A. Proofs of Section 2

Proof of proposition 1. (RANKING EQUIVALENCE) Let $h \in \mathcal{C}^0(\mathcal{X}, \mathbb{R})$ be any continuous function defined on $\mathcal{X}$ and taking values in $\mathbb{R}$ and let $f : \mathcal{X} \to \mathbb{R}$ be any function defined on $\mathcal{X}$ and taking values in $\mathbb{R}$.

($\Leftarrow$) Assume that there exists a strictly increasing function $\psi : \mathbb{R} \to \mathbb{R}$ such that $h = \psi \circ f$. Pick any $(x, x') \in \mathcal{X}^2$ and let sgn : $\mathbb{R} \to \{-1, 0, 1\}$ be the standard sign function defined by $\text{sgn}(x) = \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$. Using the fact that $h = \psi \circ f$ where ψ is a strictly increasing function directly gives that,

$$r_h(x, x') = \text{sgn}(\psi \circ f(x) - \psi \circ f(x')) = \text{sgn}(f(x) - f(x')) = r_f(x, x').$$

($\Rightarrow$) Assume that $r_f(x, x') = r_h(x, x')$ for all $(x, x') \in \mathcal{X}^2$. First case: if $r_f(x, x') = r_h(x, x') = 0$ for all $(x, x') \in \mathcal{X}^2$, $f = c$ and $h = c'$ are constant over $\mathcal{X}$ and therefore $h = \psi \circ f$ where $\psi : x \mapsto x + (c' - c)$ is a strictly increasing function. Second case: assume that h is not a constant function and introduce the function $M : \mathcal{X} \to [0, 1]$ defined by

$$M(x) = \int_{x' \in \mathcal{X}} \mathbb{1}\{r_f(x, x') < 0\}\, dx' = \mu(\{x' \in \mathcal{X} : f(x') < f(x)\}).$$

We start to show that there exists a strictly increasing function $\psi : \mathbb{R} \to \mathbb{R}$ such that $f = \psi \circ M$. To properly define this function, we first show (by contradiction) that for any pair of points $(x_1, x_2) \in \mathcal{X}^2$ satisfying $M(x_1) = M(x_2)$ necessarily $f(x_1) = f(x_2)$. Let $(x_1, x_2) \in \mathcal{X}^2$ be a pair of points satisfying $M(x_1) = M(x_2)$ and assume that $f(x_1) < f(x_2)$. The equality of the rankings implies that (i) $h(x_1) < h(x_2)$, (ii) $M(x_1) = \mu(\{x : h(x) < h(x_1)\})$ and (iii) $M(x_2) = \mu(\{x : h(x) < h(x_2)\})$. Putting (i), (ii) and (iii) altogether and using the continuity of the function h leads to the next contradiction:

$$M(x_1) = \mu(\{x \in \mathcal{X} : h(x) < h(x_1)\}) < \mu(\{x \in \mathcal{X} : h(x) < h(x_2)\}) = M(x_2).$$

Assuming that $f(x_2) < f(x_1)$ leads to a similar contradiction. We deduce that for any $(x_1, x_2) \in \mathcal{X}^2$ satisfying $M(x_1) = M(x_2)$, necessarily $f(x_1) = f(x_2)$. As a direct consequence, for any $y \in \text{Im}(M)$, the function f is constant over the iso level set $M^{-1}(y) = \{x \in \mathcal{X} : M(x) = y\}$ of the function M. We are now ready to introduce the function $\bar{\psi} : \text{Im}(M) \to \mathbb{R}$ defined by,

$$\bar{\psi} : y \mapsto f(x) \text{ where } x \in M^{-1}(y).$$

Now, note that for any $x \in \mathcal{X}$ we have that $\bar{\psi}(M(x)) = f(x)$. Using the same statement gives us that $\bar{\psi}(y_1) < \bar{g}(y_2)$ for any $y_1 < y_2 \in \text{Im}(M)^2$. Therefore, $f = \psi \circ M$ where $\psi : \mathbb{R} \to \mathbb{R}$ is any strictly increasing extension of the function $\bar{\psi} : \text{Im}(M) \to \mathbb{R}$ over $\mathbb{R}$. Reproducing the same steps with the function h gives us that there exists a strictly increasing function $\psi' : \mathbb{R} \to \mathbb{R}$ such that $h = \psi' \circ M$. Combining those results, we finally get that $h = \psi' \circ M = (\psi' \circ \psi^{-1}) \circ f$ where $(\psi' \circ \psi^{-1})$ is a strictly increasing function. □

B. Proofs of Section 3

In the first subsection, we recall the main definitions and provide some technical lemmas that will be used to prove the main statements presented in the second subsection.

B.1. RankOpt process, PAS process and technical lemmas

Definition 5. (RANKOPT PROCESS) *Let $\mathcal{X} \subset \mathbb{R}^d$ be any compact and convex set, let $\mathcal{R} \subseteq \mathcal{R}_\infty$ be any set of rankings and let $f : \mathcal{X} \to \mathbb{R}$ be any function such that $r_f \in \mathcal{R}$. The sequence $\{X_i\}_{i=1}^n$ is distributed as a* RANKOPT *process if the sequence has the same distribution as the process defined by:*

$$\begin{cases} X_1 \sim \mathcal{U}(\mathcal{X}) \\ X_{t+1} |\, \{X_i\}_{i=1}^t \sim \mathcal{U}(\mathcal{X}_t) \qquad \forall t \in \{1 \dots n-1\} \end{cases}$$

where $\mathcal{X}_t = \{x \in \mathcal{X} : \exists r \in \mathcal{R}_t,\;\; r(x, X_{\hat{\imath}_t}) \geq 0\}$ denotes the sampling area of the RANKOPT *algorithm at iteration $t+1$, $\mathcal{R}_t = \{r \in \mathcal{R} : L_t(r) = 0\}$ denotes the set of rankings that are still consistent with the sample and $\hat{\imath}_t \in \arg\max_{i=1\dots t} f(X_i)$ denotes the index of the best value observed so far.*

Lemma 1. *Using the same notations and assumptions as in Definition 5, for all $t \in \{1 \dots n\}$, we have that*

$$\mathcal{X}_{f(X_{\hat{i}_t})} \subseteq \mathcal{X}_t \subseteq \mathcal{X},$$

where $\mathcal{X}_t$ denotes the sampling area of the RANKOPT *process after t iterations (see Definition 5) and $\mathcal{X}_{f(X_{\hat{i}_t})} = \{x \in \mathcal{X} : f(x) \geq f(X_{\hat{i}_t})\}$ denotes the level set of the best value observed so far.*

Proof. Noticing that $\mathcal{X}_t$ is a subset of $\mathcal{X}$ gives the first inclusion. We now state the second inclusion. Since the true ranking r_f always perfectly ranks the sample, we have that $r_f \in \{r \in \mathcal{R} : L_t(r) = 0\} = \mathcal{R}_t$ for any $t \in \{1 \dots n\}$. Now, pick any $x \in \{x \in \mathcal{X} : f(x) \geq f(X_{\hat{i}_t})\}$. Using the definition of the true ranking r_f, we get that $r_f(x, X_{\hat{i}_t}) = \operatorname{sgn}(f(x) - f(X_{\hat{i}_t})) \geq 0$. Therefore, there exists $r = r_f \in \mathcal{R}_t$ such that $r(x, X_{\hat{i}_t}) \geq 0$ and we deduce that $\{x \in \mathcal{X} : f(x) \geq f(X_{\hat{i}_t})\} \subseteq \{x \in \mathcal{X} : \exists r \in \mathcal{R}_t \text{ s.t. } r(x, X_{\hat{i}_t}) \geq 0\}$. □

Definition 6. (PURE ADAPTIVE SEARCH PROCESS *(Zabinsky & Smith, 1992)*). *Let $f : \mathcal{X} \to \mathbb{R}$ be any function such that $r_f \in \mathcal{R}$. The sequence $\{X_i^\star\}_{i=1}^n$ is distributed as a* PURE ADAPTIVE SEARCH (PAS) *process if the sequence has the same distribution as the Markov process defined by:*

$$\begin{cases} X_1^\star \sim \mathcal{U}(\mathcal{X}) \\ X_{t+1}^\star | \, X_t^\star \sim \mathcal{U}(\mathcal{X}_t^\star) \quad \textit{for } \; t \in \{1 \dots n-1\}, \end{cases}$$

where $\mathcal{X}_t^\star = \{x \in \mathcal{X} : f(x) \geq f(X_t^\star)\}$ denotes the sampling area of the PAS *process (which is also the level set of the best value observed so far).*

