflated level sets $\big\{|f-a| \le r\big\}$ can be at all scales $r \in (0,1)$, the parameters $C^\star$ and $d^\star$ quantify the difficulty of the level set approximation problem. In contrast, scales $r \ge 1$ are not informative since in this case the packing number in (19) is always 1. To see this, simply note that if $r \ge 1$, no more than 1 strictly r-separated point can be packed in $\big\{|f-a| \le r\big\}$, which is included in $[0,1]^d$, that has diameter 1 (in the sup-norm).

The dimension d of the domain is always a NLS dimension of any function $f\colon [0,1]^d \to \mathbb{R}$ (with $C^\star = 2^d$; we add a proof of this claim in Section B, Lemma 5). Hence, it is sufficient to consider NLS dimensions $d^\star \le d$, as we do in our Definition 9. While (as we will see in Section F.2) $d^\star$ is in general strictly smaller than d, bounds expressed in terms of a NLS dimension should only be considered slight refinements of worst-case bounds expressed in terms of d. Indeed, the following result shows that, with the exceptions of sets of minimizers and maximizers, level sets $\{f=a\}$ of continuous functions f have NLS dimension at least $d-1$.

Theorem 6 (Theorem 1)**.** *Let $f\colon [0,1]^d \to \mathbb{R}$ be a non-constant continuous function, and $a \in \mathbb{R}$ be any level such that $\min_{\boldsymbol{x}\in[0,1]^d} f(\boldsymbol{x}) < a < \max_{\boldsymbol{x}\in[0,1]^d} f(\boldsymbol{x})$. Then, there exists $C^\star > 0$ such that, for all $r>0$,*

$$\mathcal{N}\big(\{f=a\},\ r\big) \ge C^\star \left(\frac{1}{r}\right)^{d-1} .$$

Proof. For all $d_0 \in \mathbb{N}^*$, $\boldsymbol{z} \in \mathbb{R}^{d_0}$, and $\rho > 0$, we denote by $B_{d_0}(\boldsymbol{z},\rho)$ the closed d_0-dimensional Euclidean ball $\{\boldsymbol{x} \in \mathbb{R}^{d_0} : \|\boldsymbol{x}-\boldsymbol{z}\|_2 \le \rho\}$ with center $\boldsymbol{z}$ and radius ρ. Since a is neither the maximum nor the minimum of f and f is continuous, then the two sets $\{f<a\}$ and $\{f>a\}$ are non-empty and open. Therefore, we claim that there exist two points $\boldsymbol{x} \in \{f<a\}$, $\boldsymbol{y} \in \{f>a\}$, and a radius $\rho>0$, such that $B_d(\boldsymbol{x},\rho) \subseteq \{f<a\} \cap (0,1)^d$ and $B_d(\boldsymbol{y},\rho) \subseteq \{f>a\} \cap (0,1)^d$ (Figure 2). To see this, note that if f were identically equal to a on $(0,1)^d$, then, by continuity, f would be identically equal to a on the whole $[0,1]^d$, contradicting the assumption that it is non-constant. Hence there exists an $\boldsymbol{x} \in (0,1)^d$ such that $f(\boldsymbol{x}) \ne a$. Assume that $f(\boldsymbol{x}) < a$ (for the opposite case, proceed analogously). Then, being $\{f<a\} \cap (0,1)^d$ open, there exists a radius $\rho_1 > 0$ such that $B_d(\boldsymbol{x},\rho_1) \subseteq \{f<a\} \cap (0,1)^d$. Now, if f were lower than or equal to a on $(0,1)^d$, then, by continuity, f would be lower than or equal to a on the whole $[0,1]^d$ (and so would be its maximum), contradicting the assumption that $a < \max(f)$. Hence, there exists an $\boldsymbol{y} \in (0,1)^d$ such that $f(\boldsymbol{y}) > a$. Then, being $\{f>a\} \cap (0,1)^d$ open, there exists a radius $\rho_2 > 0$ such that $B_d(\boldsymbol{y},\rho_2) \subseteq \{f>a\} \cap (0,1)^d$. The claim is therefore proven by letting $\rho := \min(\rho_1,\rho_2)$.

Now, for all $r \geq \rho/\sqrt{d}$, we have

$$\mathcal{N}\big(\{f = a\},\ r\big) \geq 1 \geq \left(\frac{\rho}{\sqrt{d}}\right)^{d-1} \left(\frac{1}{r}\right)^{d-1}$$

and the result is proven with $C^\star = \big(\rho/\sqrt{d}\big)^{d-1}$.

Fix now an arbitrary $r \in \big(0, \rho/\sqrt{d}\big)$. Consider the line $\mathcal{L} := \big\{(1-t)\boldsymbol{x} + t\boldsymbol{y} : t \in \mathbb{R}\big\}$ passing through $\boldsymbol{x}$ and $\boldsymbol{y}$ and the two hyperplanes $H_{\boldsymbol{x}}$ and $H_{\boldsymbol{y}}$ orthogonal to $\mathcal{L}$ and passing through $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively. We denote, for each $\boldsymbol{z} \in \mathbb{R}^d$ and $E \subseteq \mathbb{R}^d$, the Minkowski sum $\{\boldsymbol{z} + \boldsymbol{u} : \boldsymbol{u} \in E\}$ of $\{\boldsymbol{z}\}$ and E by $\boldsymbol{z} + E$. Note that, by construction, $(\boldsymbol{y} - \boldsymbol{x}) + \big(H_{\boldsymbol{x}} \cap B_d(\boldsymbol{x}, \rho)\big) = H_{\boldsymbol{y}} \cap B_d(\boldsymbol{y}, \rho)$, and there is a rigid transformation $T \colon \mathbb{R}^d \to \mathbb{R}^d$ that maps $H_{\boldsymbol{x}} \cap B_d(\boldsymbol{x}, \rho)$ into the $(d-1)$-dimensional Euclidean ball $B_{d-1}(\boldsymbol{0}, \rho) = \big\{\boldsymbol{z} \in \mathbb{R}^{d-1} : \|\boldsymbol{z}\|_2 \leq \rho\big\}$ of $\mathbb{R}^{d-1}$ (where, with a slight abuse of notation, we identify from here on out $\mathbb{R}^{d-1}$ with the subspace $\big\{(z_1, \ldots, z_d) \in \mathbb{R}^d : z_d = 0\big\}$ of $\mathbb{R}^d$). By the symmetry of the Euclidean balls, for all $\boldsymbol{z}' \in \big(H_{\boldsymbol{x}} \cap B_d(\boldsymbol{x}, \rho)\big)$ and $\rho' > 0$, the transformed through the rigid transformation T of the intersection $B_d(\boldsymbol{z}', \rho') \cap \big(H_{\boldsymbol{x}} \cap B_d(\boldsymbol{x}, \rho)\big)$ of an arbitrary d-dimensional Euclidean ball $B_d(\boldsymbol{z}', \rho')$ centered at $H_{\boldsymbol{x}} \cap B_d(\boldsymbol{x}, \rho)$ and $H_{\boldsymbol{x}} \cap B_d(\boldsymbol{x}, \rho)$ itself is simply the intersection $B_{d-1}(\boldsymbol{z}'', \rho') \cap B_{d-1}(\boldsymbol{0}, \rho)$ between the ball $B_{d-1}(\boldsymbol{0}, \rho)$ and a $(d-1)$-dimensional ball $B_{d-1}(\boldsymbol{z}'', \rho')$ with some center $\boldsymbol{z}'' \in B_{d-1}(\boldsymbol{0}, \rho)$ and the same radius ρ' of $B_d(\boldsymbol{z}', \rho')$. We recall that for any dimension $d_0 \in \mathbb{N}^*$, norm $\|\cdot\|$ on $\mathbb{R}^{d_0}$, scale $r_0 > 0$, and non-empty subset E_0 of $\mathbb{R}^{d_0}$, a set $P \subseteq E_0$ is an r_0-packing of E_0 in $\mathbb{R}^{d_0}$ with respect to $\|\cdot\|$ if each two distinct points $\boldsymbol{z}_1, \boldsymbol{z}_2 \in P$ satisfy $\|\boldsymbol{z}_1 - \boldsymbol{z}_2\| > r_0$, and a set $C \subseteq E_0$ is an r_0-covering of E_0 in $\mathbb{R}^{d_0}$ with respect to $\|\cdot\|$ if for all $\boldsymbol{z} \in E_0$ there exists $\boldsymbol{c} \in C$ such that $\|\boldsymbol{z} - \boldsymbol{c}\| \leq r_0$; we denote by $\mathcal{N}_{d_0, \|\cdot\|}(E_0, r_0)$ the largest cardinality of an r_0-packing of E_0 in $\mathbb{R}^{d_0}$ with respect to $\|\cdot\|$, and by $\mathcal{M}_{d_0, \|\cdot\|}(E_0, r_0)$ the smallest cardinality of an r_0-covering of E_0 in $\mathbb{R}^{d_0}$ with respect to $\|\cdot\|$. By the previous observation, then, for all $r_0 > 0$, a set is an r_0-packing (resp., covering) of $H_{\boldsymbol{x}} \cap B_d(\boldsymbol{x}, \rho)$ in $\mathbb{R}^d$ with respect to the d-dimensional Euclidean norm if and only if its transformed under T is an r_0-packing (resp., covering) of $B_{d-1}(\boldsymbol{0}, \rho)$ in $\mathbb{R}^{d-1}$ with respect to the $(d-1)$-dimensional Euclidean norm. Hence

