Appendix A. Preliminary results

We provide here two geometric results (Corollary 26 and Lemma 27) and a stochastic result (Proposition 28) that are used repeatidly in the computations. We start with the definition of covering numbers.

Definition 24 (Covering number and ϵ-cover) *For any compact and convex set $\mathcal{X} \subset \mathbb{R}^d$ and any $\epsilon > 0$, we say that a sequence $x_1, \ldots, x_n$ of n points in $\mathcal{X}$ defines an ϵ-cover of $\mathcal{X}$ if and only if $\mathcal{X} \subseteq \bigcup_{i=1}^n B(x_i, \epsilon)$. The covering number $\mathcal{N}_\epsilon(\mathcal{X})$ of $\mathcal{X}$ is then defined as the minimal size of a sequence defining an ϵ-cover of $\mathcal{X}$, i.e.*

$$\mathcal{N}_\epsilon(\mathcal{X}) := \inf\left\{ n \in \mathbb{N}^\star : \exists (x_1, \ldots, x_n) \in \mathcal{X}^n \ s.t. \ \mathcal{X} \subseteq \bigcup_{i=1}^n B(x_i, \epsilon) \right\}.$$

The next result provides an upper bound on the covering numbers of hypercubes.

Proposition 25 (Covering number of hypercubes) *Let $[0, R]^d$ be an hypercube of dimensionality $d \geq 1$ whose side has length $R > 0$. Then, for all $\epsilon > 0$, we have that*

$$\mathcal{N}_\epsilon([0, R]^d) \leq (\sqrt{d}R/2\epsilon)^d \vee 1.$$

Proof Observe first that since $[0, R]^d \subseteq B(c, \sqrt{d}R/2)$ where c denotes the center of the hypercube, then the result trivially holds for any $\epsilon \geq \sqrt{d}R/2$. Fix any $\epsilon < \sqrt{d}R/2$, set $N_\epsilon = \lceil \sqrt{d}R/2\epsilon \rceil$ and define for all $I \in \{0, \ldots, N_\epsilon - 1\}^d$ the series $H_I = I \times R/N_\epsilon + [0, R/N_\epsilon]^d$ of N_ϵ^d hypercubes which cover $[0, R]^d := \bigcup_{I \in \{0, \ldots, N_\epsilon - 1\}^d} H_I$. Dentoting by c_I the center of H_I and since $\max_{x \in H_I} \|x - c_I\|_2 \leq \epsilon$, it necessarily follows that $H_I \subseteq B(c_I, \epsilon)$ which implies that $[0, R]^d \subseteq \bigcup_{I \in \{0, \ldots, N_\epsilon - 1\}^d} B(c_I, \epsilon)$ and proves that $\mathcal{N}_\epsilon([0, R]^d) \leq N_\epsilon^d \leq (\sqrt{d}R/2\epsilon)^d$. $\square$

This result can be extended to any compact and convex set of $\mathbb{R}^d$ as shown below.

Corollary 26 (Covering number of a convex set) *For any bounded compact and convex set $\mathcal{X} \subset \mathbb{R}^d$, we have that $\forall \epsilon > 0$,*

$$\mathcal{N}_\epsilon(\mathcal{X}) \leq (\sqrt{d} \operatorname{diam}(\mathcal{X}) / \epsilon)^d \vee 1.$$

Proof First, we show that $\mathcal{N}_\epsilon(\mathcal{X}) \leq \mathcal{N}_\epsilon([0, 2 \operatorname{diam}(\mathcal{X})]^d)$ and then, we use the bound of Proposition 25 to conclude the proof. By definition of $\operatorname{diam}(\mathcal{X})$, we know that there exists some $x \in \mathbb{R}^d$ such that $\mathcal{X} \subseteq x + [0, 2 \operatorname{diam}(\mathcal{X})]^d$. Hence, we know from Proposition 25 that there exists a sequence $c_1, \ldots, c_{N_\epsilon}$ of $N_\epsilon \leq \mathcal{N}_\epsilon([0, 2 \operatorname{diam}(\mathcal{X})]^d)$ points in $[0, 2 \operatorname{diam}(\mathcal{X})]^d$ forming an ϵ-cover of $\mathcal{X}$:

$$\mathcal{X} \subseteq [0, 2 \operatorname{diam}(\mathcal{X})]^d \subseteq \bigcup_{i=1}^{N_\epsilon} B(c_i, \epsilon). \tag{1}$$

However, we do not have the guarantee at this point that the centers $c_1, \ldots c_{N_\epsilon}$ belong to $\mathcal{X}$. To build an ϵ-cover of $\mathcal{X}$, we project each of those centers on $\mathcal{X}$. More precisely, we show that $\mathcal{X} \subseteq \bigcup_{i=1}^{N_\epsilon} B(\Pi_{\mathcal{X}}(c_i), \epsilon)$ where $\Pi_{\mathcal{X}} : x \in \mathbb{R}^d \mapsto \arg\min_{x' \in \mathcal{X}} \|x - x'\|_2 \in \mathcal{X}$ denotes the projection over the compact and convex set $\mathcal{X}$. Starting from (1), it sufficient to show that $B(c_i, \epsilon) \cap \mathcal{X} \subseteq B(\Pi_{\mathcal{X}}(c_i), \epsilon)$, for all $i \in \{1, \dots, N_\epsilon\}$ to prove that

$$\mathcal{X} \subseteq \bigcup_{i=1}^{N_\epsilon} B(c_i, \epsilon) \cap \mathcal{X} \subseteq \bigcup_{i=1}^{N_\epsilon} B(\Pi_{\mathcal{X}}(c_i), \epsilon).$$

Pick any $c \in \{c_1, \dots, c_{N_\epsilon}\}$ and consider the following cases on the distance $\|c - \Pi_{\mathcal{X}}(c)\|_2$ between the center and its projection: (i) if $\|c - \Pi_{\mathcal{X}}(c)\|_2 = 0$, then $c = \Pi_{\mathcal{X}}(c)$ and we have $B(c, \epsilon) \cap \mathcal{X} \subseteq B(\Pi_{\mathcal{X}}(c), \epsilon)$, (ii) If $\|c - \Pi_{\mathcal{X}}(c)\|_2 > \epsilon$, then $\mathcal{X} \cap B(c, \epsilon) = \emptyset$, and we have $\mathcal{X} \cap B(c, \epsilon) \subseteq B(\Pi_{\mathcal{X}}(c), \epsilon)$. We now consider the non-trivial case where $\|c - \Pi_{\mathcal{X}}(c)\|_2 \in (0, \epsilon)$. Pick any $x \in B(c, \epsilon) \cap \mathcal{X}$ and note that since $x \in B(c, \epsilon)$, then

$$\begin{aligned}
\epsilon^2 &\geq \|x - c\|_2^2 \\
&= \|x - \Pi_{\mathcal{X}}(c) + \Pi_{\mathcal{X}}(c) - c\|_2^2 \\
&= \|x - \Pi_{\mathcal{X}}(c)\|_2^2 + \|c - \Pi_{\mathcal{X}}(c)\|_2^2 + 2 \cdot \langle x - \Pi_{\mathcal{X}}(c), \Pi_{\mathcal{X}}(c) - c \rangle
\end{aligned}$$

which combined with the fact that $\|c - \Pi_{\mathcal{X}}(c)\|_2^2 \geq 0$ gives

$$\|x - \Pi_{\mathcal{X}}(c)\|_2^2 \leq \epsilon^2 - 2 \cdot \langle x - \Pi_{\mathcal{X}}(c), \Pi_{\mathcal{X}}(c) - c \rangle. \tag{2}$$

We will simply show that the inner product $\langle x - \Pi_{\mathcal{X}}(c), \Pi_{\mathcal{X}}(c) - c \rangle$ cannot be stricly negative to prove that $\|x - \Pi_{\mathcal{X}}(c)\|_2 \leq \epsilon$. Assume by contradiction that $\langle x - \Pi_{\mathcal{X}}(c), \Pi_{\mathcal{X}}(c) - c \rangle < 0$. Since $\Pi_{\mathcal{X}}(c) \in \mathcal{X}$ and $x \in \mathcal{X}$, it follows the convexity of $\mathcal{X}$ implies that $\forall \lambda \in [0, 1]$, $x_\lambda = \Pi_{\mathcal{X}}(c) + \lambda \cdot (x - \Pi_{\mathcal{X}}(c)) \in \mathcal{X}$. However, for all $\lambda \in (0, 1)$ we have that

$$\begin{aligned}
\|x_\lambda - c\|_2^2 &= \|\Pi_{\mathcal{X}}(c) - c + \lambda \cdot (x - \Pi_{\mathcal{X}}(c))\|_2^2 \\
&= \|\Pi_{\mathcal{X}}(c) - c\|_2^2 + \lambda^2 \|x - \Pi_{\mathcal{X}}(c)\|_2^2 + 2\lambda \cdot \langle \Pi_{\mathcal{X}}(c) - c, x - \Pi_{\mathcal{X}}(c) \rangle \\
&= \|\Pi_{\mathcal{X}}(c) - c\|_2^2 + \lambda \cdot (\lambda \|x - \Pi_{\mathcal{X}}(c)\|_2^2 + 2 \cdot \langle \Pi_{\mathcal{X}}(c) - c, x - \Pi_{\mathcal{X}}(c) \rangle).
\end{aligned}$$

Therefore, taking any $0 < \lambda^\star < |\langle \Pi_{\mathcal{X}}(c) - c, x - \Pi_{\mathcal{X}}(c) \rangle)| / \|\Pi_{\mathcal{X}}(c) - c\|_2^2 \wedge 1$ so that the second term of the right hand term of the previous equation is strictly negative gives that $\|x_{\lambda^\star} - c\|_2^2 < \|\Pi_{\mathcal{X}}(c) - c\|_2^2$ leads us to the following contradiction $\min_{x \in \mathcal{X}} \|x - c\|_2 \leq \|x_{\lambda^\star} - c\|_2 < \|\Pi_{\mathcal{X}}(c) - c\|_2 = \min_{x \in \mathcal{X}} \|x - c\|_2$. Hence, $\langle x - \Pi_{\mathcal{X}}(c), \Pi_{\mathcal{X}}(c) - c \rangle \geq 0$ and we deduce from (2) that $\mathcal{X} \cap B(c, \epsilon) \subseteq B(\Pi_{\mathcal{X}}(c), \epsilon)$, which completes the proof. □

The next inequality will be useful to bound to bound volume of the intersection of a ball and a convex set.