Lemma 2. *Let $\{X_i^\star\}_{i=1}^n$ be a sequence of n random variables distributed as the* PAS *process (Definition 6). Then, for any $n \in \mathbb{N}^\star$, we have that,*

$$\mathbb{P}\left(\frac{\mu(\mathcal{X}_n^\star)}{\mu(\mathcal{X})} \leq u\right) \leq \mathbb{P}\left(\prod_{i=1}^n U_i \leq u\right), \;\; \forall u \in [0, 1],$$

where $\mathcal{X}_n^\star = \{x \in \mathcal{X} : f(x) \geq f(X_n^\star)\}$ denotes the sampling area of the PAS *process after n iterations and $\{U_i\}_{i=1}^n$ is a sequence of n independent copies of $U \sim \mathcal{U}([0, 1])$.*

Proof. The lemma is proved by induction. We start to set some notations. For any $u \in [0, 1]$, define $\bar{u} = \min\{u' \in \operatorname{Im}(f) : \mu(\mathcal{X}_{u'}) \leq u \cdot \mu(\mathcal{X})\}$ and let $\mathcal{X}_{\bar{u}} = \{x \in \mathcal{X} : f(x) \geq \bar{u}\}$ be the corresponding level set. By definition, we have that $\mu(\mathcal{X}_{\bar{u}}) \leq u \cdot \mu(\mathcal{X})$. We are now ready to prove the statement.

•$(n = 1)$ Fix any $u \in [0, 1]$ and let $n = 1$. Since $X_1^\star \sim \mathcal{U}(\mathcal{X})$, we have that,

$$\mathbb{P}\left(\frac{\mu(\mathcal{X}_1^\star)}{\mu(\mathcal{X})} \leq u\right) = \mathbb{P}(X_1^\star \in \mathcal{X}_{\bar{u}}) = \frac{\mu(\mathcal{X}_{\bar{u}})}{\mu(\mathcal{X})} \leq u.$$

Now, let $U_1 \sim \mathcal{U}([0, 1])$ be a random variable uniformly distributed over $[0, 1]$. By definition, we have that $u = \mathbb{P}(U_1 \leq u)$, therefore the result holds for $n = 1$.

•(Induction) Assume that the statement holds for a fixed $n \in \mathbb{N}^\star$ and fix any $u \in [0, 1]$. Conditioning on $X_n^\star$ and using the fact that $X_{n+1}^\star | X_n^\star \sim \mathcal{U}(\mathcal{X}_n^\star)$ (Definition 6) gives us that,

$$\mathbb{P}\left(\frac{\mu(\mathcal{X}_{n+1}^\star)}{\mu(\mathcal{X})} \leq u\right) = \mathbb{E}[\mathbb{P}(X_{n+1}^\star \in \mathcal{X}_{\bar{u}} | X_n^\star)] = \mathbb{E}\left[\frac{\mu(\mathcal{X}_{\bar{u}} \cap \mathcal{X}_n^\star)}{\mu(\mathcal{X}_n^\star)}\right].$$

By definition and using the inclusion of the level sets, we have the following equivalences on the events, $\{f(X_n^\star) \geq \bar{u}\} = \{\mathcal{X}_n^\star \subseteq \mathcal{X}_{\bar{u}}\}$ and $\{f(X_n^\star) \leq \bar{u}\} = \{\mathcal{X}_{\bar{u}} \subseteq \mathcal{X}_n^\star\}$. Therefore, using those equivalences gives us that $\mu(\mathcal{X}_{\bar{u}} \cap \mathcal{X}_n^\star)/\mu(\mathcal{X}_n^\star) = \min(1, \mu(\mathcal{X}_{\bar{u}})/\mu(\mathcal{X}_n^\star))$ and so,

$$\mathbb{P}\left(\frac{\mu(\mathcal{X}_{n+1}^\star)}{\mu(\mathcal{X})} \leq u\right) = \mathbb{E}\left[\min\left(1, \frac{\mu(\mathcal{X}_{\bar{u}})}{\mu(\mathcal{X}_n^\star)}\right)\right].$$

Now, let $U_{n+1} \sim \mathcal{U}([0,1])$ be a random variable uniformly distributed over $[0,1]$, independent of $X_n^\star$. By definition, we have that $\mathbb{P}(U_{n+1} \leq \mu(\mathcal{X}_{\bar{u}})/\mu(\mathcal{X}_n^\star)|X_n^\star) = \min(1, \mu(\mathcal{X}_{\bar{u}})/\mu(\mathcal{X}_n^\star))$. Therefore, using independence, we get that,

$$\mathbb{P}\left(\frac{\mu(\mathcal{X}_{n+1}^\star)}{\mu(\mathcal{X})} \leq u\right) = \mathbb{E}\left[\mathbb{P}\left(U_{n+1} \leq \frac{\mu(\mathcal{X}_{\bar{u}})}{\mu(\mathcal{X}_n^\star)} \mid X_n^\star\right)\right] \overset{\text{ind.}}{=} \mathbb{P}\left(U_{n+1} \cdot \frac{\mu(\mathcal{X}_n^\star)}{\mu(\mathcal{X})} \leq \frac{\mu(\mathcal{X}_{\bar{u}})}{\mu(\mathcal{X})}\right) \leq \mathbb{P}\left(U_{n+1} \cdot \frac{\mu(\mathcal{X}_n^\star)}{\mu(\mathcal{X})} \leq u\right).$$

Plugging the induction assumption into the last equation and using independence gives the result. □

Lemma 3. *Let $\{U_i\}_{i=1}^n$ be a sequence of n i.i.d. copies of $U \sim \mathcal{U}([0,1])$. Then, for any $\delta \in (0,1)$, we have that,*

$$\mathbb{P}\left(\prod_{i=1}^n U_i < \delta \cdot e^{-n-\sqrt{2n\ln(1/\delta)}}\right) < \delta.$$

Proof. The result is a consequence of concentration results of sub-gamma random variables. Taking the logarithm on both sides of the inequality gives us that $\prod_{i=1}^n U_i < \delta \cdot e^{-n-\sqrt{2n\ln(1/\delta)}} \Leftrightarrow \sum_{i=1}^n -\ln(U_i) > n + \sqrt{2n\ln(1/\delta)} + \ln(1/\delta)$. Since $U_i \sim \mathcal{U}([0,1])$, we have that $-\ln(U_i) \sim \text{Exp}(1)$ and therefore $\sum_{i=1}^n -\ln(U_i) \sim \text{Gamma}(n,1)$ by independence. We finally get the result by applying concentration results of sub-gamma random variables (see (Boucheron et al., 2013)). □

Lemma 4. *Let $\mathcal{X} \subset \mathbb{R}^d$ be any compact and convex set, let $f \in \mathcal{C}^0(\mathcal{X}, \mathbb{R})$ be any continuous function that satisfies the level set assumption (α, c_α) and fix any $r \in (0, \max_{x\in\mathcal{X}} \|x^\star - x\|_2)$. Then, denoting $\mathcal{S}_r = \{x \in \mathcal{X} : \|x^\star - x\|_2 = r\}$ the intersection of $\mathcal{X}$ with the ℓ_2-sphere of radius r centered around $x^\star$, we have that,*

$$\mathcal{X} \cap B(x^\star, (r/c_\alpha)^{1+\alpha}) \subseteq \{x \in \mathcal{X} : f(x) \geq \min_{x_r \in \mathcal{S}_r} f(x_r)\} \subseteq B(x^\star, c_\alpha \cdot r^{1/(1+\alpha)}).$$