$$\begin{aligned} &\mathcal{N}_{d, \|\cdot\|_2}\Big(H_{\boldsymbol{x}} \cap B_d(\boldsymbol{x}, \rho),\ \sqrt{d}\, r\Big) \\ &\quad = \mathcal{N}_{d-1, \|\cdot\|_2}\big(B_{d-1}(\boldsymbol{0}, \rho),\ \sqrt{d}\, r\big) \\ &\quad \geq \mathcal{M}_{d-1, \|\cdot\|_2}\big(B_{d-1}(\boldsymbol{0}, \rho),\ \sqrt{d}\, r\big) \geq \left(\frac{\rho}{\sqrt{d}\, r}\right)^{d-1}, \end{aligned}$$

where the first inequality follows from the fact that each packing that is maximal with respect to the inclusion is also a covering, and the second one is a known lower bound on the number of balls with the same radius that are needed to cover a ball with a bigger radius, expressed in terms of a ratio of volumes (see, e.g., (Wainwright, 2019, Lemma 5.7)). Thus, we determined a $\big(\sqrt{d}\, r\big)$-packing P of $H_{\boldsymbol{x}} \cap B_d(\boldsymbol{x}, \rho)$ in $\mathbb{R}^d$ with respect to the d-dimensional Euclidean norm consisting of $C^\star (1/r)^{d-1}$ points, where again $C^\star := \big(\rho/\sqrt{d}\big)^{d-1}$. For all $\boldsymbol{p} \in P$, consider the segment $[\boldsymbol{p}, \boldsymbol{p} + \boldsymbol{y} - \boldsymbol{x}]$. By construction, all these segments are parallel, with an endpoint in $H_{\boldsymbol{x}} \cap B_d(\boldsymbol{x}, \rho) \subseteq \{f < a\}$ and the other in $H_{\boldsymbol{y}} \cap B_d(\boldsymbol{y}, \rho) \subseteq \{f > a\}$. Thus, the d-dimensional Euclidean distance between any two points belonging to distinct segments is at least equal to the minimum distance between the corresponding lines, which is strictly greater than $\sqrt{d}\, r$ by construction. By the continuity of f, then, for each $\boldsymbol{p} \in P$ there exists a $\boldsymbol{p}_a$ belonging to the segment $[\boldsymbol{p}, \boldsymbol{p} + \boldsymbol{y} - \boldsymbol{x}]$ such that $f(\boldsymbol{p}_a) = a$ which, together with the previous remark, implies that the family $P_a := \bigcup_{\boldsymbol{p} \in P} \boldsymbol{p}_a$ obtained this way is a $\big(\sqrt{d} r\big)$-packing of $\{f = a\}$ in $\mathbb{R}^d$ with respect to the d-dimensional Euclidean norm. Since the two norms $\|\cdot\|_\infty$ and $\|\cdot\|_2$ on $\mathbb{R}^d$ satisfy $\|\cdot\|_\infty \geq \|\cdot\|_2/\sqrt{d}$, then P_a is also an r-packing of $\{f = a\}$ in $\mathbb{R}^d$ with respect to the sup-norm $\|\cdot\|_\infty$, therefore its cardinality $|P_a| = C^\star (1/r)^{d-1}$ is smaller than or equal to the largest cardinality $\mathcal{N}\big(\{f = a\}, r\big)$ of an r-packing of $\{f = a\}$ with respect to the sup-norm $\|\cdot\|_\infty$. This concludes the proof. □

We remark that our definition in Equation (19) could be refined by considering variable $d^\star(r)$ and $C^\star(r)$ at different scales r. This would take into account that at different scales, the inflated level sets could have smaller size. Notably, our general result (Theorem 2) would naturally adapt to this finer definition as they are stated in terms of packing numbers at decreasing scales. For the sake of clarity, in this work we will stick to our worst-case definition of NLS dimension and we begin by showing how Theorem 2 has an immediate corollary in terms of $d^\star$. Note that in the following results, our BA instances are oblivious to the NLS dimension. Also, recall from the comment before Theorem 6 that typical level sets have NLS dimension $d^\star \geq d-1$.

Corollary 2 (of Theorem 2)**.** *Consider a Bisect and Approximate algorithm (Algorithm 2) run with input a, k, b, β. Let $f\colon [0,1]^d \to \mathbb{R}$ be an arbitrary function with level set $\{f = a\} \neq \varnothing$ and let $d^\star \in [0,d]$ be a NLS dimension of $\{f = a\}$ (Definition 9). Assume that the approximators $g_{C'}$ (defined at line 10) are (b,β)-accurate approximations of f (Definition 4), with $\beta \geq 1$. Fix any accuracy $\varepsilon > 0$. Then, for all $n > m(\varepsilon)$, the output S_n returned after the n-th query is an ε-approximation of $\{f = a\}$, where*

$$m(\varepsilon) := \begin{cases} \kappa_1 + \kappa_2 \log_2\left(\dfrac{1}{\varepsilon^{1/\beta}}\right)^+ & \textit{if } d^\star = 0\,, \\ \kappa(d^\star)\dfrac{1}{\varepsilon^{d^\star/\beta}} & \textit{if } d^\star > 0\,, \end{cases}$$

for $\kappa_1, \kappa_2, \kappa(d^\star) \geq 0$ independent of ε, that depend exponentially on the dimension d, where $x^+ = \max\{x, 0\}$ for all $x \in \mathbb{R}$.

Proof. Since all the conditions of Theorem 2 are met by assumption, we have that for all $n > n(\varepsilon)$, the output S_n returned after the n-th query is an ε-approximation of $\{f = a\}$, where $n(\varepsilon)$ is

$$4^d\, k \sum_{i=0}^{i(\varepsilon)-1} \lim_{\delta \to 1^-} \mathcal{N}\Big(\big\{|f-a| \leq 2\, b\, 2^{-\beta i}\big\},\ \delta\, 2^{-i}\Big) \tag{20}$$

and $i(\varepsilon) := \big\lceil (1/\beta)\log_2(2b/\varepsilon)\big\rceil$. If $\varepsilon \geq 2b$, than the sum in the definition of $n(\varepsilon)$ ranges from 0 to a *negative* value, thus $n(\varepsilon) = 0$ by definition of sum over an empty set and the result is true with $\kappa_1 = \kappa_2 = \kappa(d^\star) = 0$. Assume then that $\varepsilon < 2b$ so that such sum is not trivially zero. Being $\beta \geq 1$, we can further upper bound $n(\varepsilon)$ by

$$4^d\, k \sum_{i=0}^{i(\varepsilon)-1} \lim_{\delta \to 1^-} \mathcal{N}\Big(\big\{|f-a| \leq 2b\, 2^{-i}\big\},\ \delta\, 2^{-i}\Big)\,.$$

By Lemma 3, the packing number is at most

$$\left(1 + 4\,\frac{2b}{\delta}\,\mathbb{I}_{2b>\delta}\right)^d \mathcal{N}\Big(\big\{|f-a| \leq 2b\, 2^{-i}\big\},\ 2b\, 2^{-i}\Big)\,.$$

Taking the limit for $\delta \to 1^-$, the first term becomes $\big(1 + 8b\,\mathbb{I}_{2b\geq 1}\big)^d$, while our NLS assumption (19) implies that the packing number is smaller than, or equal to

$$\mathbb{I}_{2b2^{-i}\geq 1} + C^\star\left(\frac{1}{2b\,2^{-i}}\right)^{d^\star} \mathbb{I}_{2b2^{-i}<1} \tag{21}$$

A direct computation shows that the sum over i of the first term in (21) is

$$\sum_{i=0}^{i(\varepsilon)-1} \mathbb{I}_{2b2^{-i}\geq 1} \leq \log_2(4b)\,\mathbb{I}_{2b\geq 1}\,. \tag{22}$$

For the sum over i of second term in (21), we upper bound the indicator function $\mathbb{I}_{2b2^{-i}<1}$ with 1 for all i and study separately the two cases $d^\star = 0$ and $d^\star > 0$. If $d^\star = 0$, then, by definition of $i(\varepsilon)$,

$$\sum_{i=0}^{i(\varepsilon)-1} \big(2^{d^\star}\big)^i = i(\varepsilon) \leq \log_2\left(\frac{1}{\varepsilon^{1/\beta}}\right) + \log_2\big(2\,(2b)^{1/\beta}\big)\,.$$

Hence, the result follows by defining the additive and multiplicative terms κ_1 and κ_2, respectively, by $\kappa' \, k \left(\log_2(4b)\, \mathbb{I}_{2b\geq 1} + C^\star \log_2\left(2\,(2b)^{1/\beta}\right)\right)$ and $\kappa' \, k \, \frac{C^\star}{(2b)^{d^\star}}$, where $\kappa' := \left(4 + 32\, b\, \mathbb{I}_{2b\geq 1}\right)^d$.