Lemma 27 *(From Zabinsky and Smith (1992), see Appendix Section therein). For any compact and convex set $\mathcal{X} \subset \mathbb{R}^d$ with non-empty interior, we have that for any $x^\star \in \mathcal{X}$ and $\epsilon \in (0, \mathrm{diam}(\mathcal{X}))$,*

$$\frac{\mu(B(x^\star, \epsilon) \cap \mathcal{X})}{\mu(\mathcal{X})} \geq \left(\frac{\epsilon}{\mathrm{diam}(\mathcal{X})} \right)^d.$$

Proof We point out that a detailed proof of this result can be found in the Appendix Section of (Zabinsky and Smith (1992)). Nonetheless, we provide here a proof with less details for completeness. Introduce the similarity transformation $S : \mathbb{R}^d \to \mathbb{R}^d$ defined by

$$S : x \mapsto x^\star + \frac{r}{\mathrm{diam}(\mathcal{X})}(x - x^\star)$$

and let $S(\mathcal{X}) := \{S(x) : x \in \mathcal{X}\}$ be the image of $\mathcal{X}$ by S. Since $x^\star \in \mathcal{X}$ and $\max_{x\in\mathcal{X}} \|x - x^\star\|_2 \leq \mathrm{diam}(\mathcal{X})$ by definition, it follows from the convexity of $\mathcal{X}$ that $S(\mathcal{X}) \subseteq B(x^\star, r) \cap \mathcal{X}$ which implies that $\mu(B(x^\star, r) \cap \mathcal{X}) \geq \mu(S(\mathcal{X}))$. However, as S is a similarity transformation conserves the ratios of the volumes before/after transformation, we thus deduce that

$$\frac{\mu(B(x^\star, r) \cap \mathcal{X})}{\mu(\mathcal{X})} \geq \frac{\mu(S(\mathcal{X}))}{\mu(\mathcal{X})} = \frac{\mu(S(B(x^\star, \mathrm{diam}(\mathcal{X}))))}{\mu(B(x^\star, \mathrm{diam}(\mathcal{X})))} = \frac{\mu(B(x^\star, r))}{\mu(B(x^\star, \mathrm{diam}(\mathcal{X})))}$$

and the result follows using the fact that $\forall r \geq 0$, $\mu(B(x^\star, r)) = \pi^{d/2} r^d / \Gamma(d/2 + 1)$ where $\Gamma(\cdot)$ stands for the standard gamma function. □

Proposition 28 (Pure Random Search) *Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact and convex set with non-empty interior and let $f \in \mathrm{Lip}(k)$ be a k-Lipschitz functions defined on $\mathcal{X}$ for some $k \geq 0$. Then, for any $n \in \mathbb{N}^\star$ and $\delta \in (0, 1)$, we have with probability at least $1 - \delta$,*

$$\max_{x\in\mathcal{X}} f(x) - \max_{i=1\dots n} f(X_i) \leq k \cdot \mathrm{diam}(\mathcal{X}) \cdot \left(\frac{\ln(1/\delta)}{n}\right)^{\frac{1}{d}}$$

where $X_1, \dots, X_n$ denotes a sequence of n independent copies of $X \sim \mathcal{U}(\mathcal{X})$.

Proof Fix any $n \in \mathbb{N}^\star$ and $\delta \in (0, 1)$, let $\epsilon = k\,\mathrm{diam}(\mathcal{X})\,(\ln(1/\delta)/n)^{1/d}$ be the value of the upper bound and $\mathcal{X}_\epsilon = \{x \in \mathcal{X} : f(x) \geq \max_{x\in\mathcal{X}} f(x) - \epsilon\}$ the corresponding level set. As the result trivially holds whenever $n \leq \ln(1/\delta)$, we consider that $n > \ln(1/\delta)$. Observe now that since $f \in \mathrm{Lip}(k)$, then for any $x^\star \in \arg\max_{x\in\mathcal{X}} f(x)$, we have that $\mathcal{X} \cap B(x^\star, \epsilon/k) \subseteq \mathcal{X}_\epsilon$ since $|f(x) - f(x^\star)| \leq k \cdot \|x - x^\star\|_2 = \epsilon$ for all $x \in B(x^\star, \epsilon/k) \cap \mathcal{X}$. Therefore, by picking any $x^\star \in \arg\max_{x\in\mathcal{X}} f(x)$, one gets

$$\begin{aligned}
\mathbb{P}\left(\max_{i=1\dots n} f(X_i) \geq \max_{x\in\mathcal{X}} f(x) - \epsilon\right) &= \mathbb{P}\left(\bigcup_{i=1}^{n} \{X_i \in \mathcal{X}_\epsilon\}\right) && \text{(def. of } \mathcal{X}_\epsilon) \\
&= 1 - \mathbb{P}\left(X_1 \notin \mathcal{X}_\epsilon\right)^n && \text{(i.i.d. r.v.)} \\
&\geq 1 - \mathbb{P}(X_1 \notin \mathcal{X} \cap B(x^\star, \epsilon/k))^n && (\mathcal{X} \cap B(x^\star, \epsilon/k) \subseteq \mathcal{X}_\epsilon) \\
&= 1 - \left(1 - \left(\frac{\mu(\mathcal{X} \cap B(x^\star, \epsilon/k))}{\mu(\mathcal{X})}\right)^d\right)^n && (X_1 \sim \mathcal{U}(\mathcal{X})) \\
&\geq 1 - \left(1 - \left(\frac{\epsilon}{k\,\mathrm{diam}(\mathcal{X})}\right)^d\right)^n && \text{(Lemma 27)} \\
&= 1 - \left(1 - \frac{\ln(1/\delta)}{n}\right)^n && \text{(def. of } \epsilon) \\
&\geq 1 - \delta. && (1 + x \leq e^x)
\end{aligned}$$

□

Appendix B. Proofs of Section 3

In this section, we provide the proofs of Propositions 3, 5, 6 and Example 4.

Proof of proposition 3. ($\Leftarrow$) Let A be any global optimization algorithm such that $\forall f \in \bigcup_{k\geq 0} \mathrm{Lip}(k)$, $\sup_{x\in\mathcal{X}} \min_{i=1\dots n} \|X_i - x\|_2 \xrightarrow{p} 0$. Pick any $\epsilon > 0$, any $f \in \bigcup_{k\geq 0} \mathrm{Lip}(k)$ and let $\mathcal{X}_\epsilon = \{x \in \mathcal{X} : f(x) \geq \max_{x\in\mathcal{X}} f(x) - \epsilon\}$ be the corresponding level set. As $\mathcal{X}_\epsilon$ is non-empty, there necessarily exists some $x_\epsilon \in \mathcal{X}$ and $r_\epsilon > 0$ such that $B(x_\epsilon, r_\epsilon) \cap \mathcal{X} \subseteq \mathcal{X}_\epsilon$. Thus, if $X_1, \dots, X_n$ denotes a sequence a sequence of n evaluation points generated by A over f, we directly obtain from the convergence in probability of the mesh grid that

$$\begin{aligned}
\mathbb{P}\left(\max_{x\in\mathcal{X}} f(x) - \max_{i=1\dots n} f(X_i) > \epsilon\right) &= \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \notin \mathcal{X}_\epsilon\}\right) \\
&\leq \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \notin B(x_\epsilon, r_\epsilon)\}\right) \\
&= \mathbb{P}\left(\min_{i=1\dots n} \|X_i - x_\epsilon\|_2 > r_\epsilon\right) \\
&\leq \mathbb{P}\left(\sup_{x\in\mathcal{X}} \min_{i=1\dots n} \|X_i - x\|_2 > r_\epsilon\right) \xrightarrow[n\to\infty]{} 0.
\end{aligned}$$

($\Rightarrow$) Let A be any global optimization algorithm consistent over the set of Lipschitz functions and assume by contradiction that there exists some $f^\star \in \bigcup_{k\geq 0} \mathrm{Lip}(k)$ such that $\sup_{x\in X} \min_{i=1\dots n} \|x - X_i\|_2 \overset{p}{\nrightarrow} 0$. The implication is proved in two steps: first, we show that there exists a ball $B(c^\star, \epsilon)$ for some $c^\star \in \mathcal{X}$ which is almost never hit by the algorithm and second, we build a Lipschitz function which admits its maximum over this ball.