Proof. Fix any $x_r \in \arg\min_{x\in\mathcal{S}_r} f(x)$, fix any $y \in [f(x_r), f(x^\star)]$ and let $f^{-1}(y) = \{x \in \mathcal{X} : f(x) = y\}$ be the corresponding iso level set. We first show that there exists $x_y \in f^{-1}(y)$ such that $\|x^\star - x_y\| \leq r$. Introduce the function $F_{x_r,x^\star} : [0,1] \to \mathbb{R}$ that returns the value of the function f over the segment $[x^\star, x_r]$, defined by,

$$F_{(x^\star,x_r)} : \lambda \mapsto f((1-\lambda)x^\star + \lambda x_r).$$

The convexity of the subset $\mathcal{X}$ and the continuity of the function f imply that the function $F_{(x^\star,x_r)}$ is well-defined and continuous. Since $F_{(x^\star,x_r)}(0) = f(x^\star)$ and $F_{(x^\star,x_r)}(1) = f(x_r)$, applying the intermediate value theorem gives us that there exists $\lambda_y \in [0,1]$ such that $F_{x_r}(\lambda_y) = y$. Therefore, there exists $x_y = \lambda_y x^\star + (1-\lambda_y)x_r \in f^{-1}(y)$ such that $\|x^\star - x_y\|_2 \leq \|x^\star - x_r\|_2 = r$. We now show the second inclusion (by contradiction). Assume that there exist $x_y' \in f^{-1}(y)$ such that $\|x^\star - x_y'\|_2 > c_\alpha r^{1/(1+\alpha)}$. It implies that $\max_{x\in f^{-1}(y)} \|x^\star - x\|_2 \geq \|x^\star - x_y'\|_2 > c_\alpha r^{1/(1+\alpha)}$. On the other hand, we have that $c_\alpha \min_{x\in f^{-1}(y)} \|x^\star - x\|_2^{1/(1+\alpha)} \leq c_\alpha \|x^\star - x_y\|_2^{1/(1+\alpha)} \leq c_\alpha r^{1/(1+\alpha)}$. Combining the previous inequalities together with the level set assumption leads us to a contradiction :

$$\max_{x\in f^{-1}(y)} \|x^\star - x\|_2 \leq c_\alpha \cdot \min_{x\in f^{-1}(y)} \|x^\star - x\|_2^{1/(1+\alpha)} < \max_{x\in f^{-1}(y)} \|x^\star - x\|_2.$$

The contradiction holds for any $y \in [f(x_r), f(x^\star)]$. We deduce that $\{x \in \mathcal{X} : f(x) \geq \min_{x\in S_r} f(x_r)\} \subseteq B(x^\star, c_\alpha \cdot r^{1/(1+\alpha)})$.

We use similar arguments to prove the first inclusion. Assume that there exists $x' \in \mathcal{X} \cap B(x^\star, (r/c_\alpha)^{1+\alpha})$ such that $f(x) < f(x_r)$. Introducing the function $F_{(x^\star,x')} : \lambda \mapsto f((1-\lambda)x^\star + \lambda x')$ and reproducing the same steps as previously gives us that there exists $x_r' \in f^{-1}(f(x_r))$ such that $\|x^\star - x_r'\|_2 < (r/c_\alpha)^{1+\alpha}$. Hence, $c_\alpha \cdot \min_{x\in f^{-1}(f(x_r))} \|x^\star - x\|_2^{1/(1+\alpha)} \leq c_\alpha \|x^\star - x_r'\|_2^{1/(1+\alpha)} < r$. On the other hand, we have that $\max_{x\in f^{-1}(f(x_r))} \|x^\star - x\|_2 \geq \|x^\star - x_r\|_2 = r$. We get a similar contradiction and we deduce that $\mathcal{X} \cap B(x^\star, (d/c_\alpha)^{1+\alpha}) \subseteq \{x \in \mathcal{X} : f(x) \geq \min_{x_r\in\mathcal{S}_r} f(x_r)\}$. □

Lemma 5. *(From (Zabinsky & Smith, 1992)). Let $\mathcal{X} \subset \mathbb{R}^d$ be any compact and convex set. Then, for any $x^\star \in \mathcal{X}$ and any $r \in (0, \operatorname{diam}(\mathcal{X}))$, we have that,*

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

Proof. Introduce the similarity transformation $\lambda : \mathbb{R}^d \to \mathbb{R}^d$ defined by

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

Denote by $\lambda(\mathcal{X}) = \{\lambda(x) : x \in \mathcal{X}\}$ the image of the subset $\mathcal{X}$ after the similarity transformation. Using the convexity of $\mathcal{X}$ and the fact that $\max_{x \in \mathcal{X}} \|x^\star - x\|_2 \leq \operatorname{diam}(\mathcal{X})$, we have that $\lambda(\mathcal{X}) \subseteq B(x^\star, r) \cap \mathcal{X}$. Hence, $\mu(B(x^\star, r) \cap \mathcal{X}) \geq \mu(\lambda(\mathcal{X}))$. Now, using the fact that λ is a similarity transformation and conserves the ratios of the volumes before/after transformation, we get that,

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

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

B.2. Proofs of the main results

Proof of Proposition 2. The statement is proved by induction. Since $X_1 \sim \mathcal{U}(\mathcal{X})$, the result trivially holds for $n = 1$. Assume that the statement holds for a fixed $n \in \mathbb{N}^\star$. For any $y \leq \min_{x \in \mathcal{X}} f(x)$ and any $y \geq \max_{x \in \mathcal{X}} f(x)$, the result also trivially holds. Now, fix any $y \in (\min_{x \in \mathcal{X}} f(x), \max_{x \in \mathcal{X}} f(x))$, let $\mathcal{X}_y = \{x \in \mathcal{X} : f(x) \geq y\}$ be the corresponding level set and let $\{X_i\}_{i=1}^{n+1}$ be a sequence distributed as the RankOpt process (see Definition 5). Applying the Bayes rule gives us that,

$$\mathbb{P}(f(X_{\hat{\imath}_{n+1}}) \geq y) = \mathbb{P}(\{X_{n+1} \in \mathcal{X}_y\} \bigcap_{i=1}^{n} \{X_i \notin \mathcal{X}_y\}) + \mathbb{P}(\bigcup_{i=1}^{n} \{X_i \notin \mathcal{X}_y\}). \tag{1}$$

We start to bound the first term. Conditioning on $\{X_i\}_{i=1}^n$ and using the fact that $X_{n+1} | \{X_i\}_{i=1}^n \sim \mathcal{U}(\mathcal{X}_n)$ (see Definition 5), we have that,

$$\mathbb{P}(\{X_{n+1} \in \mathcal{X}_y\} \bigcap_{i=1}^{n} \{X_i \notin \mathcal{X}_y\}) = \mathbb{E}[\mathbb{1}\{\bigcap_{i=1}^{n} \{X_i \notin \mathcal{X}_y\}\} \cdot \mathbb{P}(X_{n+1} \in \mathcal{X}_y | \{X_i\}_{i=1}^n)] = \mathbb{E}\left[\mathbb{1}\{\bigcap_{i=1}^{n} \{X_i \notin \mathcal{X}_y\}\} \cdot \frac{\mu(\mathcal{X}_n \cap \mathcal{X}_y)}{\mu(\mathcal{X}_n)}\right]$$

On the event $\bigcap_{i=1}^n \{X_i \notin \mathcal{X}_y\}_{i=1}^n = \{f(X_{\hat{\imath}_n}) < y\}$, we have that $\{\mathcal{X}_y \subseteq \mathcal{X}_n \subseteq \mathcal{X}\}$ (see Lemma 1). Hence,

$$\mathbb{P}(\{X_{n+1} \in \mathcal{X}_y\} \bigcap_{i=1}^{n} \{X_i \notin \mathcal{X}_y\}) \geq \frac{\mu(\mathcal{X}_y)}{\mu(\mathcal{X})} \cdot \mathbb{E}[\mathbb{1}\{\bigcap_{i=1}^{n} \{X_i \notin \mathcal{X}_y\}\}]. \tag{2}$$

Plugging (2) into (1) and noticing that $\mathbb{E}[\mathbb{1}\{\bigcap_{i=1}^n \{X_i \notin \mathcal{X}_y\}\}] = 1 - \mathbb{P}(f(X_{\hat{\imath}_n}) \geq y)$ give us that

$$\mathbb{P}(f(X_{\hat{\imath}_{n+1}}) \geq y) \geq \mathbb{P}(f(X_{\hat{\imath}_n}) \geq y) + \frac{\mu(\mathcal{X}_y)}{\mu(\mathcal{X})} \cdot (1 - \mathbb{P}(f(X_{\hat{\imath}_n}) \geq y)). \tag{3}$$

Now, let $\{X_i'\}_{i=1}^{n+1}$ be a sequence of $(n+1)$ i.i.d. random variables uniformly distributed over $\mathcal{X}$. Reproducing the same steps as previously and using the fact that $\mathbb{P}(X_{n+1}' \in \mathcal{X}_y) = \mu(\mathcal{X}_y)/\mu(\mathcal{X})$, we get that,

$$\mathbb{P}\left(\max_{i=1 \ldots n+1} f(X_i') \geq y\right) = \mathbb{P}\left(\max_{i=1 \ldots n} f(X_i') \geq y\right) + \frac{\mu(\mathcal{X}_y)}{\mu(\mathcal{X})} \cdot \left(1 - \mathbb{P}\left(\max_{i=1 \ldots n} f(X_i') \geq y\right)\right). \tag{4}$$