If on the other hand, $d^\star > 0$, recognizing the geometric sum below, we have, by definition of $i(\varepsilon)$

$$\sum_{i=0}^{i(\varepsilon)-1} \left(2^{d^\star}\right)^i = \frac{\left(2^{d^\star}\right)^{i(\varepsilon)} - 1}{2^{d^\star} - 1} \leq \frac{2^{d^\star}\,(2\,b)^{d^\star/\beta}}{2^{d^\star} - 1} \frac{1}{\varepsilon^{d^\star/\beta}} \,.$$

Thus, if $2b < 1$ or if simultaneously $2b \geq 1$ and $\varepsilon \leq 1/\left(\log_2(4b)\right)^{\beta/d^\star}$ —so that the term $\log_2(4b)\, \mathbb{I}_{2b\geq 1}$ in (22) can be upper bounded by $1/\varepsilon^{d^\star/\beta}\, \mathbb{I}_{2b\geq 1}$— the result follows by defining $\kappa(d^\star)$ as

$$\left(4 + 32\, b\, \mathbb{I}_{2b\geq 1}\right)^d k \, \frac{C^\star}{(2b)^{d^\star}} \left(\mathbb{I}_{2b\geq 1} + \frac{2^{d^\star}\,(2\,b)^{d^\star/\beta}}{2^{d^\star} - 1}\right) .$$

Finally, we consider the case in which $2b \geq 1$ and $\varepsilon > 1/\left(\log_2(4b)\right)^{\beta/d^\star}$. In this simpler instance, we upper bound $n(\varepsilon)$ as in the proofs of Theorems 3 and 4. Look back at Equation (20). Upper-bounding, for any $\delta \in (0,1)$ and all $i \geq 0$,

$$\mathcal{N}\Big(\left\{|f-a| \leq 2\,b\,2^{-\beta i}\right\},\ \delta\, 2^{-i}\Big) \leq \mathcal{N}\left([0,1]^d,\ \delta\, 2^{-i}\right) \overset{(\dagger)}{\leq} \left(2^i/\delta + 1\right)^d \leq (2/\delta)^d\, 2^{di}$$

(for completeness, we include a proof of the known upper bound (†) in Section B, Lemma 4) and recognizing the geometric sum below, we have

$$\begin{aligned} n(\varepsilon) &\leq 8^d\, k \sum_{i=0}^{\lceil (1/\beta) \log_2(2b/\varepsilon)\rceil - 1} \left(2^d\right)^i \\ &= 8^d k \frac{2^{d\lceil (1/\beta)\log_2(2b/\varepsilon)\rceil} - 1}{2^d - 1} \leq 8^d k \frac{2^{d((1/\beta)\log_2(2b/\varepsilon)+1)}}{2^d - \left(2^d/2\right)} \\ &= 2\, 8^d\, k\, (2\,b)^{d/\beta} \,\frac{1}{\varepsilon^{d/\beta}} \,. \end{aligned}$$

Finally, using the assumption $\varepsilon > 1/\left(\log_2(4b)\right)^{\beta/d^\star}$, we can upper bound the term $1/\varepsilon^{d/\beta}$ with

$$\begin{aligned} \frac{1}{\varepsilon^{d/\beta}} &= \left(\frac{1}{\varepsilon}\right)^{(d-d^\star)/\beta} \frac{1}{\varepsilon^{d^\star/\beta}} \\ &< \left(\left(\log_2(4b)\right)^{\beta/d^\star}\right)^{(d-d^\star)/\beta} \frac{1}{\varepsilon^{d^\star/\beta}} \\ &= \left(\log_2(4b)\right)^{(d-d^\star)/d^\star} \frac{1}{\varepsilon^{d^\star/\beta}} \,, \end{aligned}$$

and the results follows after defining the constant $\kappa(d^\star) := 2\, 8^d\, k\, (2\,b)^{d/\beta} \left(\log_2(4b)\right)^{(d-d^\star)/d^\star}$. □

The previous result has the following immediate consequence for BAG algorithms. Recall from the comment before Theorem 6 that typical level sets have NLS dimension $d^\star \geq d-1$.

Corollary (Corollary 1). *Consider the BAG algorithm (Algorithm 4) run with input a, c_1, γ_1. Let $f\colon [0,1]^d \to \mathbb{R}$ be an arbitrary (c_1,γ_1)-gradient-Hölder function with level set $\{f=a\} \neq \varnothing$ and let $d^\star \in [0,d]$ be a NLS dimension of $\{f=a\}$ (Definition 9). Fix any accuracy $\varepsilon > 0$. Then, for all $n > m(\varepsilon)$, the output S_n returned after the n-th query is an ε-approximation of $\{f=a\}$, where*

$$m(\varepsilon) := \begin{cases} \kappa_1 + \kappa_2 \log_2\left(\dfrac{1}{\varepsilon^{1/(1+\gamma_1)}}\right) & \text{if } d^\star = 0\,, \\ \kappa(d^\star) \dfrac{1}{\varepsilon^{d^\star/(1+\gamma_1)}} & \text{if } d^\star > 0\,, \end{cases}$$

for $\kappa_1, \kappa_2, \kappa(d^\star) \geq 0$ independent of ε, that depend exponentially on the dimension d.

Proof. The result follows immediately from Corollary 2 and Lemma 1. □

The two previous corollaries suggest a general method for solving the level set approximation problem for a given class $\mathcal{F}$, obtaining bounds that are slightly more refined than the worst-case ones that we saw in Sections 4 and 5. First, determine a family of approximators that accurately approximate the functions in $\mathcal{F}$. Second, obtain for the resulting choice of BA algorithm a sample complexity bound in terms of packing numbers of inflated level sets (as in Theorem 2). Third, find a NLS dimension of an arbitrary $f \in \mathcal{F}$. Importantly, both steps one and three of this process are decoupled from the task of determining approximations of level sets, and as such, they can be investigated independently. For step two, we can simply plug in Theorem 2.

In the next section, we will discuss the notable convex case, in which the estimation of the NLS dimension is non-trivial. As it turns out, this also leads to a rate-optimal sample complexity for BA algorithms.

F.2 Upper Bound for Convex gradient-Hölder Functions

In this section we show a non trivial application of the theory presented so far. We will prove that our BAG algorithm is rate-optimal for approximating the level set of convex gradient-Lipschitz functions.

For the sake of simplicity, we will focus on the approximation of what we call *proper* level sets (Definition 10, below). Informally, a level set is proper if it is non-empty, bounded away from the set of minimizers (where the problem collapses into a simpler, standard minimization problem) and it is not cropped by the boundary of $[0,1]^d$.

Definition 10 (Proper level sets)**.** Fix any level $a \in \mathbb{R}$, a function $f\colon [0,1]^d \to \mathbb{R}$, and a margin $\Delta > 0$. We say that $\{f = a\}$ is a Δ-*proper* level set (for f), if $\{f = a\} \neq \varnothing$ and

$$\min_{\boldsymbol{x}\in[0,1]^d} f(\boldsymbol{x}) + \Delta \le a \le \min_{\boldsymbol{x}\in\partial[0,1]^d} f(\boldsymbol{x}) \;,$$

where we denoted by $\partial[0,1]^d$ the boundary of $[0,1]^d$. When we need not explicitly refer to the margin Δ, we simply say that $\{f = a\}$ is a *proper* level set.

In this section, we present an upper bound on the number of samples that our BAG algorithm needs in order to guarantee that its output is an approximation of the target level set of a convex gradient-Hölder function. As we discussed in Section F, now that we established a method on how to get these types of results, we only need to determine a NLS dimension $d^\star$ (Definition 9) of the level set of an arbitrary convex gradient-Hölder function. The following results shows that $d^\star = d-1$.

Proposition 5. *Fix any level $a \in \mathbb{R}$, two Hölder constants $c > 0, \gamma \in (0,1]$, and an arbitrary convex (c,γ)-Hölder function $f\colon [0,1]^d \to \mathbb{R}$ with proper level set $\{f = a\}$. Then, there exists a constant $C^\star > 0$ such that*

$$\forall r \in (0,1) \;,\; \mathcal{N}\Big(\big\{|f-a| \le r\big\},\; r\Big) \le C^\star \left(\frac{1}{r}\right)^{d-1} \;.$$

Proof. Let $\Delta > 0$ be a margin such that $\{f = a\}$ is Δ-proper. Fix any $r \in (0,1)$. If $r > \Delta/2$, we can simply apply Lemma 5 in Section B and use the lower bound on r to obtain

$$\mathcal{N}\Big(\big\{|f-a| \le r\big\},\; r\Big) \le 2^d \left(\frac{1}{r}\right)^d \le \frac{2^{d+1}}{\Delta} \left(\frac{1}{r}\right)^{d-1} \;.$$

Hence, without loss of generality, we can (and do) assume that $r \in (0, \Delta/2)$. In the following, we denote by $\mathcal{S}^{d-1}$ the $(d-1)$-dimensional unit sphere $\big\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 \le 1\big\}$ with respect to the Euclidean norm $\|\cdot\|_2$. Let $\boldsymbol{x}^\star$ be a minimizer of f. Note that, being $\{f = a\}$ a proper level set (Definition 10), we have that $f(\boldsymbol{x}^\star) < a \le \min_{\boldsymbol{x}\in\partial[0,1]^d} f(\boldsymbol{x})$, therefore $\boldsymbol{x}^\star$ belongs to the interior $(0,1)^d$ of $[0,1]^d$. Now, for each $\boldsymbol{z} \in \mathcal{S}^{d-1}$, let $\boldsymbol{p}_{\boldsymbol{z}}$ be the unique element of $\partial[0,1]^d$ such that $(\boldsymbol{p}_{\boldsymbol{z}} - \boldsymbol{x}^\star)/\|\boldsymbol{p}_{\boldsymbol{z}} - \boldsymbol{x}_m\|_2 = \boldsymbol{z}$ (Figure 3) and define

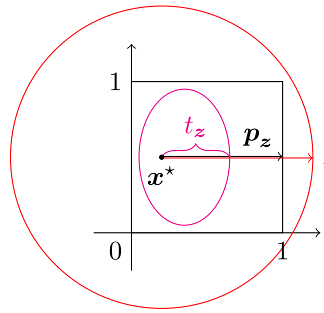

Figure 3: In black, the unit hypercube $[0,1]^d$; in red, the unit sphere centered at the minimizer $\boldsymbol{x}^\star$; in magenta, the level set $\{f = a\}$.