First step. Let $\{X_i\}_{i\in\mathbb{N}^\star}$ be a sequence of evaluation points generated by A over $f^\star$. Observe first that since for all $\epsilon > 0$, the series $n \in \mathbb{N}^\star \mapsto \mathbb{P}(\sup_{x\in X} \min_{i=1\dots n} \|x - X_i\|_2 > \epsilon)$ is non-increasing, then the contradiction assumption necessarily implies that

$$\exists \epsilon_1, \epsilon_2 > 0 \text{ such that } \forall n \in \mathbb{N}^\star,\ \mathbb{P}\left(\sup_{x\in\mathcal{X}} \min_{i=1\dots n} \|x - X_i\|_2 > \epsilon_1\right) > \epsilon_2. \tag{3}$$

Consider now any sequence $c_1, \dots, c_{N_1}$ of $N_1 = \mathcal{N}_{\epsilon_1}(\mathcal{X})$ points in $\mathcal{X}$ defining an ϵ_1-cover of $\mathcal{X}$ and suppose by contradiction that

$$\forall c \in \{c_1, \dots, c_{N_1}\},\ \exists n_c \in \mathbb{N}^\star \text{ such that } \mathbb{P}\left(\bigcap_{i=1}^{n_c} \{X_i \notin B(c, \epsilon_1) \cap \mathcal{X}\}\right) \leq \frac{\epsilon_2}{2N_1}$$

which gives by setting $N_2 = \max_{c\in\{c_1,\cdots,c_{N_1}\}} n_c$ that

$$\forall c \in \{c_1, \dots, c_{N_1}\},\ \mathbb{P}\left(\bigcap_{i=1}^{N_2} \{X_i \notin B(c, \epsilon) \cap \mathcal{X}\}\right) \leq \frac{\epsilon_2}{2N_1}.$$

However, as $c_1, \dots, c_{N_1}$ form an ϵ_1-cover of $\mathcal{X}$, it follows that

$$\begin{aligned}
\mathbb{P}\left(\sup_{x\in\mathcal{X}} \min_{i=1\dots N_2} \|x - X_i\|_2 \le \epsilon_1\right) &\ge \mathbb{P}\left(\bigcap_{j=1}^{N_1}\bigcup_{i=1}^{N_2}\{X_i \in B(c_j,\epsilon_1)\cap\mathcal{X}\}\right)\\
&= 1 - \mathbb{P}\left(\bigcup_{j=1}^{N_1}\bigcap_{i=1}^{N_2}\{X_i \notin B(c_j,\epsilon_1)\cap\mathcal{X}\}\right)\\
&\ge 1 - \sum_{j=1}^{N_1}\mathbb{P}\left(\bigcap_{i=1}^{N_2}\{X_i \notin B(c_j,\epsilon)\cap\mathcal{X}\}\right)\\
&\ge 1 - N_1 \times \frac{\epsilon_2}{2N_1}\\
&= 1 - \frac{\epsilon_2}{2}
\end{aligned}$$

which contradicts (3). Hence, we deduce that

$$\exists c^\star \in \{c_1, \dots, c_{N_\epsilon}\} \text{ such that } \forall n \in \mathbb{N}^\star,\ \mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \notin B(c^\star,\epsilon_1)\cap\mathcal{X}\}\right) \ge \frac{\epsilon_2}{2N_1}.$$

Second Step. Based on this center $c^\star \in \mathcal{X}$, one can introduce the function $\tilde{f} : \mathcal{X} \mapsto \mathbb{R}$ defined for all $x \in \mathcal{X}$ by

$$\tilde{f}(x) = \begin{cases} f^\star(x) + 3\left(1 - \frac{\|c^\star - x\|_2}{\epsilon_1}\right) \times (\max_{x\in\mathcal{X}} f^\star(x) - \min_{x\in\mathcal{X}} f^\star(x)) & \text{if } x \in B(c^\star,\epsilon_1)\\ f^\star(x) & \text{otherwise}\end{cases}$$

which is maximized over $B(c^\star,\epsilon_1)$ and Lipschitz continuous as both $f^\star$ and $x \mapsto \|c^\star - x\|_2$ are Lipschitz. However, since $\tilde{f}$ and $f^\star$ can not be distinguished over $\mathcal{X}/B(c,\epsilon_1)$, we have that $\forall n \in \mathbb{N}^\star$,

$$\begin{aligned}
\mathbb{P}\left(\max_{x\in\mathcal{X}} \tilde{f}(x) - \max_{i=1\dots n}\tilde{f}(X'_i) > \max_{x\in\mathcal{X}} f(x)\right) &\ge \mathbb{P}\left(\bigcap_{i=1}^{n}\{X'_i \notin B(c,\epsilon_2)\cap\mathcal{X}\}\right)\\
&= \mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \notin B(c,\epsilon_2)\cap\mathcal{X}\}\right)\\
&\ge \epsilon_2/(2N_1)\\
&> 0
\end{aligned}$$

where $X'_1, \dots, X'_n$ denotes a sequence of evaluation points generated by A over $\tilde{f}$, and we deduce that there exists $\tilde{f} \in \bigcup_{k\ge0}\mathrm{Lip}(k)$ such that $\max_{i=1\dots n}\tilde{f}(X'_i) \overset{p}{\nrightarrow} \max_{x\in\mathcal{X}}\tilde{f}(x)$. Hence, it contradicts the fact that A is consistent over $\bigcup_{k\ge0}\mathrm{Lip}(k)$ and we deduce that, necessarily, $\sup_{x\in\mathcal{X}}\min_{i=1\dots n}\|X_i - x\| \xrightarrow{p} 0$ for all $f \in \bigcup_{k\ge0}\mathrm{Lip}(k)$. □

Proof of Example 4. Fix any $n \in \mathbb{N}^\star$, set $\delta \in (0,1)$, define $\epsilon = \mathrm{diam}(\mathcal{X}) \cdot ((\ln(n/\delta) + d\ln(d))/n)^{1/d}$ and let $X_1, \dots, X_n$ be a sequence of n independent copies of

$X \sim \mathcal{U}(\mathcal{X})$. Since the result trivially holds whenever $(\ln(n/\delta) + d\ln(d))/n \geq 1$, we consider the case where $(\ln(n/\delta) + d\ln(d))/n < 1$. From Proposition 26, we know that there exists a sequence $x_1, \ldots, x_{N_\epsilon}$ of $N_\epsilon = \mathcal{N}_\epsilon(\mathcal{X})$ points in $\mathcal{X}$ such that $\mathcal{X} \subseteq \bigcup_{j=1}^{N_\epsilon} B(x_j, \epsilon)$. Therefore, using the bound on the covering number $\mathcal{N}_\epsilon(\mathcal{X})$ of Corollary 26, we obtain that

$$
\begin{aligned}
\mathbb{P}\left(\sup_{x\in\mathcal{X}} \min_{i=1\ldots n} \|x - X_i\|_2 \leq \epsilon\right) &\geq \mathbb{P}\left(\bigcap_{j=1}^{N_\epsilon} \bigcup_{i=1}^{n} \{X_i \in B(x_j, \epsilon) \cap \mathcal{X}\}\right) \\
&= 1 - \mathbb{P}\left(\bigcup_{j=1}^{N_\epsilon} \bigcap_{i=1}^{n} \{X_i \notin B(x_j, \epsilon) \cap \mathcal{X}\}\right) \\
&\geq 1 - \sum_{j=1}^{N_\epsilon} \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \notin B(x_j, \epsilon) \cap \mathcal{X}\}\right) \\
&\geq 1 - N_\epsilon \times \max_{j=1\ldots N_\epsilon} \mathbb{P}(X_1 \notin B(x_j, \epsilon) \cap \mathcal{X})^n \\
&= 1 - N_\epsilon \times \max_{j=1\ldots N_\epsilon} \left(1 - \frac{\mu(\mathcal{X} \cap B(x_j, \epsilon))}{\mu(\mathcal{X})}\right)^n \\
&\geq 1 - N_\epsilon \times \left(1 - \left(\frac{\epsilon}{\mathrm{diam}(\mathcal{X})}\right)^d\right)^n \\
&\geq 1 - \left(\frac{\sqrt{d}\,\mathrm{diam}(\mathcal{X})}{\epsilon}\right)^d \times \left(1 - \left(\frac{\epsilon}{\mathrm{diam}(\mathcal{X})}\right)^d\right)^n \\
&\geq 1 - \delta
\end{aligned}
$$

and the proof is complete. □

Proof of Proposition 5. The proof heavily builds upon the arguments used in the proof of the Theorem 1 in (Bull (2011)). Pick any algorithm $A \in \mathcal{A}$ and any constant $C > 0$. Fix any $n \in \mathbb{N}^\star$ and $\delta \in (0,1)$ and set $N_\delta = \lceil (n/\delta)^{1/d} \rceil$. By definition of $\mathrm{rad}(\mathcal{X})$, we know there exists some $x \in \mathcal{X}$ such that $x + [0, 2\,\mathrm{rad}(\mathcal{X})/\sqrt{d}]^d \subseteq \mathcal{X}$. One can then define for all $I \in \{1, \ldots, N_\delta\}^d$, the centers c_I of the hypercubes H_I whose side are equal to $D = 2\,\mathrm{rad}(\mathcal{X})/(\sqrt{d}N_\delta)$ and cover $\mathcal{X}$, *i.e.*, $\bigcup_I H_I = x + [0, 2\,\mathrm{rad}(\mathcal{X})/\sqrt{d}]^d \subseteq \mathcal{X}$. Now, let $X_1, \ldots, X_n$ be a sequence of n evaluation points generated by the algorithm A over the constant function $f_0 : x \in \mathcal{X} \mapsto 0$ and define for all $I \in \{1, \ldots, N_\delta\}^d$ the event

$$
E_I = \bigcap_{i=1}^{n} \{X_i \notin \mathrm{Int}(H_I)\}.
$$

As the interiors of the N_δ^d hypercubes are disjoint and we have n points, it necessarily follows that

$$
N_\delta^d \times \max_I \mathbb{P}(E_I) \geq \sum_I \mathbb{P}(E_I) = \mathbb{E}\left[\sum_I \mathbb{I}\{E_I\}\right] \geq N_\delta^d - n.
$$