Plugging the induction assumption into (3) and comparing the result with (4) gives us $\mathbb{P}(f(X_{\hat{\imath}_{n+1}}) \geq y) \leq \mathbb{P}(\max_{i=1\dots n+1} f(X'_i) \geq y)$. □

Proof of Corollary 1. (CONSISTENCY) We use the same notations and assumptions as in Definition 5 and we further assume that the maximum of the function f is identifiable. Fix any $\varepsilon > 0$, let $\mathcal{X}_{f^\star-\varepsilon} = \{x \in \mathcal{X} : f(x) \geq \max_{x\in\mathcal{X}} f(x) - \varepsilon\}$ be the corresponding level set and fix any $n \in \mathbb{N}^\star$. Applying Proposition 2 gives that $\mathbb{P}(f(X_{\hat{\imath}_n}) < \max_{x\in\mathcal{X}} f(x) - \varepsilon) \leq \mathbb{P}(\max_{i=1\dots n} f(X'_i) < \max_{x\in\mathcal{X}} f(x) - \varepsilon)$ where $\{X'_i\}_{i=1}^n$ are n independent copies of $X' \sim \mathcal{U}(\mathcal{X})$. Using the independence of the $\{X'_i\}_{i=1}^n$, we get that,

$$\mathbb{P}(f(X_{\hat{\imath}_n}) < \max_{x\in\mathcal{X}} f(x) - \varepsilon) \leq \mathbb{P}(X'_1 \notin \mathcal{X}_{f^\star-\varepsilon})^n = \left(1 - \frac{\mu(\mathcal{X}_{f^\star-\varepsilon})}{\mu(\mathcal{X})}\right)^n.$$

Since $\mu(\mathcal{X}_{f^\star-\varepsilon}) > 0$ by the identifiability condition, we have that $\mathbb{P}(f(X_{\hat{\imath}_n}) < \max_{x\in\mathcal{X}} f(x) - \varepsilon) \underset{n\to\infty}{\longrightarrow} 0$. □

Proof of Theorem 1. (UPPER BOUND) We use the same notations and assumptions as in Definition 5 and we further assume that the function f has (α, c_α)-regular level set. Since the true ranking $r_f \in \mathcal{R} \subseteq \mathcal{R}_\infty$ is continuous, there exists a continuous function $h \in \mathcal{C}^0(\mathcal{X}, \mathbb{R})$ such that $r_h = r_f$ (see Proposition 1). Therefore, without loss of generality, one can assume that the f is continuous (all the arguments used in the proofs only use function comparisons). We also set some additional notations: let $r_{\delta,n} = c_\alpha c_\alpha^{1/(1+\alpha)} \operatorname{diam}(\mathcal{X})^{1/(1+\alpha)^2} (\ln(1/\delta)/n)^{1/d(1+\alpha)^2}$, $r'_{\delta,n} = (r_{\delta,n}/c_\alpha)^{1+\alpha}$ and $r''_{\delta,n} = (r'_{\delta,n}/c_\alpha)^{1+\alpha} = \operatorname{diam}(\mathcal{X}) (\ln(1/\delta)/n)^{1/d}$. First, note that the result trivially holds when $r_{\delta,n} \geq \max_{x\in\mathcal{X}} \|x - x^\star\|_2$. Now, assume that $r_{\delta,n} < \max_{x\in\mathcal{X}} \|x - x^\star\|_2$ (it also implies, thanks to the level set assumption, that $\ln(1/\delta) < n$). Using the second inclusion of Lemma 4 and Theorem 1, we have that,

$$\begin{aligned}
\mathbb{P}(\|X_{\hat{\imath}_n} - x^\star\|_2 \leq r_{\delta,n}) &= \mathbb{P}(X_{\hat{\imath}_n} \in B(x^\star, r_{\delta,n})) \\
&\geq \mathbb{P}(f(X_{\hat{\imath}_n}) \geq \min_{x\in\mathcal{S}_{r'_{\delta,n}}} f(x)) && \text{(Lemma 4)} \\
&\geq \mathbb{P}(\max_{i=1\dots n} f(X'_i) \geq \min_{x\in\mathcal{S}_{r'_{\delta,n}}} f(x)), && \text{(Theorem 1)}
\end{aligned}$$

where $\{X'_i\}_{i=1}^n$ are n i.i.d. copies of $X \sim \mathcal{U}(\mathcal{X})$. Now, using the first inclusion of Lemma 4 and the fact that $\{X'_i\}_{i=1}^n$ are n independent random variables uniformly distributed over $\mathcal{X}$, we have that,

$$\mathbb{P}(\|X_{\hat{\imath}_n} - x^\star\|_2 \leq r_{\delta,n}) \geq \mathbb{P}\left(\bigcup_{i=1}^n \{X'_i \in \mathcal{X} \cap B(x^\star, r''_{\delta,n})\}\right) \overset{\text{i.i.d.}}{=} 1 - \left(1 - \frac{\mu(\mathcal{X} \cap B(x^\star, r''_{\delta,n}))}{\mu(\mathcal{X})}\right)^n.$$

Applying lemma 5 gives us that $\mu(\mathcal{X} \cap B(x^\star, r''_{\delta,n}))/\mu(\mathcal{X}) \leq (r''_{\delta,n}/\operatorname{diam}(\mathcal{X}))^d = \ln(1/\delta)/n$. Therefore,

$$\mathbb{P}(\|X_{\hat{\imath}_n} - x^\star\|_2 \leq r_{\delta,n}) \geq 1 - \left(1 - \frac{\ln(1/\delta)}{n}\right)^n.$$

Finally, using the fact that $1 - x \leq e^{-x}$ for any $x \in \mathbb{R}$ gives us that $\mathbb{P}(\|X_{\hat{\imath}_n} - x^\star\|_2 \leq r_{\delta,n}) \geq 1 - \delta$. □

Proof of Proposition 3. We prove the proposition by induction and we use the same notations and assumptions as in Definition 6. Since X_1 and $X_1^\star$ are uniformly distributed over $\mathcal{X}$, the result trivially holds for $n = 1$. Assume that the statement holds for a fixed $n \in \mathbb{N}^\star$. As mentioned in the proof of Proposition 2, the result trivially holds for any $y \leq \min_{x\in\mathcal{X}} f(x)$ and any $y \geq \max_{x\in\mathcal{X}} f(x)$. Fix any $y \in (\min_{x\in\mathcal{X}} f(x), \max_{x\in\mathcal{X}} f(x))$ and let $\mathcal{X}_y = \{x \in \mathcal{X} : f(x) \geq y\}$ be the corresponding level set. Let $\{X_i\}_{i=1}^n$ be a sequence distributed as the RANKOPT process (Definition 5), let $\mathcal{X}_n$ be the corresponding sampling area and let $\mathcal{X}_{f(X_{\hat{\imath}_n})} = \{x \in \mathcal{X} : f(x) \geq f(X_{\hat{\imath}_n})\}$ be the level set of the highest value observed so far. Reproducing the same steps as in the proof of Proposition 2 gives us that,

$$\mathbb{P}(f(X_{\hat{\imath}_{n+1}}) \geq y) = \mathbb{E}\left[\mathbb{1}\{f(X_{\hat{\imath}_n}) \geq y\} + \frac{\mu(\mathcal{X}_y \cap \mathcal{X}_n)}{\mu(\mathcal{X}_n)} \cdot \mathbb{1}\{f(X_{\hat{\imath}_n}) < y\}\right]$$

Applying Lemma 1 gives us that on the event $\{f(X_{\hat{\imath}_n}) < y\}$, we also have that $\{\mathcal{X}_y \subseteq \mathcal{X}_{f(X_{\hat{\imath}_n})} \subseteq \mathcal{X}_n\}$. Therefore, we have that,

$$\mathbb{P}(f(X_{\hat{\imath}_{n+1}}) \geq y) \leq \mathbb{E}\left[\mathbb{1}\{f(X_{\hat{\imath}_n}) \geq y\} + \frac{\mu(\mathcal{X}_y)}{\mu(\mathcal{X}_{f(X_{\hat{\imath}_n})})}\mathbb{1}\{f(X_{\hat{\imath}_n}) < y\}\right] = \mathbb{E}\left[\min\left(1, \frac{\mu(\mathcal{X}_y)}{\mu(\mathcal{X}_{f(X_{\hat{\imath}_n})})}\right)\right]$$