the convex univariate function

$$f_{\boldsymbol{z}} \colon \big[0, \|\boldsymbol{x}^\star - \boldsymbol{p}_{\boldsymbol{z}}\|_2\big] \to \mathbb{R}$$
$$t \mapsto f(\boldsymbol{x}^\star + t\boldsymbol{z}) \ .$$

Being $\{f = a\}$ a proper level set, for all $\boldsymbol{z} \in \mathcal{S}^{d-1}$, the function $f_{\boldsymbol{z}}$ satisfies

$$\min_{\boldsymbol{x} \in [0,1]^d} f(\boldsymbol{x}) = f_{\boldsymbol{z}}(0) < a \le \min_{\boldsymbol{x} \in \partial [0,1]^d} f(\boldsymbol{x}) \le f_{\boldsymbol{z}}\big(\|\boldsymbol{x}^\star - \boldsymbol{p}_{\boldsymbol{z}}\|_2\big) \ .$$

Thus, for each $\boldsymbol{z} \in \mathcal{S}^{d-1}$, by the convexity and continuity of $f_{\boldsymbol{z}}$, there exists a unique value $t_{\boldsymbol{z}} \in \big[0, \|\boldsymbol{x}^\star - \boldsymbol{p}_{\boldsymbol{z}}\|_2\big]$ such that $f_{\boldsymbol{z}}(t_{\boldsymbol{z}}) = a$ (Figure 3), which we use to define the following function on the unit sphere

$$s \colon \mathcal{S}^{d-1} \to \mathbb{R}$$
$$\boldsymbol{z} \mapsto s(\boldsymbol{z}) := t_{\boldsymbol{z}} \ .$$

In words, $t_{\boldsymbol{z}}$ is the distance between the minimizer $\boldsymbol{x}^\star$ and the level set $\{f = a\}$ in the direction of $\boldsymbol{z}$. We show now that s is Lipschitz with respect to the geodesic distance θ on $\mathcal{S}^{d-1}$, i.e., that there exists a constant $\ell > 0$ such that, for all $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathcal{S}^{d-1}$,

$$\big|s(\boldsymbol{z}_1) - s(\boldsymbol{z}_2)\big| \le \ell \, \theta(\boldsymbol{z}_1, \boldsymbol{z}_2) \ ,$$

where $\theta(\boldsymbol{z}_1, \boldsymbol{z}_2) = \arccos\big(\langle \boldsymbol{z}_1, \boldsymbol{z}_2 \rangle\big)$ is the angle between the two unit vectors $\boldsymbol{z}_1, \boldsymbol{z}_2$. Fix two arbitrary $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathcal{S}^{d-1}$ with geodesic distance $\theta := \theta(\boldsymbol{z}_1, \boldsymbol{z}_2) \in (0, \pi]$. If $\theta \ge \pi/6$, we have

$$\frac{\big|s(\boldsymbol{z}_1) - s(\boldsymbol{z}_2)\big|}{\theta} \le \frac{6\sqrt{d}}{\pi}.$$

Assume now that $\theta < \pi/6$. Consider the two-dimensional plane containing the triangle with vertices $\boldsymbol{x}^\star$, $\boldsymbol{v}_1 := \boldsymbol{x}^\star + s(\boldsymbol{z}_1)\boldsymbol{z}_1$ and $\boldsymbol{v}_2 := \boldsymbol{x}^\star + s(\boldsymbol{z}_2)\boldsymbol{z}_2$ (note that the three points are not aligned). Let $\boldsymbol{v}$ be the orthogonal projection of $\boldsymbol{x}^\star$ on the line containing $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Assume first that $\boldsymbol{v}$ belongs to the segment $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ (Figure 4, left). Then, the function

$$g_{\boldsymbol{v}_1, \boldsymbol{v}_2} \colon \mathbb{R} \to [0, +\infty)$$
$$t \mapsto g_{\boldsymbol{v}_1, \boldsymbol{v}_2}(t) := \Big\|\boldsymbol{x}^\star - \big(\boldsymbol{v}_1 + t(\boldsymbol{v}_2 - \boldsymbol{v}_1)\big)\Big\|_2^2$$

has its unique minimum at some $t^\star \in [0,1]$. For all $t \in \mathbb{R}$, we have

$$\begin{aligned} g_{\boldsymbol{v}_1, \boldsymbol{v}_2}(t) &= \big\|(1-t)(\boldsymbol{v}_1 - \boldsymbol{x}^\star) + t(\boldsymbol{v}_2 - \boldsymbol{x}^\star)\big\|_2^2 \\ &= (1-t)^2 s(\boldsymbol{z}_1)^2 + t^2 s(\boldsymbol{z}_2)^2 + 2t(1-t)s(\boldsymbol{z}_1)s(\boldsymbol{z}_2)\cos(\theta) \\ &= t^2\big(s(\boldsymbol{z}_1)^2 + s(\boldsymbol{z}_2)^2 - 2s(\boldsymbol{z}_1)s(\boldsymbol{z}_2)\cos(\theta)\big) + t\big(2s(\boldsymbol{z}_1)s(\boldsymbol{z}_2)\cos(\theta) - 2s(\boldsymbol{z}_1)^2\big) + s(\boldsymbol{z}_1)^2 \ . \end{aligned}$$

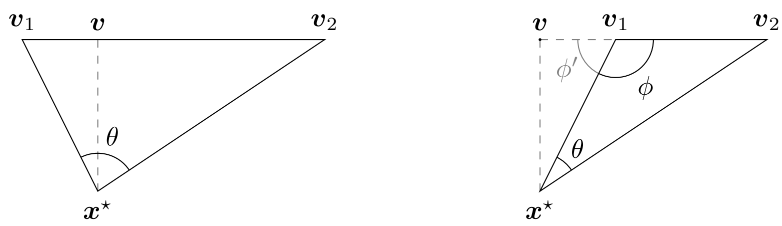

Figure 4: In the left (resp., right) picture, $\boldsymbol{v}$ belongs (resp., does not belong) to the segment $[\boldsymbol{v}_1, \boldsymbol{v}_2]$.

The derivative of this function is given, for all $t \in \mathbb{R}$, by

$$g'_{\boldsymbol{v}_1,\boldsymbol{v}_2}(t) = 2t\big(s(\boldsymbol{z}_1)^2 + s(\boldsymbol{z}_2)^2 - 2s(\boldsymbol{z}_1)s(\boldsymbol{z}_2)\cos(\theta)\big) + \big(2s(\boldsymbol{z}_1)s(\boldsymbol{z}_2)\cos(\theta) - 2s(\boldsymbol{z}_1)^2\big) \ .$$

Hence we have

$$0 \le t^\star = \frac{2s(\boldsymbol{z}_1)^2 - 2s(\boldsymbol{z}_1)s(\boldsymbol{z}_2)\cos(\theta)}{2\big(s(\boldsymbol{z}_1)^2 + s(\boldsymbol{z}_2)^2 - 2s(\boldsymbol{z}_1)s(\boldsymbol{z}_2)\cos(\theta)\big)} .$$

Since the above denominator is strictly positive, we obtain

$$2s(\boldsymbol{z}_1)^2 \ge 2s(\boldsymbol{z}_1)s(\boldsymbol{z}_2)\cos(\theta) \ ,$$

thus, being $s(\boldsymbol{z}_1)$ and $s(\boldsymbol{z}_2)$ also strictly positive,

$$\cos(\theta) \le 1 - \frac{s(\boldsymbol{z}_2) - s(\boldsymbol{z}_1)}{s(\boldsymbol{z}_2)}$$

and in turn, since $s(\boldsymbol{z}) \le \sqrt{d}$ for all $\boldsymbol{z} \in \mathcal{S}^{d-1}$,

$$s(\boldsymbol{z}_2) - s(\boldsymbol{z}_1) \le \sqrt{d}\big(1 - \cos(\theta)\big) \ .$$

Being $\theta > 0$, we have $1 - \cos(\theta) \le \theta$ and thus

$$\frac{s(\boldsymbol{z}_2) - s(\boldsymbol{z}_1)}{\theta} \le \sqrt{d} \ .$$

Swapping the roles of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ (i.e., considering the function $g_{\boldsymbol{v}_2,\boldsymbol{v}_1}$) we obtain similarly

$$\frac{s(\boldsymbol{z}_1) - s(\boldsymbol{z}_2)}{\theta} \le \sqrt{d} \ .$$

Hence, when $\boldsymbol{v}$ belongs to the segment $[\boldsymbol{v}_1, \boldsymbol{v}_2]$, we obtained

$$\frac{\big|s(\boldsymbol{z}_1) - s(\boldsymbol{z}_2)\big|}{\theta} \le \sqrt{d} \ .$$

Consider now the last case where $\boldsymbol{v}$ does not belong to the segment $[\boldsymbol{v}_1, \boldsymbol{v}_2]$ (Figure 4, right). Without loss of generality, we can (and do) assume that $s(\boldsymbol{z}_2) > s(\boldsymbol{z}_1)$, and thus that $\boldsymbol{v}$ is closer to $\boldsymbol{v}_1$ than to $\boldsymbol{v}_2$. By convexity of f on the line containing $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, we have $f(\boldsymbol{v}) \ge a$. Using the fact that the level set $\{f = a\}$ is Δ-proper and the (c, γ)-Hölderness of f, we get

$$\Delta \le a - \min_{\boldsymbol{x} \in [0,1]^d} f(\boldsymbol{x}) \le f(\boldsymbol{v}) - f(\boldsymbol{x}^\star) \le c\|\boldsymbol{v} - \boldsymbol{x}^\star\|_\infty^\gamma \le c\|\boldsymbol{v} - \boldsymbol{x}^\star\|_2^\gamma \ ,$$

which in turn implies

$$\|\boldsymbol{v} - \boldsymbol{x}^\star\|_2 \ge \left(\frac{\Delta}{c}\right)^{1/\gamma} . \tag{23}$$

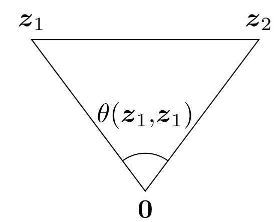

Figure 5: The isosceles triangle $\boldsymbol{z}_1, \mathbf{0}, \boldsymbol{z}_2$.