Hence, there exsits some fixed $I^\star$ only depending on $\mathcal{A}$ which maximizes the above probability and thus satisfies

$$\mathbb{P}(E_{I^\star}) \geq \frac{N_\delta^d - n}{N_\delta^d} = 1 - \frac{n}{\lceil (n/\delta)^{1/d} \rceil^d} \geq 1 - \delta.$$

Now, using the center $c_{I^\star}$ of the hypercube $H_{I^\star}$, one can then introduce the function $\tilde{f} \in \bigcup_{k \geq 0} \mathrm{Lip}(k)$ defined for all $x \in \mathcal{X}$ by

$$\tilde{f}(x) = \begin{cases} C \times (1 - 2\left\|c_{I^\star} - x\right\|_2 / D) & \text{if } \left\|c_{I^\star} - x\right\|_2 \leq D/2 \\ 0 & \text{otherwise.} \end{cases}$$

However, since the functions $\tilde{f}$ and f_0 can not be distinguished over $\mathcal{X}/H_{I^\star}$, we have that

$$\mathbb{P}\left(\max_{x \in X} \tilde{f}(x) - \max_{i=1\ldots n} \tilde{f}(X_i') \geq C \right) \geq \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i' \notin \mathrm{Int}(H_{I^\star})\} \right) = \mathbb{P}(E_{I^\star}) \geq 1 - \delta$$

where $X_1', \ldots, X_n'$ denotes a sequence of evaluation points generated by $\mathcal{A}$ over $\tilde{f}$, which proves the result. □

Proof of Proposition 6. (Lower bound). Pick any $n \in \mathbb{N}^\star$ and set $D = 2\,\mathrm{rad}(\mathcal{X})/(\sqrt{d}\lceil (2n)^{1/d} \rceil)$. It can easily be shown by reproducing the same steps as in the proof of Proposition 5 with δ set to 1/2, that for any global optimization algorithm A, there exists a function $\tilde{f}_A \in \mathrm{Lip}(k)$ defined by

$$\tilde{f}_A(x) = \begin{cases} kD/2 - k \cdot \left\|c_A - x\right\|_2 & \text{if } \left\|c_A - x\right\|_2 \leq D/2 \\ 0 & \text{otherwise,} \end{cases}$$

for some center $c_A \in \mathcal{X}$ only depending on A, for which we have $\mathbb{P}(\max_{x \in X} \tilde{f}_A(x) - \max_{i=1\ldots n} \tilde{f}_A(X_i) \geq k \cdot D/2) \geq 1/2$ where $X_1, \ldots, X_n$ is a sequence of n evaluation points generated by A over $\tilde{f}_A$. Therefore, using the definition of the supremum and Markov's inequality gives that $\forall A \in \mathcal{A}$:

$$\begin{aligned} \sup_{f \in \mathrm{Lip}(k)} \mathbb{E}\left[\max_{x \in \mathcal{X}} f(x) - \max_{i=1\ldots n} f(X_i) \right] &\geq \mathbb{E}\left[\max_{x \in \mathcal{X}} \tilde{f}_A(x) - \max_{i=1\ldots n} \tilde{f}_A(X_i) \right] \\ &\geq \frac{kD}{2} \times \mathbb{P}\left(\max_{x \in X} \tilde{f}_A(x) - \max_{i=1\ldots n} \tilde{f}_A(X_i) \geq k \cdot \frac{D}{2} \right) \\ &\geq k \cdot \frac{\mathrm{rad}(\mathcal{X})}{8\sqrt{d}} \cdot n^{-\frac{1}{d}}. \end{aligned}$$

As the previous inequaliy holds true for any algorithm A, the proof is complete.

(Upper bound). Sequentially using the fact that (i) the infinimum minimax loss taken over all algorithms is necessarily upper bounded by the loss suffered by a Pure Random

Search, (ii) for any positive random variable, $\mathbb{E}[X] = \int_{t=0}^{\infty} \mathbb{P}(X \geq t)\mathrm{d}t$, (iii) Proposition 28 and (iv) the change of variable $u = n(t/\operatorname{diam}(\mathcal{X}))^{1/d}$, we obtain that

$$\begin{aligned}\inf_{A\in\mathcal{A}} \sup_{f\in\mathrm{Lip}(k)} \mathbb{E}\left[\max_{x\in\mathcal{X}} f(x) - \max_{i=1\dots n} f(X_i)\right] &\leq \sup_{f\in\mathrm{Lip}(k)} \mathbb{E}\left[\max_{x\in\mathcal{X}} f(x) - \max_{i=1\dots n} f(X_i')\right] \\ &\leq \int_0^{\infty} \exp\left\{-n(t/k\cdot\operatorname{diam}(\mathcal{X}))^{1/d}\right\}\mathrm{d}t \\ &= k\cdot\operatorname{diam}(\mathcal{X})\cdot n^{-d}\cdot d\cdot\int_0^{\infty} u^{d-1}e^{-u}\mathrm{d}u \\ &= k\cdot\operatorname{diam}(\mathcal{X})\cdot n^{-d}\cdot d\cdot\Gamma(d)\end{aligned}$$

where $X_1',\dots,X_n'$ denotes a sequence of n independent copies of $X' \sim \mathcal{U}(\mathcal{X})$ and $\Gamma(\cdot)$ the Euler's Gamma function. Recalling that $\Gamma(d) = (d-1)!$ for all $d \in \mathbb{N}^\star$ completes the proof. $\square$

Appendix C. Proofs of Section 3

In this section, we provide the proofs for Lemma 9, Proposition 11, Proposition 12, Corollary 13, Proposition 14, Theorem 15 and Theorem 16.

Proof of Lemma 9. The first implication ($\Rightarrow$) is a direct consequence of the definition of $\mathcal{X}_{k,t}$. Noticing that the function $\hat{f} : x \mapsto \min(\max_{i=1\dots t} f(X_i), \min_{i=1\dots t} f(X_i)+k\left\|x - X_i\right\|_2)$ belongs to $\mathcal{F}_{k,t}$ and that $\arg\max_{x\in\mathcal{X}} \hat{f}(x) = \{x \in \mathcal{X} : \min_{i=1\dots t} f(X_i) + k\left\|x - X_i\right\|_2 \geq \max_{i=1\dots t} f(X_i)\}$ proves the second implication. $\square$

Proof of Proposition 11. Fix any $f \in \mathrm{Lip}(k)$, pick any $n \in \mathbb{N}^\star$, set $\epsilon > 0$ and let $\mathcal{X}_\epsilon = \{x \in \mathcal{X} : f(x) \geq \max_{x\in\mathcal{X}} f(x) - \epsilon\}$ be the corresponding level set. Denoting by $X_1',\dots,X_n'$ a sequence of n random variable uniformly distributed over $\mathcal{X}$ and observing that $\mu(\mathcal{X}_\epsilon) > 0$, we directly obtain from Proposition 12 that

$$\begin{aligned}\mathbb{P}\left(\max_{x\in\mathcal{X}} f(x) - \max_{i=1\dots n} f(X_i) > \epsilon\right) &\leq \mathbb{P}\left(\max_{x\in\mathcal{X}} f(x) - \max_{i=1\dots n} f(X_i') > \epsilon\right) \\ &= \mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i' \notin \mathcal{X}_\epsilon\}\right) \\ &\leq \left(1 - \frac{\mu(\mathcal{X}_\epsilon)}{\mu(\mathcal{X})}\right)^n \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

$\square$

Proof of Proposition 12. The proof is similar to the one of Proposition 12 in (Malherbe and Vayatis (2016)). $\square$

Proof of Corollary 13. Combining Proposition 12 and Proposition 28 stated at the begining of the Appendix Section gives the result. $\square$

Proof of Proposition 14. Fix any $\delta \in (0,1)$, set $n \in \mathbb{N}^\star$ and let $r_{\delta,n} = \operatorname{rad}(\mathcal{X})\,(\delta/n)^{\frac{1}{d}}$ be the value of the lower bound divided by k. As $\text{rad}(\mathcal{X}) > 0$, there necessarily exists some point $x^\star \in \mathcal{X}$ such that $B(x^\star, \text{rad}(\mathcal{X})) \subseteq \mathcal{X}$. Based on this point, one can then introduce the function $\tilde{f} \in \text{Lip}(k)$ defined for all $x \in \mathcal{X}$ by

$$\tilde{f}(x) = \begin{cases} k \cdot r_{\delta,n} - k \cdot \|x - x^\star\|_2 & \text{if } x \in B(x^\star, r_{\delta,n}) \\ 0 & \text{otherwise.} \end{cases}$$