Using the fact that $\mathbb{E}[X] = \int_0^1 \mathbb{P}(X > t)\, dt$ for any random variable $X \in [0, 1]$, we get that,

$$\mathbb{P}(f(X_{\hat{\imath}_{n+1}}) \geq y) \leq \int_0^1 \mathbb{P}\left(\min\left(1, \frac{\mu(\mathcal{X}_y)}{\mu(\mathcal{X}_{f(X_{\hat{\imath}_n})})}\right) > t\right)\, dt = \frac{\mu(\mathcal{X}_y)}{\mu(\mathcal{X})} + \int_{\frac{\mu(\mathcal{X}_y)}{\mu(\mathcal{X})}}^1 \mathbb{P}\left(\mu(\mathcal{X}_{f(X_{\hat{\imath}_n})}) < \frac{\mu(\mathcal{X}_y)}{t}\right)\, dt \tag{5}$$

Now, let $\{X_i^\star\}_{i=1}^{n+1}$ be a sequence distributed as a PAS process (Definition 6). Reproducing the same steps, we get that

$$\mathbb{P}(f(X_{n+1}^\star) \geq y) = \mathbb{E}\left[\min\left(1, \frac{\mu(\mathcal{X}_y)}{\mu(\mathcal{X}_{f(X_n^\star)})}\right)\right] = \frac{\mu(\mathcal{X}_y)}{\mu(\mathcal{X})} + \int_{\frac{\mu(\mathcal{X}_y)}{\mu(\mathcal{X})}}^1 \mathbb{P}\left(\mu(\mathcal{X}_{f(X_n^\star)}) < \frac{\mu(\mathcal{X}_y)}{t}\right)\, dt. \tag{6}$$

Now, fix any $t \in (\mu(\mathcal{X}_y)/\mu(\mathcal{X}), 1)$ and let $\bar{y}_t = \min\{y' \in \mathrm{Im}(f) : \mu(\mathcal{X}_{y'}) < \mu(\mathcal{X}_y)/t\}$. Using the induction assumption, we get that,

$$\mathbb{P}\left(\mu(\mathcal{X}_{f(X_{\hat{\imath}_n})}) < \frac{\mu(\mathcal{X}_y)}{t}\right) = \mathbb{P}(f(X_{\hat{\imath}_n}) \geq \bar{y}_t) \leq \mathbb{P}(f(X_n^\star) \geq \bar{y}_t) = \mathbb{P}\left(\mu(\mathcal{X}_{f(X_n^\star)}) < \frac{\mu(\mathcal{X}_y)}{t}\right). \tag{7}$$

Since the previous inequality holds for any $t \in (\mu(\mathcal{X}_y)/\mu(\mathcal{X}), 1)$, plugging equation (7) into inequality (5) and comparing the result with equation (6) gives that $\mathbb{P}(f(X_{\hat{\imath}_n}) \geq y) \leq \mathbb{P}(f(X_n^\star) \geq y)$. □

Proof of Theorem 2. (LOWER BOUND) We use the same notations and assumptions as in Definition 5 and we further assume that the function f has (α, c_α)-regular level sets. As mentioned in the proof of Theorem 1, since the true ranking $r_f \in \mathcal{R} \subseteq \mathcal{R}_\infty$ is continuous, there exists a continuous function $h \in \mathcal{C}^0(\mathcal{X}, \mathbb{R})$ such that $r_h = r_f$ (see Proposition 1). Therefore, one can assume, without loss of generality, that the function f is continuous (all the arguments used in the proofs only use function comparisons). We also set some additional notations: let $r_{\delta,n} = c_\alpha^{-(1+\alpha)(2+\alpha)}\, \mathrm{rad}(\mathcal{X})^{(1+\alpha)^2}\, \delta^{(1+\alpha)^2/d} \exp(-(1+\alpha)^2(n+\sqrt{2n\ln(1/\delta)}/d)$, $r'_{\delta,n} = c_\alpha r_{\delta,n}^{1/(1+\alpha)}$ and $r''_{\delta,n} = c_\alpha r'^{1/(1+\alpha)}_{\delta,n} = \mathrm{rad}(\mathcal{X})\, \delta^{1/d} \exp(-(n + \sqrt{2n\ln(1/\delta)})/d)$. Using the first inclusion of Lemma 4 and Proposition 3 gives that,

$$\begin{aligned}
\mathbb{P}(\|X_{\hat{\imath}_n} - x^\star\|_2 \leq r_{\delta,n}) &= \mathbb{P}(X_{\hat{\imath}_n} \in B(x^\star, r_{\delta,n}) \cap \mathcal{X}) \\
&\leq \mathbb{P}(f(X_{\hat{\imath}_n}) \geq \min_{x \in \mathcal{S}_{r'_{\delta,n}}} f(x)) && \text{(Lemma 4)} \\
&\leq \mathbb{P}(f(X_n^\star) \geq \min_{x \in \mathcal{S}_{r'_{\delta,n}}} f(x)), && \text{(Proposition 3)}
\end{aligned}$$

where $\{X_i^\star\}_{i=1}^n$ is a sequence distributed as the PAS process (Definition 6). Now, denoting by $\mathcal{X}_n^\star = \{x \in \mathcal{X} : f(x) \geq f(X_n^\star)\}$ the sampling area of the PAS process after n iterations and using the second inclusion of Lemma 4 gives that,

$$\mathbb{P}(\|X_{\hat{\imath}_n} - x^\star\|_2 \leq r_{\delta,n}) \leq \mathbb{P}\left(\mu(\mathcal{X}_n^\star) \leq \mu(\{x \in \mathcal{X} : f(x) \geq \min_{x \in \mathcal{S}_{r'_{\delta,n}}} f(x)\})\right) \leq \mathbb{P}\left(\frac{\mu(\mathcal{X}_n^\star)}{\mu(\mathcal{X})} \leq \frac{\mu(B(x^\star, r''_{\delta,n}))}{\mu(\mathcal{X})}\right).$$

Using the definition of $\mathrm{rad}(\mathcal{X})$, we know that there exists $x' \in \mathcal{X}$ such that $B(x', \mathrm{rad}(\mathcal{X})) \subseteq \mathcal{X}$. Hence, $\mu(\mathcal{X}) \geq \mu(B(x', \mathrm{rad}(\mathcal{X})))$. Therefore, using the volume of the ball (see Lemma 5) we get that,

$$\mathbb{P}(\|X_{\hat{\imath}_n} - x^\star\|_2 \leq r_{\delta,n}) \leq \mathbb{P}\left(\frac{\mu(\mathcal{X}_n^\star)}{\mu(\mathcal{X})} \leq \frac{\mu(B(x^\star, r'_{\delta,n}))}{\mu(B(x', \mathrm{rad}(\mathcal{X})))}\right) = \mathbb{P}\left(\frac{\mu(\mathcal{X}_n^\star)}{\mu(\mathcal{X})} \leq \left(\frac{r''_{\delta,n}}{\mathrm{rad}(\mathcal{X})}\right)^d\right).$$

Applying Lemma 2 and using the fact that $(r''_{\delta,n}/\operatorname{rad}(\mathcal{X}))^d = \delta \exp\left(-n - \sqrt{2n\ln(1/\delta)}\right)$ gives that

$$\mathbb{P}(\|X_{\hat{\imath}_n} - x^\star\|_2 \leq r_{\delta,n}) = \mathbb{P}\left(\prod_{i=1}^{n} U_i \leq \delta \cdot e^{-n-\sqrt{2n\ln(1/\delta)}}\right)$$

where $\{U_i\}_{i=1}^n$ is a sequence of n i.i.d. copies of $U \sim \mathcal{U}([0,1])$. Applying Lemma 3 finally gives us that $\mathbb{P}(\|X_{\hat{\imath}_n} - x^\star\|_2 \leq r_{\delta,n}) < \delta$. □