Let ϕ be the angle between $\boldsymbol{x}^\star - \boldsymbol{v}_1$ and $\boldsymbol{v}_2 - \boldsymbol{v}_1$. Applying the sine rule to the triangle $\boldsymbol{x}^\star, \boldsymbol{v}_1, \boldsymbol{v}_2$, we obtain

$$\frac{\sin(\theta)}{\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_2} = \frac{\sin(\phi)}{s(\boldsymbol{z}_2)}$$

and thus

$$\sin(\phi) = \frac{s(\boldsymbol{z}_2)\sin(\theta)}{\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_2} \; . \tag{24}$$

Let $\phi' = \pi - \phi$ be the angle between $\boldsymbol{x}^\star - \boldsymbol{v}_1$ and $\boldsymbol{v} - \boldsymbol{v}_1$. Note that the angle between $\boldsymbol{x}^\star - \boldsymbol{v}$ and $\boldsymbol{v}_1 - \boldsymbol{v}$ is $\pi/2$, being $\boldsymbol{v}$ the orthogonal projection of $\boldsymbol{x}^\star$ on the line containing $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Hence, applying the sine rule to the triangle $\boldsymbol{x}^\star, \boldsymbol{v}_1, \boldsymbol{v}$ we obtain

$$\frac{\sin(\phi')}{\|\boldsymbol{v} - \boldsymbol{x}^\star\|_2} = \frac{\sin(\pi/2)}{s(\boldsymbol{z}_1)}$$

and thus

$$\|\boldsymbol{v} - \boldsymbol{x}^\star\|_2 = s(\boldsymbol{z}_1)\sin(\phi) \; . \tag{25}$$

From (23), (25), and (24), we obtain

$$\frac{s(\boldsymbol{z}_1)s(\boldsymbol{z}_2)\sin(\theta)}{\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_2} \geq \left(\frac{\Delta}{c}\right)^{1/\gamma} .$$

The triangle inequality yields

$$\big|s(\boldsymbol{z}_1) - s(\boldsymbol{z}_2)\big| = \big|\|\boldsymbol{v}_1 - \boldsymbol{x}^\star\|_2 - \|\boldsymbol{v}_2 - \boldsymbol{x}^\star\|_2\big| \leq \big\|(\boldsymbol{v}_1 - \boldsymbol{x}^\star) - (\boldsymbol{v}_2 - \boldsymbol{x}^\star)\big\|_2 = \|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_2$$

and thus

$$\big|s(\boldsymbol{z}_1) - s(\boldsymbol{z}_2)\big| \leq \left(\frac{c}{\Delta}\right)^{1/\gamma} d\sin(\theta) \leq \left(\frac{c}{\Delta}\right)^{1/\gamma} d\theta \; ,$$

where we used again $s(\boldsymbol{z}) \leq d^{1/2}$ for any $\boldsymbol{z} \in \mathcal{S}^{d-1}$. Putting everything together, we have shown that

$$\frac{\big|s(\boldsymbol{z}_1) - s(\boldsymbol{z}_2)\big|}{\theta} \leq \max\left(\frac{6}{\pi}\sqrt{d},\; \left(\frac{c}{\Delta}\right)^{1/\gamma} d\right) := \ell \; , \tag{26}$$

for all $\theta \in (0, \pi]$, i.e., that s is ℓ-Lipschitz on $\mathcal{S}^{d-1}$ with respect to the geodesic distance.

Consider now a covering of $\mathcal{S}^{d-1}$ with respect to the geodesic distance, with radius βr, where $\beta \in (0, 1]$ will be selected later. This is a set of points $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n \in \mathcal{S}^{d-1}$ such that the union of all the balls (with respect to the geodesic distance θ) with radius βr centered at these points contains the whole $\mathcal{S}^{d-1}$. We show now how such a covering can be taken using order of $1/r^{d-1}$ points. Fix any two distinct $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathcal{S}^{d-1}$ with geodesic distance $\theta(\boldsymbol{z}_1, \boldsymbol{z}_2) \in (0, \pi/2]$ and consider the isosceles triangle $\boldsymbol{z}_1, \mathbf{0}, \boldsymbol{z}_2$ with angles $\angle(\boldsymbol{z}_1\,\mathbf{0}\,\boldsymbol{z}_2) = \theta(\boldsymbol{z}_1, \boldsymbol{z}_2)$ and $\angle(\mathbf{0}\,\boldsymbol{z}_2\,\boldsymbol{z}_1) = \angle(\boldsymbol{z}_2\,\boldsymbol{z}_1\,\mathbf{0}) = \big(\pi - \theta(\boldsymbol{z}_1, \boldsymbol{z}_2)\big)/2 = \pi/2 - \theta(\boldsymbol{z}_1, \boldsymbol{z}_2)/2$ (Figure 5). The sine rule yields

$$\frac{\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|_2}{\sin\big(\theta(\boldsymbol{z}_1, \boldsymbol{z}_2)\big)} = \frac{1}{\cos\big(\theta(\boldsymbol{z}_1, \boldsymbol{z}_2)/2\big)}$$

or, equivalently stated,

$$\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|_2 = \frac{\sin\big(\theta(\boldsymbol{z}_1, \boldsymbol{z}_2)\big)}{\cos\big(\theta(\boldsymbol{z}_1, \boldsymbol{z}_2)/2\big)} \ .$$

Using the fact that $\sin(x) \geq (2/\pi)x$, for all $x \in [0, \pi/2]$, the equality above gives

$$\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|_2 \geq \sin\big(\theta(\boldsymbol{z}_1, \boldsymbol{z}_2)\big) \geq \frac{2}{\pi}\, \theta(\boldsymbol{z}_1, \boldsymbol{z}_2) \ .$$

Therefore, if $x \leq \pi/2$, each ball with center $\boldsymbol{c}$ radius $(2/\pi)\rho$ with respect to the Euclidean distance is included in the corresponding ball with center $\boldsymbol{c}$ and radius ρ with respect to the geodesic distance. Thus, being $r < 1 \leq \pi/2$, in order to cover $\mathcal{S}^{d-1}$ with balls with radius βr with respect to the geodesic distance, it is enough to cover $\mathcal{S}^{d-1}$ with balls with radius $(2/\pi)\beta r$ with respect to the Euclidean distance. Moreover, since for any two points $\boldsymbol{x}, \boldsymbol{y} \in \partial[-1,1]^d$ on the boundary of the hypercube $[-1,1]^d$, their Euclidean distance $\|\boldsymbol{x} - \boldsymbol{y}\|_2$ is larger than the Euclidean distance $\big\|\boldsymbol{x}/\|\boldsymbol{x}\|_2 - \boldsymbol{y}/\|\boldsymbol{y}\|_2\big\|_2$ between their projections on the unit sphere, and since any point in the unit sphere can be reached this way, in order to cover the unit sphere with balls with radius $(2/\pi)\beta r$ with respect to the Euclidean distance it is sufficient to cover the boundary $\partial[-1,1]^d$ of $[-1,1]^d$ with balls with radius $(2/\pi)\beta r$ with respect to the Euclidean distance. This is easy to do, as each one of the $2d$ faces $\big\{[-1,1] \times \ldots \times [-1,1] \times \{-1,1\} \times [-1,1] \times \ldots \times [-1,1]\big\}$ of $\partial[-1,1]^d$ can be covered with the same number of balls of radius $(2/\pi)\beta r$ with respect to the $(d-1)$-dimensional Euclidean distance that cover the hypercube $[-1,1]^{d-1}$. This can be done, e.g., by taking a uniform grid of $(2/\pi)\beta r$-spaced points. Projecting these points onto $\mathcal{S}^{d-1}$ gives a covering $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ of $\mathcal{S}^{d-1}$ with respect to the geodesic distance, with radius βr, and with a number of points n that is at most

$$n \leq 2d\left(1 + \left\lceil \frac{\pi}{2\beta r} \right\rceil\right)^{d-1} \leq 2d\left(2 + \frac{\pi}{2\beta r}\right)^{d-1} \leq 2d\left(\frac{3}{2}\pi\right)^{d-1}\left(\frac{1}{\beta r}\right)^{d-1} \ . \tag{27}$$