Denoting now by $X_1, \dots, X_n$ a sequence of n evaluation points generated by LIPO tuned with a parameter k over $\tilde{f}$ and observing that (i) X_1 is uniformly distributed over $\mathcal{X}$ and (ii) X_{i+1} is also uniformly distributed over $\mathcal{X}$ for $i \geq 1$ as soon as only constant evaluations have been recorded (*i.e.* $\mathcal{X}_{k,i+1} = \mathcal{X}$ on the event $\bigcap_{t \leq i}\{X_t \notin B(x^\star, r_{\delta,n})\}$), we have that

$$\begin{aligned}
\mathbb{P}\left(\max_{x \in \mathcal{X}} \tilde{f}(x) - \max_{i=1\dots n} \tilde{f}(X_i) \geq k \cdot r_{\delta,n}\right) &\geq \mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \notin B(x^\star, r_{\delta,n})\}\right) \\
&= \Bigg[\mathbb{P}(X_1 \notin B(x^\star, r_{\delta,n}))\times \\
&\quad \prod_{i=1}^{n-1} \mathbb{P}\left(X_{i+1} \notin B(x^\star, r_{\delta,n}) \mid \bigcap_{t=1}^{i}\{X_t \notin B(x^\star, r_{\delta,n})\}\right)\Bigg] \\
&= \left(1 - \frac{\mu(B(x^\star, r_{\delta,n}) \cap \mathcal{X})}{\mu(\mathcal{X})}\right)^n \\
&\geq \left(1 - \left(\frac{r_{\delta,n}}{\text{rad}(\mathcal{X})}\right)^d\right)^n \\
&= \left(1 - \frac{\delta}{n}\right)^n \\
&\geq 1 - \delta.
\end{aligned}$$

□

Proof of Theorem 15. Pick any $n \in \mathbb{N}^\star$, fix any $\delta \in (0,1)$ and let $X_1, \dots, X_n$ be a sequence of n evaluation points generated by the LIPO algorithm over f after n iterations. To clarify the proof, we set some specific notations: let $D = \max_{x \in \mathcal{X}} \|x - x^\star\|_2$, set

$$M = \begin{cases} \left\lfloor \left(\frac{c_\kappa}{8k}\right)^d \cdot \frac{n}{\ln(n/\delta) + 2(2\sqrt{d})^d} \right\rfloor & \text{if } \kappa = 1 \\[2ex] \left\lfloor \frac{1}{\ln(2)d(\kappa - 1)} \ln\left(1 + \left(\frac{c_\kappa D^\kappa}{8kD}\right)^d \frac{n(2^{d(\kappa-1)} - 1)}{\ln(n/\delta) + 2(2\sqrt{d})^d}\right) \right\rfloor & \text{otherwise,} \end{cases}$$

define for all $m \in \{1 \dots M\}$ the series of integers:

$$N_m := \left\lceil \sqrt{d} \cdot \left(\frac{8kD}{c_\kappa D^\kappa}\right) \cdot 2^{m(\kappa-1)} \right\rceil^d \quad \text{and} \quad N'_m := \left\lceil \ln(M/\delta) \cdot \left(\frac{8kD}{c_\kappa D^\kappa}\right)^d \cdot 2^{md(\kappa-1)} \right\rceil$$

and let $\tau_0, \ldots, \tau_M$ be the series of stopping times initialized by $\tau_0 = 0$ and defined for all $m \geq 1$ by

$$\tau_m := \inf \left\{ t \geq \tau_{m-1} \mid \sum_{i=\tau_{m-1}+1}^{t} \mathbb{I}\{X_i \in B(x^\star, 2 \cdot D \cdot 2^{-m})\} = N'_m \right\}.$$

The stopping time τ_m correspond to the time after τ_{m-1} where we have recorded at least N'_m random evaluation points inside the ball $B(x^\star, 2 \cdot D \cdot 2^{-m})$. To prove the result, we show that each of the following events:

$$E_m := \left\{ \max_{i=1\ldots\tau_m} f(X_i) \geq \max_{x \in \mathcal{X}} f(x) - \frac{c_\kappa}{2} \cdot \left(\frac{D}{2^m}\right)^\kappa \right\} \cap \left\{ \tau_m \leq N'_1 + \sum_{l=1}^{m-1} (N'_{l+1} + N_l) \right\}.$$

holds true with probability at least $1 - \delta/M$ on the event $\bigcap_{l=1}^{m-1} E_l$ for all $m \in \{2, \ldots, M\}$ so that:

$$\mathbb{P}\left(E_M\right) \geq \mathbb{P}(E_1) \times \prod_{m=1}^{M-1} \mathbb{P}\left(E_{m+1} | \bigcap_{l=1}^{m} E_l\right) \geq \left(1 - \frac{\delta}{M}\right)^M \geq 1 - \delta \tag{4}$$

that will leads us to the result by analyzing E_M.

Analysis of $\mathbb{P}(E_1)$. Observe first that since $\mathcal{X} \subseteq B(x^\star, D)$, then $\tau_1 = N'_1$. Using now the fact that (i) the algorithm is faster than a Pure Random Search (Proposition 12) and (ii) the bound of Proposition 28, we directly get that with probability at least $1 - \delta/M$,

$$\begin{aligned}
\max_{x \in \mathcal{X}} f(x) - \max_{i=1\ldots\tau_1} f(X_i) &\leq k \cdot 2D \cdot \left(\frac{\ln(M/\delta)}{N'_1}\right)^{\frac{1}{d}} \\
&\leq k \cdot 2D \cdot \left(\frac{\ln(M/\delta)}{\ln(M/\delta) 2^{d(\kappa-1)}} \left(\frac{c_\kappa D^\kappa}{8kD}\right)^d\right)^{\frac{1}{d}} \\
&= \frac{c_\kappa}{2} \cdot \left(\frac{D}{2}\right)^\kappa
\end{aligned}$$

which proves that $\mathbb{P}(E_1) \geq 1 - \delta/M$.

Analysis of $\mathbb{P}(E_{m+1} | \cap_{l=1}^{m} E_l)$. To bound this term, we use (i) a deterministic covering argument to control the stopping time τ_{m+1} (Lemma 29 and Corollary 30) and (ii) a stochastic argument to bound the maximum $\max_{i=1\ldots\tau_{m+1}} f(X_i)$ (Lemma 31 and Corollary 32). The following lemma states that after τ_m and on the event E_m there will be at most N_m evaluation points that will fall inside the area $B(x^\star, 2D \cdot 2^{-m}) / B(x^\star, D \cdot 2^{-m})$.

Lemma 29 *For all $m \in \{1, \ldots, M-1\}$, we have on the event E_m,*

$$\sum_{t=\tau_m+1}^{n} \mathbb{I}\{X_t \in B(x^\star, 2D \cdot 2^{-m}) / B(x^\star, D \cdot 2^{-m})\} \leq N_m.$$

Proof Fix $m \in \{1, \ldots, M-1\}$ and assume that $E_m = \{\max_{i=1 \ldots \tau_m} f(X_i) \geq \max_{x \in \mathcal{X}} f(x) - c_k/2 \cdot (D/2^m)^\kappa\}$ holds true. Setting $N = \lceil \sqrt{d} 8kD2^{m(\kappa-1)}/(c_k D^\kappa) \rceil$ and observing that $B(x^\star, 2D \cdot 2^{-m}) \subseteq x^\star + 2D \cdot 2^{-m} \cdot [-1, +1]^d$, one can then introduce the sequence H_I, with $I \in \{1, \ldots, N\}^d$, of the $N^d = N_m$ hypercubes whose side have length $4D \cdot 2^{-m}/N$ and cover $x^\star + 2D \cdot 2^{-m} \times [-1, +1]^d$, so that

$$B(x^\star, 2D \cdot 2^{-m})/B(x^\star, D \cdot 2^{-m}) \subseteq x^\star + 2D \cdot 2^{-m} \cdot [-1, +1]^d = \bigcup_I H_I.$$

Based on these hypercubes, one can define the set

$$I_t = \{I \in \{1, \ldots, N\}^d : H_I \cap B(x^\star, 2D \cdot 2^{-m})/B(x^\star, 2D \cdot 2^{-m}) \cap \mathcal{X}_{k,t} \neq \emptyset\}$$

which contains the indexes of the hypercubes that still intersect the set of potential maximizers $\mathcal{X}_{k,t}$ at time t and the target area $B(x^\star, 2D \cdot 2^{-m})/B(x^\star, 2D \cdot 2^{-m})$. We show by contradiction that there cannot be more than $N^d = N_m$ evaluation points falling inside this area, otherwise it would be empty. Suppose that, after τ_m, there exists a sequence

$$\tau_m < t_1 < t_2 < \cdots < t_{N^d+1} \leq n$$

of $N^d + 1$ strictly increasing indexes for which the evaluation points X_{t_j}, $j \geq 1$, belong to the target area, *i.e,*

$$\forall j \in \{1, \ldots, N^d + 1\}, \; X_{t_j} \in B(x^\star, 2D \cdot 2^{-m})/B(x^\star, D \cdot 2^{-m}).$$