C. Proofs of section 4

C.1. AdaRank process and additional results

Definition 7. (ADARANKOPT PROCESS) *Fix any $p \in (0,1)$, let $\{\mathcal{R}_N\}_{N\in\mathbb{N}^\star}$ be any sequence of nested sets of rankings and let $f : \mathcal{X} \to \mathbb{R}$ be any function such that $r_f \in \mathcal{R}_\infty$. We say that the sequence $\{X_i\}_{i=1}^n$ is distributed as the* ADARANKOPT$(n, f, \mathcal{X}, p, \{\mathcal{R}_N\}_{N\in\mathbb{N}^\star})$ *process if the sequence has the same distribution as the process defined by:*

$$\begin{cases} X_1 \sim \mathcal{U}(\mathcal{X}) \\ X_{t+1} |\, B_{t+1}, \{X_i\}_{i=1}^t \sim B_{t+1} \cdot \mathcal{U}(\mathcal{X}) + (1 - B_{t+1}) \cdot \mathcal{U}(\mathcal{X}_t) \qquad \forall t \in \{1 \dots n-1\} \end{cases}$$

where $\mathcal{X}_t = \{x \in \mathcal{X} : \exists r \in \mathcal{R}_{N_t},\ r(x, X_{\hat{\imath}_t}) \geq 0\}$ denotes the sampling area, $N_t = \min\{N \in \mathbb{N}^\star : \min_{r\in\mathcal{R}_N} L_t(r) = 0\}$ denotes the index of the smallest set of rankings that may contain r_f, $\hat{\imath}_t \in \arg\max_{i=1\dots t} f(X_i)$ denotes the index of the best value observed so far and B_{t+1} is a Bernoulli random variable of parameter p (i.e. $p = \mathbb{P}(B_{t+1} = 1) = 1 - \mathbb{P}(B_{t+1} = 0)$), independent of $\{X_i\}_{i=1}^t$ and $\{B_i\}_{i=1}^t$.

The next proposition will be needed later.

Proposition 9. *[From (Clémençon et al., 2010)] Let $\{X_i\}_{i=1}^n$ be a sequence of n independent copies of $X \sim \mathcal{U}(\mathcal{X})$, let $\mathcal{R}$ be any set of rankings and let f be any function defined on $\mathcal{X}$ taking values in $\mathbb{R}$. Define the Rademacher average*

$$R_n(\mathcal{R}) = \sup_{r\in\mathcal{R}} \frac{1}{\lfloor n/2 \rfloor} \left| \sum_{i=1}^{\lfloor n/2 \rfloor} \epsilon_i \cdot \mathbb{1}\{r_f(X_i, X_{\lfloor n/2 \rfloor + i}) \neq r(X_i, X_{\lfloor n/2 \rfloor + i})\} \right|$$

where $\{\epsilon_i\}_{i=1}^{\lfloor n/2 \rfloor}$ are $\lfloor n/2 \rfloor$ independent Rademacher random variables (i.e., random symmetric sign variables), independent of $\{X_i\}_{i=1}^n$. Then, for any $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\forall r \in \mathcal{R},\ |L_n(r) - L(r)| \leq 2\mathbb{E}[R_n(\mathcal{R})] + 2\sqrt{\frac{\log(1/\delta)}{n-1}},$$

where $L(r) = \mathbb{P}(r_f(X, X') \neq r(X, X'))$ denotes the true ranking loss and (X, X') is a couple of independent random variables uniformly distributed over $\mathcal{X}$.

Corollary 3. *Assume that there exists $V > 0$ such that $\mathbb{E}[R_n(\mathcal{R})] \leq \sqrt{V/n}$ for any $n \in \mathbb{N}^\star$ and assume that $\inf_{r\in\mathcal{R}} L(r) > 0$. Let $\{X_i\}_{i\in\mathbb{N}^\star}$ be a sequence of i.i.d. random variables uniformly distributed over $\mathcal{X}$ and let $\tau = \min\{n \in \mathbb{N}^\star : \min_{r\in\mathcal{R}} L_n(r) > 0\}$ be the stopping time corresponding to the number of samples required to be certain that $r_f \notin \mathcal{R}$ where L_n denotes the empirical ranking loss taken over the first n samples $\{X_i\}_{i=1}^n$. Then, for any $\delta \in (0,1)$, with probability at least $1-\delta$,*

$$\tau \leq 10 \cdot \left(\frac{V + \log(1/\delta)}{\inf_{r\in\mathcal{R}} L(r)^2}\right).$$

Proof. Fix any $\delta \in (0,1)$, let $n_\delta = \lfloor 10 \cdot (V + \log(1/\delta))/(\inf_{r\in\mathcal{R}} L(r)^2) \rfloor$ be the integer part of the upper bound and let L_{n_δ} be the empirical ranking loss taken over the first n_δ i.i.d. samples of the sequence $\{X_i\}_{i=1}^{n_\delta}$. By definition, we have that $\mathbb{P}(\tau \leq n_\delta) = \mathbb{P}(\min_{r\in\mathcal{R}} L_{n_\delta}(r) > 0)$. Now, since $L(r) \geq \min_{r\in\mathcal{R}} L(r)$ for any $r \in \mathcal{R}$, applying Proposition 9 gives us that, with probability at least $1-\delta$,

$$\min_{r\in\mathcal{R}} L_{n_\delta}(r) \geq \inf_{r\in\mathcal{R}} L(r) - 2\sqrt{\frac{V}{n_\delta}} - 2\sqrt{\frac{\log(1/\delta)}{n_\delta - 1}}.$$

We end the proof by pointing out that the right hand term of the previous inequality is strictly positive due to the definition of n_δ. □

C.2. Proofs of the results

Proof of Proposition 4. (CONSISTENCY) We use the same notations and assumptions as in Definition 7 and we further assume that the maximum of the function f is identifiable. Fix any $\varepsilon > 0$ and let $\mathcal{X}_{f^\star-\varepsilon} = \{x \in \mathcal{X} : f(x) \geq \max_{x\in\mathcal{X}} f(x) - \varepsilon\}$ be the corresponding level set. We show by induction that, for any $n \in \mathbb{N}^\star$,

$$\mathbb{P}(f(X_{\hat{\imath}_n}) < \max_{x\in\mathcal{X}} f(x) - \varepsilon) \leq \left(1 - p\cdot\frac{\mu(\mathcal{X}_{f^\star-\varepsilon})}{\mu(\mathcal{X})}\right)^n \underset{n\to\infty}{\longrightarrow} 0.$$

• Since $X_1 \sim \mathcal{U}(\mathcal{X})$, we directly get that,

$$\mathbb{P}(f(X_{\hat{\imath}_1}) < \max_{x\in\mathcal{X}} f(x) - \varepsilon) = \mathbb{P}(X_1 \notin \mathcal{X}_{f^\star-\varepsilon}) = \left(1 - \frac{\mu(\mathcal{X}_{f^\star-\varepsilon})}{\mu(\mathcal{X})}\right) \leq \left(1 - p\cdot\frac{\mu(\mathcal{X}_{f^\star-\varepsilon})}{\mu(\mathcal{X})}\right).$$

• Assume that the statement holds for a fixed $n \in \mathbb{N}^\star$ and let $\{X_i\}_{i=1}^{n+1}$ be a sequence of $n+1$ random variables distributed as the ADARANKOPT process. Conditioning on $\{X_i\}_{i=1}^n$ gives us that

$$\mathbb{P}(f(X_{\hat{\imath}_{n+1}}) < \max_{x\in\mathcal{X}} f(x) - \varepsilon) = \mathbb{P}\left(\bigcap_{i=1}^{n+1}\{X_i \notin \mathcal{X}_{f^\star-\varepsilon}\}\right) = \mathbb{E}\left[\mathbb{P}(X_{n+1} \notin \mathcal{X}_{f^\star-\varepsilon} | \{X_i\}_{i=1}^n)\cdot \mathbb{1}\{\bigcap_{i=1}^{n}\{X_i \notin \mathcal{X}_{f^\star-\varepsilon}\}\}\right]$$

Now, using the distribution of $X_{n+1}|\{X_i\}_{i=1}^n$ (Definition 7), we get that,

$$\begin{aligned}\mathbb{P}(f(X_{\hat{\imath}_{n+1}}) < \max_{x\in\mathcal{X}} f(x) - \varepsilon) &= \mathbb{E}\left[\left(1 - p\cdot\frac{\mu(\mathcal{X}_{f^\star-\varepsilon})}{\mu(\mathcal{X})} - (1-p)\cdot\frac{\mu(\mathcal{X}_n \cap \mathcal{X}_{f^\star-\varepsilon})}{\mu(\mathcal{X}_n)}\right)\cdot \mathbb{1}\{\bigcap_{i=1}^{n}\{X_i \notin \mathcal{X}_{f^\star-\varepsilon}\}\}\right] \\ &\leq \left(1 - p\cdot\frac{\mu(\mathcal{X}_{f^\star-\varepsilon})}{\mu(\mathcal{X})}\right)\cdot\mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \notin \mathcal{X}_{f^\star-\varepsilon}\}\right).\end{aligned}$$