Fix this covering $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$. Fix also an arbitrary $\boldsymbol{x} \in \big\{|f - a| \leq r|\big\}$. Note that, being $r \leq \Delta/2$ and $\{f = a\}$ a Δ-proper level set, then the minimizer $\boldsymbol{x}^\star$ cannot belong to the set $\big\{|f - a| \leq r|\big\}$, hence $\boldsymbol{x} \neq \boldsymbol{x}^\star$. Let $\boldsymbol{z} = (\boldsymbol{x} - \boldsymbol{x}^\star)/\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2$. Similarly as before, define for all $t \in \big[0, \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2\big]$, the function $f_{\boldsymbol{z}}(t) := f(\boldsymbol{x}^\star + t\boldsymbol{z})$. Then $f_{\boldsymbol{z}}$ is convex, $f_{\boldsymbol{z}}(0) = f(\boldsymbol{x}^\star)$, $f_{\boldsymbol{z}}\big(s(\boldsymbol{z})\big) = a$ and $\big|f_{\boldsymbol{z}}\big(\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2\big) - a\big| \leq r$. If $f_{\boldsymbol{z}}\big(\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2\big) < a$, by convexity, we have $s(\boldsymbol{z}) > \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2$, hence

$$\frac{a - r - f(\boldsymbol{x}^\star)}{\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2} \leq \frac{f_{\boldsymbol{z}}\big(\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2\big) - f_{\boldsymbol{z}}(0)}{\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 - 0} \leq \frac{a - f_{\boldsymbol{z}}\big(\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2\big)}{s(\boldsymbol{z}) - \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2} \leq \frac{r}{s(\boldsymbol{z}) - \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2}$$

and recalling that $r \leq \Delta/2$ so that $a - r - f(\boldsymbol{x}^\star) \geq \Delta/2 > 0$, we have

$$s(\boldsymbol{z}) - \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 \leq r \frac{\sqrt{d}}{a - r - f(\boldsymbol{x}^\star)} \leq \left(2\frac{\sqrt{d}}{\Delta}\right) r \ ,$$

where we used $\|x - x_m\|_2 \leq \sqrt{d}$. If $f_{\boldsymbol{z}}\big(\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2\big) \geq a$, proceed similarly. By convexity of $f_{\boldsymbol{z}}$ we have $s(\boldsymbol{z}) \leq \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2$. If $s(\boldsymbol{z}) = \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2$, then trivially $\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 - s(\boldsymbol{z}) = 0 \leq \big(2\sqrt{d}/\Delta\big)r$. If on the other hand, $s(\boldsymbol{z}) < \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2$, using the convexity of $f_{\boldsymbol{z}}$ once again, we get

$$\frac{a - f_{\boldsymbol{z}}(0)}{s(\boldsymbol{z}) - 0} \leq \frac{f_{\boldsymbol{z}}\big(\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2\big) - a}{\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 - s(\boldsymbol{z})} \leq \frac{r}{\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 - s(\boldsymbol{z})}$$

and using $a - f(\boldsymbol{x}^\star) \geq \Delta > 0$ and $s(\boldsymbol{z}) \leq \sqrt{d}$, yields

$$\|\boldsymbol{x} - \boldsymbol{x}^\star\|_2 - s(\boldsymbol{z}) \leq \left(\frac{\sqrt{d}}{\Delta}\right) r \leq \left(2\frac{\sqrt{d}}{\Delta}\right) r \ .$$

Thus we proved that

$$\big|s(\boldsymbol{z}) - \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2\big| \leq \left(2\frac{\sqrt{d}}{\Delta}\right) r \ . \tag{28}$$

Furthermore, there exists $i \in \{1,\dots,n\}$ such the geodesic distance of $\boldsymbol{z}_i$ and $\boldsymbol{z}$ is smaller than or equal to βr. Therefore we have, from (28), and the ℓ-Lipschitzness of the function r with respect to the geodesic distance,

$$\left|s(\boldsymbol{z}_i) - \|\boldsymbol{x}-\boldsymbol{x}^\star\|_2\right| \le \left|s(\boldsymbol{z}_i) - s(\boldsymbol{z})\right| + \left|s(\boldsymbol{z}) - \|\boldsymbol{x}-\boldsymbol{x}^\star\|_2\right| \le (\ell\beta)\, r + \left(2\frac{\sqrt{d}}{\Delta}\right) r \; .$$

Hence, with $\gamma > 0$ to be chosen later, there exists

$$\boldsymbol{x}_i' := \boldsymbol{x}^\star + s(\boldsymbol{z}_i)\boldsymbol{z}_i + k\gamma r \boldsymbol{z}_i \; ,$$

with $k \in \mathbb{Z}$ such that

$$|k| \le \frac{\ell\beta + \frac{2\sqrt{d}}{\Delta}}{\gamma}$$

and with

$$\left| \|\boldsymbol{x}_i' - \boldsymbol{x}^\star\|_2 - \|\boldsymbol{x}-\boldsymbol{x}^\star\|_2 \right| \le \gamma r \; . \tag{29}$$

This is obtained by covering the segment $[-\ell\beta r - 2rd^{1/2}/\Delta,\ \ell\beta r + 2rd^{1/2}/\Delta]$ with points with equidistance γr. Then, we obtain

$$\begin{aligned}
\|\boldsymbol{x}-\boldsymbol{x}_i'\|_\infty &\le \|\boldsymbol{x}-\boldsymbol{x}_i'\|_2 \\
&= \left\| \left(\boldsymbol{x}^\star + \|\boldsymbol{x}-\boldsymbol{x}^\star\|_2\, \boldsymbol{z}\right) - \left(\boldsymbol{x}^\star + \|\boldsymbol{x}_i'-\boldsymbol{x}^\star\|_2\, \boldsymbol{z}_i\right) \right\|_2 \\
&= \left\| \|\boldsymbol{x}-\boldsymbol{x}^\star\|_2\, \boldsymbol{z} - \|\boldsymbol{x}_i'-\boldsymbol{x}^\star\|_2\, \boldsymbol{z}_i \right\|_2 \\
&= \Big\| \|\boldsymbol{x}-\boldsymbol{x}^\star\|_2 (\boldsymbol{z}-\boldsymbol{z}_i) \\
&\qquad\qquad - \left(\|\boldsymbol{x}_i'-\boldsymbol{x}^\star\|_2 - \|\boldsymbol{x}-\boldsymbol{x}^\star\|_2\right)\boldsymbol{z}_i \Big\|_2 \\
&\le \left| \|\boldsymbol{x}-\boldsymbol{x}^\star\|_2 - \|\boldsymbol{x}_i'-\boldsymbol{x}^\star\|_2 \right| + \sqrt{d}\|\boldsymbol{z}-\boldsymbol{z}_i\|_2 \\
&\le (\gamma + \sqrt{d}\beta) r \; ,
\end{aligned}$$

from (29). Hence, with

$$n' \le 2d\left(\frac{3}{2}\pi\right)^{d-1}\left(\frac{1}{\beta r}\right)^{d-1}\left(1 + 2\frac{\ell\beta + \frac{2\sqrt{d}}{\Delta}}{\gamma}\right)$$

points, we have obtained a covering of $\{|f-a| \le r\}$ with radius $(\gamma + \sqrt{d}\beta) r$ with respect to the sup-norm $\|\cdot\|_\infty$. Choosing $\beta := 1/(4\sqrt{d})$ and $\gamma := 1/4$ so that that $(\gamma + \sqrt{d}\beta) \le 1/2$, we therefore determined a covering of $\{|f-a| \le r\}$ with radius $r/2$ with respect to the sup-norm consisting of n' elements. Thus, n' is greater than or equal to the smallest cardinality $\mathcal{M}(\{|f-a| \le r\},\ r/2)$ of a covering of $\{|f-a| \le r\}$ with radius $r/2$ with respect to the sup-norm. For a known result relating pickings and coverings (we recall it in (12), Section B), we have

$$\mathcal{M}\Big(\{|f-a| \le r\},\ r/2\Big) \ge \mathcal{N}\Big(\{|f-a| \le r\},\ r\Big) \; ,$$

which concludes the proof. □

Theorem 7. *Consider the BAG algorithm (Algorithm 4) run with input a, c_1, γ_1. Let $f\colon [0,1]^d \to \mathbb{R}$ be an arbitrary convex (c_1,γ_1)-gradient-Hölder function with proper level set $\{f = a\}$. Fix any accuracy $\varepsilon > 0$. Then, for all*

$$n > \kappa \, \frac{1}{\varepsilon^{(d-1)/(1+\gamma_1)}}$$

the output S_n returned after the n-th query is an ε-approximation of $\{f = a\}$, where $\kappa > 0$ is a constant independent of ε that depends exponentially on the dimension d.

Proof. Being f the restriction of a differentiable function defined on on an open set containing $[0,1]^d$, it is Lipschitz on the compact $[0,1]^d$. Thus we can apply Proposition 5 to get a NLS dimension $d^\star = d-1$ for f. The result then follows directly from Corollary 1. □

F.3 Rate-optimal Sample Complexity for Convex Gradient-Lipschitz Functions

Theorem 7 applied to the special case of gradient-Lipschitz functions, states that the BAG algorithm (Algorithm 4) needs order of $1/\varepsilon^{(d-1)/2}$ queries to reliably output an ε-approximation of a gradient-Lipschitz function. The following theorem shows that this rate cannot be improved, i.e., that BAG is rate-optimal (Definition 8) for determining proper level sets of gradient-Lipschitz functions.

Theorem 8. *Fix any level $a \in \mathbb{R}$ and an arbitrary accuracy $\varepsilon > 0$. No deterministic algorithm A can guarantee to output an ε-approximation of any Δ-proper level set $\{f = a\}$ of an arbitrary convex c_1-gradient-Lipschitz functions f with $c_1 \geq 3$ and $\Delta \in (0, 1/4]$, querying less than $\kappa/\varepsilon^{(d-1)/2}$ of their values, where $\kappa > 0$ is a constant independent of ε. This implies in particular that (recall Definition 7),*

$$\inf_A \sup_f \mathfrak{n}(f, A, \varepsilon, a) \geq \kappa \frac{1}{\varepsilon^{(d-1)/2}} ,$$

where the inf *is over all deterministic algorithms A and the* sup *is over all c_1-gradient-Lipschitz functions f with Δ-proper level set $\{f = a\}$, with $c_1 \geq 3$ and $\Delta \in (0, 1/4]$.*

Proof. We will prove the equivalent statement that no algorithm can output a $(\kappa'\varepsilon)$-approximation of any Δ-proper level set $\{f = a\}$ of an arbitrary c_1-gradient-Lipschitz functions f with $c_1 \geq 3$ and $\Delta \in (0, 1/4]$, querying less than $1/\varepsilon^{(d-1)/2}$ of their values, where $\kappa' > 0$ is a constant independent of ε.