Fix any $j \geq 1$ and observe that since $X_{t_j} \notin B(x^\star, D \cdot 2^{-m})$, then we have from Condition 1 that (i) $f(X_{t_j}) < \max_{x \in \mathcal{X}} f(x) - c_k \cdot (D \cdot 2^{-m})^\kappa$. Moreover, as $X_{t_j} \in \mathcal{X}_{k,t_j-1} \cap B(x^\star, 2D \cdot 2^{-m})/B(x^\star, D \cdot 2^{-m})$, it necessarily follows from the definition of the algorithm that (ii) there exists an index $I^\star \in I_{t_j-1}$ such that $X_{t_j} \in H_{I^\star}$. Therefore, combining (i) and (ii) with E_m, gives that $\forall x \in H_{I^\star}$:

$$\begin{aligned}
f(x) &\leq f(X_{t_j}) + k \cdot \left\| X_{t_j} - x \right\|_2 && (f \in \mathrm{Lip}(k)) \\
&\leq f(X_{t_j}) + k \cdot \max_{(x,x') \in H_I^2} \|x - x'\|_2 && ((X_{t_j}, x) \in H_I^2) \\
&= f(X_{t_j}) + k \cdot \sqrt{d} \cdot 4D \cdot 2^{-m}/N && (\text{def. of } H_I) \\
&\leq f(X_{t_j}) + \frac{c_k}{2} \cdot (D \cdot 2^{-m})^\kappa && (\text{def. of } N) \\
&< \max_{x \in \mathcal{X}} f(x) - c_\kappa \cdot (D \cdot 2^{-m})^\kappa + \frac{c_k}{2} \cdot (D \cdot 2^{-m})^\kappa && (\text{i}) \\
&\leq \max_{i=1 \ldots \tau_m} f(X_i) && (E_1) \\
&\leq \max_{i=1 \ldots t_j} f(X_i). && (t_j > \tau_m)
\end{aligned}$$

It has been shown that if X_{t_j} belongs to the target area then $f(x) < \max_{i=1 \ldots t_j} f(X_i)$ for all $x \in H_{I^\star}$, which combined with the definition of the set of potential maximizers $\mathcal{X}_{k,t_j}$ at time t_j implies that $H_{I^\star} \notin \mathcal{X}_{\tau_j}$. Hence, once an evaluation has been made in $H_{I^\star}$, there will not be any future evaluation point falling inside this cube. We thus deduce that $|I_{t_j}| \leq |I_{t_j-1}| - 1$ for all $j \geq 1$ which leads us to the following contradiction:

$$0 \leq |I_{t_{N^d+1}}| = |I_{\tau_m}| + \sum_{j=\tau_m+1}^{t_{N^d+1}} |I_{t_j}| - |I_{t_{j-1}}| \leq |I_{\tau_m}| - (N^d+1) \leq N^d - (N^d+1) < 0$$

and proves the statement. $\square$

Based on this lemma, one might then derive a bound on the stopping time τ_{m+1}.

Corollary 30 *For all $m \in \{1, \ldots, M-1\}$, we have on the event $\bigcap_{l=1}^m E_l$ that*

$$\tau_{m+1} \leq N_1' + \sum_{l=1}^{m} \left(N_{l+1}' + N_l\right).$$

Proof The result is proved by induction. We start with the case where $m = 1$. Assuming that E_1 holds true and observing that (i) $\tau_1 = N_1'$ and (ii) $\mathcal{X} \subseteq B(x^\star, D) = B(x^\star, D/2) \cup B(x^\star, D)/B(x^\star, D/2)$, one can then write:

$$\begin{aligned}\tau_2 &= \tau_1 + \sum_{i=\tau_1+1}^{\tau_2} \mathbb{I}\{X_i \in B(x^\star, D)\} \\ &= N_1' + \sum_{i=\tau_1+1}^{\tau_2} \mathbb{I}\{X_i \in B(x^\star, D/2)\} + \sum_{i=\tau_1+1}^{\tau_2} \mathbb{I}\{X_i \in B(x^\star, D)/B(x^\star, D/2)\}.\end{aligned}$$

However, since (i) $\sum_{i=\tau_1+1}^{\tau_2} \mathbb{I}\{X_i \in B(x^\star, D/2)\} = N_2'$ by definition of τ_2 and (ii) $\sum_{i=\tau_1+1}^{\tau_2} \mathbb{I}\{X_i \in B(x^\star, D)/B(x^\star, D/2)\} \leq N_1$ by Lemma 29, the result holds true for $m = 1$. Consider now any $m \geq 2$ and assume that the statement holds true for all $l < m$. Again, observing that $\mathcal{X} \subseteq B(x^\star, D \cdot 2^{-m}) \cup \bigcup_{l=1}^m B(x^\star, D \cdot 2^{-(l-1)})/B(x^\star, D \cdot 2^{-l})$ and keeping in mind that the stopping times are bounded by the induction assumption, one can write

$$\begin{aligned}\tau_{m+1} = \tau_m &+ \sum_{i=\tau_m+1}^{\tau_{m+1}} \mathbb{I}\{X_i \in B(x^\star, D \cdot 2^{-m})\} \\ &+ \sum_{i=\tau_m+1}^{\tau_{m+1}} \sum_{l=1}^{m} \mathbb{I}\Big\{X_i \in B(x^\star, D \cdot 2^{-(l-1)})/B(x^\star, D \cdot 2^{-l})\Big\}.\end{aligned}$$

Now, combining the telescopic representation $\tau_{m+1} = \tau_1 + \sum_{l=1}^m (\tau_{l+1} - \tau_l)$ with the previous decomposition gives that

$$\begin{aligned}\tau_{m+1} = \tau_1 &+ \sum_{l=1}^{m} \sum_{\tau_l+1}^{\tau_{l+1}} \mathbb{I}\Big\{X_i \in B(x^\star, D \cdot 2^{-l})\Big\} \\ &+ \sum_{l=1}^{m} \sum_{i=\tau_l+1}^{\tau_{m+1}} \mathbb{I}\Big\{X_i \in B(x^\star, D \cdot 2^{-(l-1)})/B(x^\star, D \cdot 2^{-l})\Big\}.\end{aligned}$$

However, since (i) $\tau_1 = N'_1$, (ii) $\sum_{\tau_l+1}^{\tau_{l+1}} \mathbb{I}\left\{X_i \in B(x^\star, D\cdot 2^{-l})\right\} = N'_{l+1}$, for all $l \geq 1$ by definition of the stopping times and (iii) $\sum_{i=\tau_l+1}^{\tau_{m+1}} \mathbb{I}\left\{X_i \in B(x^\star, 2^{-(l-1)})/B(x^\star, 2^{-l})\right\} \leq N_l$, for all $l \geq 1$ on the event $\bigcap_{l=1}^m E_l$, from Lemma 29, we finally get that

$$\tau_{m+1} \leq N'_1 + \sum_{l=1}^{m}(N'_{l+1} + N_l).$$

□

As Corollary 30 gives the desired bound on τ_{m+1}, it remains to control the maximum $\max_{i=1\dots\tau_{m+1}} f(X_i)$. The next lemma shows that i.i.d. results can actually be used to bound this term.

Lemma 31 *For all $m \in \{1, \dots, M-1\}$, we have that $\forall y \in \mathrm{Im}(f)$,*

$$\mathbb{P}\left(\max_{i=1\dots\tau_{m+1}} f(X_i) \geq y \mid \bigcap_{l=1}^{m} E_l\right) \geq \mathbb{P}\left(\max_{i=1\dots N'_{m+1}} f(X'_i) \geq y\right).$$

where $X'_1 \dots X'_{N'_{m+1}}$ denotes a sequence N'_{m+1} i.i.d. copies of $X' \sim \mathcal{U}(\mathcal{X} \cap B(x^\star, D\cdot 2^{-m}))$.

Proof From Corollary 30, we know that on the event $\bigcap_{l=1}^m E_l$ the stopping time τ_{m+1} is finite. Moreover, as $\sum_{i=\tau_m+1}^{\tau_{m+1}} \mathbb{I}\{X_i \in B(x^\star, D\cdot 2^{-m})\} = N'_{m+1}$ by definition of τ_{m+1}, it can then easily be shown by reproducing the same steps as in the proof of Proposition 12 with the evaluations points falling into $B(x^\star, D\cdot 2^{-m})$ after τ_m that the algorithm is faster than a Pure Random Search performed over the subspace $\mathcal{X} \cap B(x^\star, D\cdot 2^{-m})$, which proves the result. □

As a direct consequence of this lemma, one can get the desired bound on the maxima as shown in the next corollary.