Plugging the induction assumption into the previous equation gives the result. □

Proof of Proposition 5. (STOPPING TIME UPPER BOUND) We use the same assumptions and notations as in Definition 7. Let $\{X_i\}_{i\in\mathbb{N}^\star}$ be a sequence distributed as the ADARANKOPT and let $\{B_i\}_{i\in\mathbb{N}^\star}$ be the sequence of associated random variables corresponding to the exploration/exploitation tradeoff. Assume that $1 < N^\star = \min\{N \in \mathbb{N}^\star : r_f \in \mathcal{R}_N\} < \infty$, assume that there exists a constant $V > 0$ such that $\mathbb{E}[R_n(\mathcal{R}_{N^\star-1})] \leq \sqrt{V/n}$ for any $n \in \mathbb{N}^\star$ and assume that $\inf_{r\in\mathcal{R}_{N^\star-1}L(r)} > 0$. Fix any $\delta \in (0,1)$, let $n_\delta = \lfloor 10\cdot(V + \log(2/\delta))/(p\cdot\inf_{r\in\mathcal{R}_{N^\star-1}} L(r)^2)\rfloor$ be the integer part of the upper bound, let $n'_\delta = \lfloor p\cdot n_\delta - \sqrt{n_\delta\log(2/\delta)/2}\rfloor$ and denote by L_{n_δ} the empirical ranking loss taken over the n_δ first samples $\{X_i\}_{i=1}^{n_\delta}$. Since $\{\mathcal{R}_N\}_{N\in\mathbb{N}^\star}$ forms a nested sequence $\min_{r\in\mathcal{R}_1} L_{n_\delta}(r) \leq \min_{r\in\mathcal{R}_2} L_{n_\delta}(r) \leq \ldots \leq \min_{r\in\mathcal{R}_{N^\star-1}} L_{n_\delta}(r)$. Therefore,

$$\mathbb{P}(\tau \leq n_\delta) = \mathbb{P}\left(\min_{r\in\mathcal{R}_{N^\star-1}} L_{n_\delta}(r) > 0\right) \geq \mathbb{P}\left(\min_{r\in\mathcal{R}_{N^\star-1}} L_{n_\delta}(r) > 0 \cap \{\sum_{i=1}^{n_\delta} B_i \geq n'_\delta\}\right). \tag{8}$$

We now lower bound the empirical risk by only keeping the first n'_δ (i.i.d.) explorative samples:

$$L_{n_\delta}(r) \geq \frac{2}{n_\delta(n_\delta-1)}\sum_{1\leq i<j\leq n_\delta} \mathbb{1}\{r_f(X_i,X_j) \neq r(X_i,X_j)\}\cdot\mathbb{1}\{(i,j)\in I^2\},$$

where $I = \{i \in \mathbb{N}^\star : B_i = 1 \text{ and } \sum_{j=1}^{i} B_i \leq n'_\delta\}$. Conditioning on $|I|$ and using Definition 7 gives us that $\{X_i\}_{i\in I}||I|$ is a sequence of $|I|$ independent random variables uniformly distributed over $\mathcal{X}$. Therefore, on the event $\{\sum_{i=1}^{n_\delta} B_i \geq n'_\delta\} = \{|I| = n'_\delta\}$, the right hand term of the previous inequality has the same distribution as,

$$L_{n'_\delta}(r) = \frac{2}{n_\delta(n_\delta - 1)} \sum_{1\leq i<j\leq n_{\delta'}} \mathbb{1}\{r_f(X'_i, X'_j) \neq r(X'_i, X'_j)\}$$

where $\{X'_i\}_{i=1}^{n'_\delta}$ is a sequence of n'_δ i.i.d. copies of $X \sim \mathcal{U}(\mathcal{X})$, independent of $\{B_i\}_{i=1}^{n_\delta}$. Combining the previous statement with (8) gives us that,

$$\mathbb{P}(\tau \leq n_\delta) \geq \mathbb{P}\left(\min_{r\in\mathcal{R}_{N^\star-1}} L_{n'_\delta}(r) > 0\right) \cdot \mathbb{P}\left(\sum_{i=1}^{n_\delta} B_i \geq n'_\delta\right),$$

where the empirical risk is taken over a sample of $n_{\delta'}$ independent copies of $X \sim \mathcal{U}(\mathcal{X})$. Due to the definition of n'_δ, applying Corollary 3 gives us that $\mathbb{P}(\min_{r\in\mathcal{R}_{N^\star-1}} L_{n'_\delta}(r) > 0) \geq 1 - \delta/2$ and using Hoeffding's inequality gives us that $\mathbb{P}(\sum_{i=1}^{n_\delta} B_i \geq n'_\delta) \geq 1 - \delta/2$. Noticing that $(1 - \delta/2)^2 \geq 1 - \delta$, for any $\delta \in (0, 1)$ completes the proof. □

Proof of Theorem 3. (UPPER BOUND) We use the same setting as in the previous proof: we use the same assumptions and notations as in Definition 7, we assume $1 < N^\star < \infty$, we assume that there exists $V > 0$ such that $\mathbb{E}[R_n(\mathcal{R}_{N^\star-1})] \leq \sqrt{V/n}$ for all $n \in \mathbb{N}^\star$ and we assume that $\inf_{r\in\mathcal{R}_{N^\star-1}} L(r) > 0$. Let $\{X_i\}_{i\in\mathbb{N}^\star}$ be a sequence distributed as the ADARANKOPT and let $\{N_t\}_{t\in\mathbb{N}^\star}$ be the random variables corresponding to the model selection. Fix any $\delta \in (0, 1)$, let $n_{\delta/2} = \lfloor 10(V + \ln(4/\delta))/(p \cdot \inf_{r\in\mathcal{R}_{N^\star-1}} L(r)^2)\rfloor$ be the integer part of the upper bound of proposition 5 (with probability $1 - \delta/2$) and let $r_{\delta/2,n}$ be the upper bound of the Theorem 1 (with probability $1 - \delta/2$). Fix any $n > n_{\delta/2}$, let $\{X_i\}_{i=1}^n$ be a sequence distributed as the ADARANKOPT process (Definition 7) and let $\tau_{N^\star} = \min\{t \in \{1 \ldots n\} : N_t = N^\star\}$ be the stopping time corresponding to the number of iterations required to identify the true ranking structure $\mathcal{R}_{N^\star}$. Applying the Bayes rules gives us that

$$\mathbb{P}(\|X_{\hat{\imath}_n} - x^\star\|_2 \leq r_{\delta,n}) \geq \mathbb{P}(\|X_{\hat{\imath}_n} - x^\star\|_2 \leq r_{\delta,n} \mid \tau_{N^\star} < n_{\delta/2}) \cdot \mathbb{P}(\tau_{N^\star} < n_{\delta/2}).$$

Applying proposition 5 gives us that $\mathbb{P}(\tau < n_{\delta/2}) \geq 1-\delta/2$. To lower bound the first term, we use the fact that on the event $\{\tau < n_{\delta/2}\}$, for any iteration $n > n_{\delta/2}$, the true ranking structure $\mathcal{R}_{N^\star}$ is identified. Therefore, one can bound the distance $\|X_{\hat{\imath}_n} - x^\star\|_2$ by using the $n - n_{\delta/2}$ samples with a similar technique as the one used in the RANKOPT process when the ranking structure is known (see proof of Theorem 1) and we get that $\mathbb{P}(\|X_{\hat{\imath}_n} - x^\star\|_2 \leq r_{\delta,n} \mid \tau < n_{\delta/2}) \geq 1 - \delta/2$. Noticing that $(1 - \delta/2)^2 \geq 1 - \delta$ for any $\delta \in (0, 1)$ ends the proof. □

D. Proofs of section 6

D.1. Convex hulls and technical lemmas

Definition 8. *Let $\{X_i\}_{i=1}^n$ be any set of n points of $\mathbb{R}^d$. The convex hull of $\{X_i\}_{i=1}^n$ can be defined as*

$$\text{CH}\{X_i\}_{i=1}^n = \left\{\sum_{i=1}^{n} \lambda_i X_i : (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}_+^n, \ \sum_{i=1}^{n} \lambda_i = 1\right\}, \text{ where } \mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}.$$