Let $\boldsymbol{o} := (1/2, \dots, 1/2) \in [0,1]^d$, $\boldsymbol{o}_1 := (1/2 + 1/(4d)^{1/2}, \dots, 1/2 + 1/(4d)^{1/2}) \in (0,1)^d$, and

$$\begin{aligned} f_0 \colon [0,1]^d &\to \mathbb{R} \\ \boldsymbol{x} &\mapsto f_0(\boldsymbol{x}) := a - \frac{1}{4} + \|\boldsymbol{x} - \boldsymbol{o}\|_2^2 . \end{aligned}$$

Then f_0 is the restriction to $[0,1]^d$ of the differentiable function $\boldsymbol{x} \mapsto a - \frac{1}{4} + \|\boldsymbol{x} - \boldsymbol{o}\|_2^2$ defined on $\mathbb{R}^d$, and it satisfies, for all $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^d$

$$\begin{aligned} \left\|\nabla f_0(\boldsymbol{x}) - \nabla f_0(\boldsymbol{y})\right\|_\infty &= \left\|2(\boldsymbol{x} - \boldsymbol{o}) - 2(\boldsymbol{y} - \boldsymbol{o})\right\|_\infty \\ &= 2\left\|\boldsymbol{x} - \boldsymbol{y}\right\|_\infty , \end{aligned} \tag{30}$$

i.e., it is 2-gradient-Lipschitz. Moreover, f_0 has minimum equal to $a - 1/4$ at $\boldsymbol{o}$ and satisfies $f_0(\boldsymbol{o}_1) = a$. Also, the minimum of f_0 over $\partial[0,1]^d$ is equal to $a + (1/2)^2 - 1/4 = a$. Hence $\{f_0 = a\}$ is a Δ-proper level set, with $\Delta = 1/4$.

Consider an arbitrary deterministic algorithm A applied to the level set $\{f_0 = a\}$ of f_0 and assume that only $n < 1/\varepsilon^{(d-1)/2}$ values are queried before outputting a set S_n.

Let $\mathcal{S}$ be the Euclidean sphere with center $\boldsymbol{o}$ and radius $\|\boldsymbol{o} - \boldsymbol{o}_1\|_2 = 1/2$ (Figure 6). Note that $\{f_0 = a\} = \mathcal{S}$. For any constant $\kappa_1 > 0$ and each point $\boldsymbol{x}_0$ in $\mathcal{S}$, consider the convex cone having origin $\boldsymbol{o}$, and with intersection with $\mathcal{S}$ equal to the geodesic ball on $\mathcal{S}$ with center $\boldsymbol{x}_0$ and radius $\kappa_1 \varepsilon^{1/2}$. Then we can choose κ_1 (small enough) and $\boldsymbol{x}_0$ such that this cone does not contain any points of f_0 queried by the algorithm. Fix such a κ_1. If S_n does not contain $\boldsymbol{x}_0$ then, since $f_0(\boldsymbol{x}_0) = a$, we have shown that $\{f_0 = a\} \not\subseteq S_n$, and the result follows.

Assume now that $\boldsymbol{x}_0 \in S_n$. We will define a function $f_1 \colon [0,1]^d \to \mathbb{R}$ such that the sum $f_0 + f_1$ is convex and 3-gradient-Lipschitz, the level set $\{f_0 + f_1 = a\}$ is Δ-proper, and the algorithm applied to the level set $\{f_0 + f_1 = a\}$ of $f_0 + f_1$ does not return a $(\kappa'\varepsilon)$-approximation of $\{f_0 + f_1 = a\}$. The idea is to carefully design a function f_1 that is non-zero only on the cone that has not been explored by the algorithm. This way, we can make $f_0 + f_1$ a perturbation of f_0 that is not far enough from f_0 so that the algorithm can distinguish the two, but it is different enough so that no $(\kappa'\varepsilon)$-approximation of $\{f_0 = a\}$ can be a $(\kappa'\varepsilon)$-approximation of $\{f_0 + f_1 = a\}$. The subtle part is that by construction, such an f_1 is not convex, but the sum $f_0 + f_1$ has to retain the convexity of f_0.

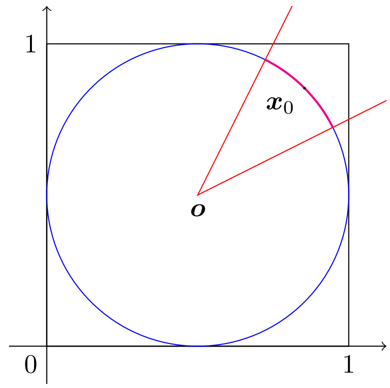

Figure 6: The blue circle is the level set $\mathcal{S}$; the geodesic ball on $\mathcal{S}$ with center $\boldsymbol{x}_0$ and radius $\kappa_1 \varepsilon^{1/2}$ is the arc in magenta and the corresponding cone is in red.

We begin by defining three non-negative auxiliary functions $\phi_1, \phi_2, \phi_3 \colon [0,1] \to \mathbb{R}$, for all $t \in [0,1]^d$, by

$$\phi_1(t) := \begin{cases} t & \text{if } t \in [0, 1/4] \\ 1/4 - (t - 1/4) & \text{if } t \in [1/4, 3/4] \\ -1/4 + (t - 3/4) & \text{if } t \in [3/4, 1] \end{cases},$$

$\phi_2(t) := \int_0^t \mathrm{d}x \int_0^x \phi_1(u)\,\mathrm{d}u$, and $\phi_3(t) := \phi_2(1-t)$. We remark that ϕ_2 is twice differentiable with second derivative ϕ_1. We see that $\phi_2(0) = 0$, $\phi_2'(0) = 0$ and $\phi_2''(0) = 0$. We see that ϕ_2' is strictly positive on $[0,1]$. Hence, $\kappa_2 := \phi_2(1) > 0$. We also see that $\phi_2'(1) = 0$ and $\phi_2''(1) = 0$. Then ϕ_3 is twice differentiable and non-negative on $[0,1]$ and satisfies $\phi_3''(0) = 0$, $\phi_3''(1) = 0$, $\phi_3'(0) = 0$, $\phi_3'(1) = 0$, $\phi_3(0) = \kappa_2 > 0$ and $\phi_3(1) = 0$.

We write $B(\boldsymbol{x}, r)$ for the closed Euclidean ball with center $\boldsymbol{x}$ and radius r intersected with $[0,1]^d$. We define the function $f_1 \colon [0,1]^d \to \mathbb{R}$, for all $\boldsymbol{x} \in [0,1]^d$, by

$$f_1(\boldsymbol{x}) := \begin{cases} \beta\varepsilon\, \phi_3\left(\dfrac{\|\boldsymbol{x} - \boldsymbol{x}_0\|_2}{\kappa_3 \varepsilon^{1/2}}\right) & \text{if } \boldsymbol{x} \in B(\boldsymbol{x}_0, \kappa_3\varepsilon^{1/2}) \\ 0 & \text{otherwise}, \end{cases}$$

with $\kappa_3, \beta > 0$ to be selected later. We can find $\kappa_3 > 0$ small enough such that $B(\boldsymbol{x}_0, \kappa_3\varepsilon^{1/2})$ is included in the cone discussed above (recall Figure 6). Fix such a κ_3. Then, f_0 and $f_0 + f_1$ differ only on this cone which is not explored by the algorithm. As a consequence, the algorithm applied to $f_0 + f_1$ returns the same set S_n, which contains $\boldsymbol{x}_0$. Since $f_0(\boldsymbol{x}_0) + f_1(\boldsymbol{x}_0) = a + \beta\varepsilon\kappa_2$, the proof will be completed (letting $\kappa' := \beta\kappa_2/2$) once we show that we can select $\beta > 0$, independently of ε, such that $f_0 + f_1$ is a convex 3-gradient-Lipschitz function with Δ-proper level set $\{f_0 + f_1 = a\}$, where $\Delta = 1/4$.

Because of the above discussed inclusion of the ball in the cone, we have $f_1(\boldsymbol{o}) = 0$. Hence

$$\min_{\boldsymbol{x} \in [0,1]^d} \big(f_0(\boldsymbol{x}) + f_1(\boldsymbol{x})\big) \le f_0(\boldsymbol{o}) + f_1(\boldsymbol{o}) = a - \frac{1}{4} \le a = \min_{\boldsymbol{x} \in \partial[0,1]^d} f_0(\boldsymbol{x}) \le \min_{\boldsymbol{x} \in \partial[0,1]^d} \big(f_0(\boldsymbol{x}) + f_1(\boldsymbol{x})\big),$$

which proves that the level set $\{f_0 + f_1 = a\}$ is Δ-proper, with $\Delta = 1/4$.