Corollary 32 *For all $m \in \{1, \dots, M-1\}$, we have that*

$$\mathbb{P}\left(\max_{i=1\dots\tau_{m+1}} f(X_i) \geq \max_{x\in\mathcal{X}} f(x) - \frac{c_k}{2}\cdot\left(\frac{D}{2^{m+1}}\right)^{\kappa} \mid \bigcap_{l=1}^{m} E_l\right) \geq 1 - \delta/M.$$

Proof Omitting the conditionning upon $\bigcap_{l=1}^m E_l$, we obtain from the combination of Lemma 31 and Proposition 28 that with probability at least $1-\delta/M$:

$$\begin{aligned}
\max_{x\in\mathcal{X}} f(x) - \max_{i=1\dots\tau_{m+1}} f(X_i) &\leq k\cdot 2D\cdot 2^{-m}\cdot\left(\frac{\ln(M/\delta)}{N'_{m+1}}\right)^{\frac{1}{d}} \\
&\leq k\cdot 2D\cdot 2^{-m}\cdot\left(\frac{\ln(M/\delta)}{\ln(M/\delta)2^{d(m+1)(\kappa-1)}}\left(\frac{c_\kappa D^\kappa}{8kD}\right)^{d}\right)^{\frac{1}{d}} \\
&= \frac{c_\kappa}{2}\cdot\left(\frac{D}{2^{m+1}}\right)^{\kappa}.
\end{aligned}$$

$\square$

At this point, we know from the combination of Corollary 30 and Corollary 32 that

$$\forall m \in \{1, \ldots, M-1\}, \quad \mathbb{P}\left(E_{m+1} | \bigcap_{l=1}^{m} E_l\right) \geq 1 - \delta/M$$

which proves from (4) that $\mathbb{P}(E_M) \geq 1 - \delta$.

Analysis of E_M. As $\max_{i=1\ldots\tau_M} f(X_i) \geq \max_{x\in\mathcal{X}} f(x) - \frac{c_\kappa}{2} \cdot D^\kappa \cdot 2^{-M\kappa}$ and $\tau_M \leq N_1' + \sum_{l=1}^{M-1}\left(N_{l+1}' + N_l\right)$ on the event E_M, it remains to show that $N_1' + \sum_{l=1}^{M-1}\left(N_{l+1}' + N_l\right) \leq n$ to conclude the proof. Consider first the case $\kappa = 1$. Setting $C = (8k/c_\kappa)^d$ and observing that (i) $N_l' \leq \ln(M/\delta)C + 1$, (ii) $N_l \leq 2 \cdot C \cdot (2\sqrt{d})^d - 1$ for all $l \leq M$ and (iii) $M \leq n$, one gets:

$$\begin{aligned}
N_1' + \sum_{l=1}^{M-1}\left(N_{l+1}' + N_l\right) &\leq C \cdot M\left(\ln(M/\delta) + 2(2\sqrt{d})^d\right) \\
&\leq n \cdot \frac{\ln(M/\delta) + 2(2\sqrt{d})^d}{\ln(n/\delta) + 2(2\sqrt{d})^d} \\
&\leq n.
\end{aligned}$$

For $\kappa > 1$, since (i) M was chosen so that $\frac{2^{d(\kappa-1)M}-1}{2^{d(\kappa-1)}-1} \leq \frac{n}{C} \cdot \frac{1}{\ln(n/\delta)+2(2\sqrt{d})^d}$ and (ii) $M \leq n$, we obtain:

$$\begin{aligned}
N_1' + \sum_{l=1}^{M-1}\left(N_{l+1}' + N_l\right) &\leq C \cdot \left(\ln(M/\delta) + 2(2\sqrt{d})^d\right)\sum_{l=1}^{M}(2^{d(\kappa-1)})^l \\
&\leq C \cdot \left(\ln(M/\delta) + 2(2\sqrt{d})^d\right) \cdot \frac{2^{d(\kappa-1)M} - 1}{2^{d(\kappa-1)} - 1} \\
&\leq n.
\end{aligned}$$

Finally, using the elementary inequality $\lfloor x \rfloor \geq x - 1$ over M and the inequality $c_\kappa D^\kappa \leq k\,\mathrm{diam}(\mathcal{X})$ (by Condition 1) leads to the desired result and completes the proof. $\square$

Proof of Theorem 16. (Lower bound) Pick any $n \in \mathbb{N}^\star$ and $\delta \in (0,1)$, set $\epsilon = c_\kappa\,\mathrm{rad}(\mathcal{X})^\kappa\,\delta^{\kappa/d}\exp(-\kappa(n - \sqrt{2n\ln(1/\delta)})/d)$, let $\mathcal{X}_\epsilon = \{x \in \mathcal{X} : f(x) \geq \max_{x\in\mathcal{X}} f(x) - \epsilon\}$ be the corresponding level set. Observe first that since (i) $\mathcal{X}_\epsilon = \{x \in \mathcal{X} : \epsilon \geq f(x^\star) - f(x)\} \subseteq \{x \in \mathcal{X} : \epsilon \geq c_\kappa \left\|x^\star - x\right\|_2^\kappa\} = \mathcal{X} \cap B(x^\star, (\epsilon/c_\kappa)^{1/\kappa}$ and (ii) there exists $x \in \mathcal{X}$ such that $B(x, \mathrm{rad}(\mathcal{X})) \subseteq \mathcal{X}$, then $\mu(\mathcal{X}_\epsilon)/\mu(\mathcal{X}) \leq ((\epsilon/c_\kappa)^{1/\kappa}/\,\mathrm{rad}(\mathcal{X}))^d = \delta e^{-n-\sqrt{2n\ln(1/\delta)}}$. It can then easily be shown by reproducing the same steps as in the proof of the Lower bound of

Theorem 17 in (Malherbe and Vayatis (2016)) that

$$\begin{aligned}\mathbb{P}\left(\max_{i=1\dots n} f(X_i) \geq \max_{x\in\mathcal{X}} f(x) - \epsilon\right) &= \mathbb{P}\left(\frac{\mu(\{x \in \mathcal{X} : f(x) \geq \max_{i=1\dots n} f(X_i)\})}{\mu(\mathcal{X})} \leq \frac{\mu(\mathcal{X}_\epsilon)}{\mu(\mathcal{X})}\right)\\
&\leq \mathbb{P}\left(\prod_{i=1}^{n} U_i \leq \frac{\mu(\mathcal{X}_\epsilon)}{\mu(\mathcal{X})}\right)\\
&\leq \mathbb{P}\left(\prod_{i=1}^{n} U_i \leq \delta \cdot e^{-n-\sqrt{2n\ln(1/\delta)}}\right)\\
&= \mathbb{P}\left(\sum_{i=1}^{n} -\ln(U_i) > n + \sqrt{2n\ln(1/\delta)} + \ln(1/\delta)\right)\\
&\leq \delta\end{aligned}$$

where $U_1, \dots, U_n$ denotes a sequence of n i.i.d. copies of $U \sim \mathcal{U}([0,1])$. We point out that a concentration results for gamma random variable was used on the last line (see Lemma 37 and Lemma 38 in (Malherbe and Vayatis (2016)) for more details). □

Appendix D. Analysis of AdaLipOpt (proofs of Section 4)

Proof of Proposition 18. Pick any $t \geq 2$, consider any non-constant $f \in \bigcup_{k\geq 0} \mathrm{Lip}(k)$ and set $i^\star = \min\{i \in \mathbb{Z} : f \in \mathrm{Lip}(k_i)\}$. To prove the result, we decorelate the sample and use the fact that $(X_1, X_{\lfloor t/2\rfloor+1}), \dots, (X_{\lfloor t/2\rfloor}, X_{2\lfloor t/2\rfloor})$ forms a sequence of $\lfloor t/2\rfloor$ i.i.d. copies of $(X, X') \sim \mathcal{U}(\mathcal{X} \times \mathcal{X})$:

$$\begin{aligned}\mathbb{P}\left(f \in \mathrm{Lip}(\hat{k}_t)\right) &= \mathbb{P}\left(\hat{k}_t = k_{i^\star}\right)\\
&= \mathbb{P}\left(\bigcup_{i\neq j}^{t} \{|f(X_i) - f(X_j)| > k_{i^\star-1} \cdot \|X_i - X_j\|_2\}\right)\\
&\geq \mathbb{P}\left(\bigcup_{i=1}^{\lfloor t/2\rfloor} \left\{\left|f(X_i) - f(X_{\lfloor t/2\rfloor+i})\right| > k_{i^\star-1} \cdot \left\|X_i - X_{\lfloor t/2\rfloor+i}\right\|_2\right\}\right)\\
&= 1 - \mathbb{P}\left(\frac{|f(X_1) - f(X_2)|}{\|X_1 - X_2\|_2} \leq k_{i^\star-1}\right)^{\lfloor t/2\rfloor}\\
&= 1 - (1 - \Gamma(f, k_{i^\star-1}))^{\lfloor t/2\rfloor}.\end{aligned}$$

It remains to show that $\Gamma(f, k_{i^\star-1}) > 0$. Observe first that since $f \in \mathrm{Lip}(k_{i^\star})$, then the function $F : (x, x') \mapsto |f(x) - f(x')| - k_{i^\star-1} \cdot \|x - x'\|_2$ is also continuous. However, as $f \notin \mathrm{Lip}(k_{i^\star-1})$, we know that there exists some $(x_1, x_2) \in \mathcal{X} \times \mathcal{X}$ such that $F(x_1, x_2) > 0$. Hence, it follows from the continuity of F that there necessarily exists some $\epsilon > 0$ such that $\forall (x, x') \in B(x_1, \epsilon) \cap \mathcal{X} \times B(x_2, \epsilon) \cap \mathcal{X}$, $F(x, x') > 0$ which proves the proof. □

Proof of Proposition 21. Combining the consistency equivalence of Proposition 3 with the upper bound on the covering rate obtained in Example 4 gives the result. □

Proof of Proposition 22. Fix any $\delta \in (0,1)$, set $N_1 = 2 + \lceil 2\ln(\delta/3)/\ln(1-\Gamma(f,k_{i^\star-1}))\rceil$ and $N_2 = \lceil((\sqrt{\ln(3/\delta)/2 + 4N_1 p} - \sqrt{\ln(3/\delta)/2})/2p)^2\rceil$. Considering any $n > N_2$, we prove the result in three steps.
Step 1. As the constant N_1 and N_2 were chosen so that Hoeffding's inequality ensures that $\mathbb{P}\left(\sum_{i=1}^{N_2} B_i \geq N_1\right) \geq 1-\delta/3$, we know that after N_2 iterations and with probability $1-\delta/3$ we have collected at least N_1 evaluation points randomly and uniformly distributed over $\mathcal{X}$ due to the exploration step.

Step 2. Using Proposition 18 and the fist N_1 evaluation points which have been sampled independently and uniformly over $\mathcal{X}$, we know that after N_2 iterations and on the event $\{\sum_{i=1}^{N_2} B_i \geq N_1\}$ the Lipschtz constant $k_{i^\star}$ has been estimated with probability at least $1-\delta/3$, *i.e.*, $\mathbb{P}\left(\forall t \geq N_2+1,\ \hat{k}_t = k_{i^\star} \mid \sum_{i=1}^{N_2} B_i \geq N_1\right) \geq 1-\delta/3$.

Step 3. Finally, as the Lipschtz constant estimate $\hat{k}_t$ satisfies $f \in \mathrm{Lip}(\hat{k}_t)$ for all $t \geq N_2+1$ on the above event, one can easily show by reproducing the same steps as in Proposition 12 that contionned upon the event $\{\forall t \geq N_2+1,\ \hat{k}_t = k_{i^\star}\} \cap \{\sum_{i=1}^{N_2} B_i \geq N_1\}$ the algorithm is always faster or equal to a Pure Random Search ran with $n - N_2$ i.i.d. copies of $X' \sim \mathcal{U}(\mathcal{X})$. Therefore, using (i) the bound of Proposition 28, (ii) the elementaries inequalities $\lceil x \rceil \leq x+1$, $\lfloor x \rfloor \geq x-1$, $\sqrt{x+y} - \sqrt{x} \leq \sqrt{y}$ and (iv) the definition of $N_2 < n$, we obtain that with probability at least $(1-\delta/3)^3 \geq 1-\delta$,

$$
\begin{aligned}
\max_{x\in\mathcal{X}} f(x) - \max_{i=1\dots n} f(X_i) &\leq k_{i^\star} \cdot \mathrm{diam}(\mathcal{X}) \cdot \left(\frac{\ln(3/\delta)}{n-N_2}\right)^{\frac{1}{d}} \\
&= k_{i^\star} \cdot \mathrm{diam}(\mathcal{X}) \cdot \left(\frac{n}{n-N_2}\right)^{\frac{1}{d}} \cdot \left(\frac{\ln(3/\delta)}{n}\right)^{\frac{1}{d}} \\
&\leq k_{i^\star} \cdot \mathrm{diam}(\mathcal{X}) \cdot (1+N_2)^{\frac{1}{d}} \left(\frac{\ln(3/\delta)}{n}\right)^{\frac{1}{d}} \\
&\leq k_{i^\star} \cdot \mathrm{diam}(\mathcal{X}) \cdot \left(\frac{5}{p} + \frac{2\ln(\delta/3)}{p\ln(1-\Gamma(f,k_{i^\star-1}))}\right)^{\frac{1}{d}} \cdot \left(\frac{\ln(3/\delta)}{n}\right)^{\frac{1}{d}}.
\end{aligned}
$$

The result is extended to the case where $n \leq N_2$ by noticing that the bound is superior to $k_{i^\star} \cdot \mathrm{diam}(\mathcal{X})$ in that case, and thus trivial. □

Proof of Theorem 23. Fix $\delta \in (0,1)$, set $N_1 = 2 + \lceil 2\ln(4/\delta)/\ln(1-\Gamma)\rceil$ and $N_2 = \lceil((\sqrt{\ln(4/\delta)/2 + 4N_1 p} - \sqrt{\ln(4/\delta)/2})/2p)^2\rceil$ and let $N_3 = N_2 + \lceil 2\ln(4/\delta)/(1-p)^2\rceil$. Picking any $n > N_3$, we proceed similarly as in the proof of Proposition 22 in four steps:

Steps 1 & 2. As in the above prove, by definition of N_1 and N_2 and due to Hoeffding's inequality and Proposition 18, we know that the following event: $\{\forall t \geq N_2+1, \hat{k}_t = k_{i^\star}\} \cap \{\sum_{i=1}^{N_2} B_i \geq N_1\}$ holds true with probability at least $(1-\delta/4)^2$.

Step 3. Again, using Hoeffding's inequality and the definition of N_2 and N_3, we know that after the iteration $N_2 + 1$ we have collected with probability at least $1-\delta/4$ at least

$(1-p)(n-N_3)/2$ exploitative evaluation points:

$$\sum_{i=N_2+1}^{n} \mathbb{I}\{B_i = 0\} \geq (1-p)(n-N_2) - \sqrt{\frac{(n-N_2)\ln(4/\delta)}{2}} \geq \frac{1-p}{2} \cdot (n-N_3).$$

Step 4. Reproducing the same steps as in the proof of the fast rate of Theorem 15 with the $(1-p)\cdot(n-N_3)/2$ previous exploitative points and putting the previous results altogether gives that with proability at least $(1-\delta/4)^4 \geq 1-\delta$,

$$\max_{x\in\mathcal{X}} f(x) - \max_{i=1\dots n} f(X_i) \leq k_{i^\star} \times \mathrm{diam}(\mathcal{X}) \times$$

$$\begin{cases} \exp\left\{ - C_{k,\kappa} \cdot \dfrac{(1-p)(n-N_3)\ln(2)}{2\ln(4n/\delta) + 4(2\sqrt{d})^d} \right\}, & \kappa = 1 \\ \\ \dfrac{2^\kappa}{2}\left(1 + C_{k,\kappa} \cdot \dfrac{(1-p)(n-N_3)(2^{d(\kappa-1)}-1)}{2\ln(4n/\delta)+4(2\sqrt{d})^d}\right)^{-\frac{\kappa}{d(\kappa-1)}}, & \kappa > 1. \end{cases}$$

We now take out the term N_3. Since $C_{k_{i^\star},\kappa}(1-p) \leq 1$, when $\kappa = 1$, we have

$$\max_{x\in\mathcal{X}} f(x) - \max_{i=1\dots n} f(X_i) \leq k_{i^\star} \cdot \mathrm{diam}(\mathcal{X}) \cdot \exp(5N_3/2) \exp\left\{ - C_{k,\kappa} \cdot \frac{(1-p)n\ln(2)}{2\ln(4n/\delta)+4(2\sqrt{d})^d}\right\}.$$

For $\kappa > 1$, setting $C = C_{k_{i^\star},\kappa}(1-p)/(2\ln(4n/\delta) + 4(2\sqrt{d})^d)$ and using the decomposition $n = (n-N_3) + N_3$, we bound the ratio:

$$\left(\frac{1 + Cn(2^{d(\kappa-1}-1)}{1 + C(n-N_3)(2^{d(\kappa-1)}-1)}\right)^{\frac{\kappa}{d(\kappa-1)}} \leq \left(1 + \frac{CN_3(2^{d(\kappa-1)}-1)}{1 + C(2^{d(\kappa-1)}-1)}\right)^{\frac{\kappa}{d(\kappa-1)}}.$$

In the case where $\kappa/d(\kappa-1) \leq 1$, one directly obtains

$$\left(1 + \frac{CN_3(2^{d(\kappa-1)}-1)}{1 + C(2^{d(\kappa-1)}-1)}\right)^{\frac{\kappa}{d(\kappa-1)}} \leq (1+N_3)^{\frac{\kappa}{d(\kappa-1)}} \leq (1+N_3) \leq e^{N_3}$$

Considering the case where $\kappa/d(\kappa-1) > 1$ and setting $\kappa = 1+\epsilon/d$ with $\epsilon \in (0,1)$, we obtain from the inequalities (i) $\kappa \leq 1 + 1/d \leq 2$, (ii) $\forall \epsilon \in (0,1)$, $2^\epsilon - 1 \leq \epsilon$ and (iii) $C \leq 1/2$ that

$$\left(1 + \frac{CN_3(2^{d(\kappa-1)}-1)}{1 + C(2^{d(\kappa-1)}-1)}\right)^{\frac{\kappa}{d(\kappa-1)}} \leq (1 + CN_3(2^\epsilon - 1))^{\frac{2}{\epsilon}} \leq (1 + CN_3\epsilon)^{\frac{2}{\epsilon}} \leq e^{2CN_3} \leq e^{N_3}$$

Finally, using standard bounds on N_3 and noticing that the previous bound is superior to $k_{i^\star}\,\mathrm{diam}(\mathcal{X})$ whenever $n \leq N_3$, the previous result remains valid for any $n \in \mathbb{N}^\star$. ☐