By definition of f_1, its gradient is, for $\boldsymbol{x} \in [0,1]^d$,

$$\nabla f_1(\boldsymbol{x}) = \beta\varepsilon\phi_3'\left(\frac{\|\boldsymbol{x} - \boldsymbol{x}_0\|_2}{\kappa_3\varepsilon^{1/2}}\right)\frac{1}{\kappa_3\varepsilon^{1/2}}\frac{\boldsymbol{x} - \boldsymbol{x}_0}{\|\boldsymbol{x} - \boldsymbol{x}_0\|_2}$$

if $\boldsymbol{x} \in B(\boldsymbol{x}_0, \kappa_3\varepsilon^{1/2})$, 0 otherwise. We remark that in the above formula, by convention, $\nabla f_1(\boldsymbol{x}_0) = 0$, which follows from the properties of ϕ_3. Next, we observe that ∇f_1 satisfies

$$\sup_{\substack{\boldsymbol{u},\boldsymbol{v}\in[0,1]^d\\ \boldsymbol{u}\neq\boldsymbol{v}}} \frac{\left\|\nabla f_1(\boldsymbol{u}) - \nabla f_1(\boldsymbol{v})\right\|_2}{\|\boldsymbol{u}-\boldsymbol{v}\|_2} \leq \sup_{\substack{\boldsymbol{u},\boldsymbol{v}\in B(\boldsymbol{x}_0,\kappa_3\varepsilon^{1/2})\\ \boldsymbol{u}\neq\boldsymbol{v}}} \frac{\left\|\nabla f_1(\boldsymbol{u}) - \nabla f_1(\boldsymbol{v})\right\|_2}{\|\boldsymbol{u}-\boldsymbol{v}\|_2} ,$$

Indeed, for $\boldsymbol{u},\boldsymbol{v} \notin B(\boldsymbol{x}_0,\kappa_3\varepsilon^{1/2})$ the gradient difference is zero while for $\boldsymbol{u} \in B(\boldsymbol{x}_0,\kappa_3\varepsilon^{1/2})$ and $\boldsymbol{v} \notin B(\boldsymbol{x}_0,\kappa_3\varepsilon^{1/2})$ the gradient difference is equal to the difference between the gradient at $\boldsymbol{u}$ and the gradient at the intersection of the segment $[u,v]$ and the boundary $\partial B(\boldsymbol{x}_0,\kappa_3\varepsilon^{1/2})$. Hence,

$$\begin{aligned}
&\sup_{\substack{\boldsymbol{u},\boldsymbol{v}\in[0,1]^d\\ \boldsymbol{u}\neq\boldsymbol{v}}} \frac{\left\|\nabla f_1(\boldsymbol{u}) - \nabla f_1(\boldsymbol{v})\right\|_2}{\|\boldsymbol{u}-\boldsymbol{v}\|_2} \\
&\quad\leq \sup_{\substack{\boldsymbol{u},\boldsymbol{v}\in B(\boldsymbol{x}_0,\kappa_3\varepsilon^{1/2})\\ \boldsymbol{u}\neq\boldsymbol{v}}} \frac{\beta\varepsilon}{\|\boldsymbol{u}-\boldsymbol{v}\|_2} \left\| \phi_3'\left(\frac{\|\boldsymbol{u}-\boldsymbol{x}_0\|_2}{\kappa_2\varepsilon^{1/2}}\right) \frac{1}{\kappa_3\varepsilon^{1/2}} \frac{\boldsymbol{u}-\boldsymbol{x}_0}{\|\boldsymbol{u}-\boldsymbol{x}_0\|_2} - \phi_3'\left(\frac{\|\boldsymbol{v}-\boldsymbol{x}_0\|_2}{\kappa_2\varepsilon^{1/2}}\right) \frac{1}{\kappa_3\varepsilon^{1/2}} \frac{\boldsymbol{v}-\boldsymbol{x}_0}{\|\boldsymbol{v}-\boldsymbol{x}_0\|_2} \right\|_2 \\
&\quad= \sup_{\substack{\boldsymbol{u},\boldsymbol{v}\in B(\boldsymbol{0},1)\\ \boldsymbol{u}\neq\boldsymbol{v}}} \frac{\frac{\beta}{\kappa_3^2}\left\| \phi_3'\big(\|\boldsymbol{u}\|_2\big)\frac{\boldsymbol{u}}{\|\boldsymbol{u}\|_2} - \phi_3'\big(\|\boldsymbol{v}\|_2\big)\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_2}\right\|_2}{\|\boldsymbol{u}-\boldsymbol{v}\|_2} .
\end{aligned}$$

Letting $\tilde{f}_1\colon B(\boldsymbol{0},1)\to\mathbb{R}$ be defined for all $t\in B(\boldsymbol{0},1)$, by $\tilde{f}_1(\boldsymbol{x}) = \phi_3\big(\|\boldsymbol{x}\|_2\big)$, we obtain

$$\sup_{\substack{\boldsymbol{u},\boldsymbol{v}\in[0,1]^d\\ \boldsymbol{u}\neq\boldsymbol{v}}} \frac{\left\|\nabla f_1(\boldsymbol{u}) - \nabla f_1(\boldsymbol{v})\right\|_2}{\|\boldsymbol{u}-\boldsymbol{v}\|_2} \leq \sup_{\substack{\boldsymbol{u},\boldsymbol{v}\in B(\boldsymbol{0},1)\\ \boldsymbol{u}\neq\boldsymbol{v}}} \frac{\frac{\beta}{\kappa_3^2}\left\|\nabla \tilde{f}_1(\boldsymbol{u}) - \nabla \tilde{f}_1(\boldsymbol{v})\right\|_2}{\|\boldsymbol{u}-\boldsymbol{v}\|_2} .$$

Since $\tilde{f}_1$ is a fixed twice differentiable function which does not depend on ε, we can choose $\beta > 0$ small enough, independently of ε, such that

$$\sup_{\substack{\boldsymbol{u},\boldsymbol{v}\in[0,1]^d\\ \boldsymbol{u}\neq\boldsymbol{v}}} \frac{\left\|\nabla f_1(\boldsymbol{u}) - \nabla f_1(\boldsymbol{v})\right\|_2}{\|\boldsymbol{u}-\boldsymbol{v}\|_2} \leq \frac{1}{\sqrt{d}} .$$

This implies that, for all $\boldsymbol{u},\boldsymbol{v}\in[0,1]^d$,

$$\left\|\nabla f_1(\boldsymbol{u}) - \nabla f_1(\boldsymbol{v})\right\|_\infty \leq \left\|\nabla f_1(\boldsymbol{u}) - \nabla f_1(\boldsymbol{v})\right\|_2 \leq \frac{1}{\sqrt{d}}\|\boldsymbol{u}-\boldsymbol{v}\|_2 \leq \|\boldsymbol{u}-\boldsymbol{v}\|_\infty . \tag{31}$$

Thus, the two bounds (30) and (31) yield

$$\sup_{\substack{\boldsymbol{u},\boldsymbol{v}\in[0,1]^d\\ \boldsymbol{u}\neq\boldsymbol{v}}} \frac{\left\|\nabla (f_0+f_1)(\boldsymbol{u}) - \nabla (f_0+f_1)(\boldsymbol{v})\right\|_\infty}{\|\boldsymbol{u}-\boldsymbol{v}\|_\infty} \leq 3 .$$

Therefore, f_0+f_1 is 3-gradient-Lipschitz. Finally, we have

$$\begin{aligned}
&\inf_{\substack{\boldsymbol{u},\boldsymbol{v}\in[0,1]^d\\ \boldsymbol{u}\neq\boldsymbol{v}}} \frac{\left\langle \nabla (f_0+f_1)(\boldsymbol{u}) - \nabla (f_0+f_1)(\boldsymbol{v}), \frac{\boldsymbol{u}-\boldsymbol{v}}{\|\boldsymbol{u}-\boldsymbol{v}\|_2}\right\rangle}{\|\boldsymbol{u}-\boldsymbol{v}\|_2} \\
&\qquad\geq \inf_{\substack{\boldsymbol{u},\boldsymbol{v}\in[0,1]^d\\ \boldsymbol{u}\neq\boldsymbol{v}}} \frac{\left\langle \nabla f_0(\boldsymbol{u}) - \nabla f_0(\boldsymbol{v}), \frac{\boldsymbol{u}-\boldsymbol{v}}{\|\boldsymbol{u}-\boldsymbol{v}\|_2}\right\rangle}{\|\boldsymbol{u}-\boldsymbol{v}\|_2} - \sup_{\substack{\boldsymbol{u},\boldsymbol{v}\in[0,1]^d\\ \boldsymbol{u}\neq\boldsymbol{v}}} \frac{\left\langle \nabla f_1(\boldsymbol{u}) - \nabla f_1(\boldsymbol{v}), \frac{\boldsymbol{u}-\boldsymbol{v}}{\|\boldsymbol{u}-\boldsymbol{v}\|_2}\right\rangle}{\|\boldsymbol{u}-\boldsymbol{v}\|_2} \geq 2 - \frac{1}{\sqrt{d}} \geq 1 .
\end{aligned}$$

Hence f_0+f_1 is 1-strongly convex and thus it is convex. In conclusion, we have eventually selected a constant $\beta > 0$, independent of ε, such that f_0+f_1 is a convex 3-gradient-Lipschitz function with Δ-proper level set $\{f_0+f_1 = a\}$, but S_n is not a $(\kappa'\varepsilon)$-approximation of $\{f_0+f_1=a\}$. This concludes the proof. ◻

We conclude this section by remarking the analogy between the problem of approximating the level set of a convex function and that of determining an approximation of a convex body in Hausdorff distance. The latter problem has been studied extensively in convex geometry. Notably, while the scope of the and the techniques used in this field differ from ours, the sample complexity results for the two problems are similar. For an overview of these results, we refer the reader to the two surveys (Kamenev, 2019; Gruber, 1993).