Lemma 6. *Fix any $\epsilon > 0$, let $\mathcal{X} \subset \mathbb{R}^d$ be any compact and convex set and let $B(\mathcal{X}, \epsilon) = \{x \in \mathbb{R}^d : d(x, \mathcal{X}) \leq \epsilon\}$ be the ϵ-ball of $\mathcal{X}$ where $d(x, \mathcal{X}) = \min_{x'\in\mathcal{X}} \|x - x'\|_2$ denotes the distance between $x \in \mathbb{R}^d$ and $\mathcal{X}$. Then, the ϵ-ball of $\mathcal{X}$ $B(\mathcal{X}, \epsilon)$ is also a convex set*

Proof. Fix any $\epsilon > 0$ and let $(b_1, b_2) \in B(\mathcal{X}, \epsilon)$ be any pair of points that belong to the ϵ-ball of $\mathcal{X}$. Since $b_1 \in B(\mathcal{X}, \epsilon)$, there exists $x_1 \in \mathcal{X}$ and $\epsilon_1 \in \mathbb{R}^d$ satisfying $\|\epsilon_1\|_2 \leq \epsilon$ such that $b_1 = x_1 + \epsilon_1$. We also know that there also exists b_2 and ϵ_2 satisfying the same conditions such that $b_2 = x_2 + \epsilon_2$. Now, fix any $\lambda \in (0, 1)$. The convexity of $\mathcal{X}$ implies that

$$(1 - \lambda)b_1 + \lambda b_2 \ = \ \underbrace{\lambda x_1 + (1 - \lambda)x_2}_{\in\mathcal{X}} \ + \ \underbrace{\lambda\epsilon_1 + (1 - \lambda)\epsilon_2}_{\|\cdot\|_2\leq\epsilon}.$$

Therefore, $(1-\lambda)b_1 + \lambda b_2 \in B(\mathcal{X}, \epsilon)$ also belongs to the ϵ-ball. Since the result holds for any $(b_1, b_2) \in B(\mathcal{X}, \epsilon)$ and any $\lambda \in (0, 1)$, we deduce that $B(\mathcal{X}, \epsilon)$ is also a convex set. □

Lemma 7. *Let $\{X_i\}_{i=1}^n$ be any set of n points of $\mathbb{R}^d$. Then, there exists a separating hyperplane $h \in \mathbb{R}^d$ such that $\langle h, X_i \rangle > 0$ for any $i \in \{1 \ldots n\}$ if and only if $\vec{0} \notin \text{CH}\{X_i\}_{i=1}^n$ where $\vec{0} = \{0 \ldots 0\} \in \mathbb{R}^d$.*

Proof. ($\Rightarrow$) Assume that there exists $h \in \mathbb{R}^d$ such that $\langle h, X_i \rangle > 0$ for all $i \in \{1 \ldots n\}$. We show (by contradiction) that $\vec{0} \notin \text{CH}\{X_i\}_{i=1}^n$. If $\vec{0} \in \text{CH}\{X_i\}_{i=1}^n$, it implies that there exists $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}_+^n$ such that $\vec{0} = \sum_{i=1}^n \lambda_i X_i$ and $\sum_{i=1}^n \lambda_i = 1$ (see Definition 8). It leads us to the contradiction

$$0 = \langle h, \vec{0} \rangle = \sum_{i=1}^n \lambda_i \langle h, X_i \rangle > 0.$$

($\Leftarrow$) Assume that $\vec{0} \notin \text{CH}\{X_i\}_{i=1}^n$. Since n and d are finite $\text{CH}\{X_i\}_{i=1}^n$ is a closed, compact and convex set. Therefore, $\min_{x \in CH} \|x\|_2 = d$ exists and the condition $\vec{0} \notin \text{CH}$ implies that $d > 0$. Let $x_d \in \text{CH}\{X_i\}_{i=1}^n$ be the (unique) point of the convex hull that satisfies $\|x_d\|_2 = d$. We show (by contradiction) that for any $x \in \text{CH}\{X_i\}_{i=1}^n$ we have that $\langle x, x_d \rangle \geq d^2$. Assume that there exists $x' \in \text{CH}\{X_i\}_{i=1}^n$ such that $\langle x', x_d \rangle < d^2$. The convexity of $\text{CH}\{X_i\}_{i=1}^n$ implies that the whole line $L = (x', x_d) \subseteq \text{CH}\{X_i\}_{i=1}^n$ also belongs to the convex hull. However, since $\|x_d\|_2 = d$ and $\langle x', x_d \rangle < \|x_d\|^2$, the line L is not tangent and intersect the ball the ball $B(\vec{0}, d)$. We deduce that there exists $x'' \in L \cap B(\vec{0}, d)$ such that $\|x''\|_2 < d$. Since $x'' \in L \subseteq \text{CH}\{X_i\}_{i=1}^n$ also belongs to the convex hull, the previous statement leads us to the contradiction

$$\min_{x \in \text{CH}\{X_i\}_{i=1}^n} \|x\|_2 \leq \|x''\|_2 < d = \min_{x \in \text{CH}\{X_i\}_{i=1}^n} \|x\|_2 .$$

Therefore, for any $x \in \text{CH}\{X_i\}_{i=1}^n$ we have that $\langle x_d, x \rangle \geq d > 0$. Finally, since $\{X_i\}_{i=1}^n \in \text{CH}\{X_i\}_{i=1}^n$, there exists $h = x_d \in \mathbb{R}^d$ such that $\langle h, X_i \rangle > 0$ for all $i \in \{1 \ldots n\}$. □

Lemma 8. *Let $\mathcal{R} \subseteq \mathcal{R}_\infty$ be any continuous ranking structure, let $\{(X_i, f(X_i))\}_{i=1}^{n+1}$ be any sample satisfying $f(X_{(1)}) < f(X_{(2)}) < \ldots < f(X_{(n+1)})$ and denote by L_{n+1} the empirical ranking loss taken over the sample. Then, we have the following equivalence:*

$$\{r \in \mathcal{R} : L_{n+1}(r) = 0\} = \{r \in \mathcal{R} : r(X_{(i+1)}, X_{(i)}) = 1, \ \forall i \in \{1 \ldots n\}\}.$$

Proof. ($\subseteq$) The first inclusion is a direct consequence of the definition of the empirical ranking loss. Let $r \in \mathcal{R}$ be any ranking such that $L_{n+1}(r) = 0$. Since r perfectly ranks the sample, it implies that for any $i \in \{1 \ldots n\}$,

$$r(X_{(i+1)}, X_{(i)}) = r_f(X_{(i+1)}, X_{(i)}) = \text{sgn}(f(X_{i+1}) - f(X_{(i)})) = 1.$$

($\supseteq$) The second inlusion is a consequence of the transitivity of the rankings. Let $r \in \mathcal{R} \subseteq \mathcal{R}_\infty$ be any ranking such that $r(X_{(i+1)}, X_{(i)}) = 1, \ \forall i \in \{1 \ldots n\}$. Since r is acontinuous ranking, there exists a function $f' \in \mathcal{C}^0(\mathcal{X}, \mathbb{R})$ such that $r(x, x') = \text{sgn}(f(x) - f(x)) \ \forall (x, x') \in \mathcal{X}^2$. Now, fix any $j \in \{2 \ldots n+1\}$ and fix any $k < j$. Using the function f' and the assumption, we have that

$$r(X_{(j)}, X_{(k)}) = \text{sgn}(f'(X_{(j)}) - f'(X_{(k)})) = \text{sgn}\left(\sum_{i=k}^{j-1} \left| f'(X_{(i+1)}) - f'(X_{(i)}) \right| \cdot \text{sgn}(f'(X_{(i+1)}) - f'(X_{(i)}))\right) = 1.$$

Switching k and j gives us that $r(X_{(j)}, X_{(k)}) = -1$ for any $j < k$. Therefore, $r(X_{(j)}, X_{(k)}) = \text{sgn}(f(X_{(j)}) - f(X_{(k)}))$ for any $(j, k) \in \{1 \ldots n+1\}^2$ and we deduce that r perfectly ranks the sample so $L_{n+1}(r) = 0$. □

D.2. Proofs of the results

Proof of Proposition 6. Let $\{(X_i, f(X_i))\}_{i=1}^n$ be any sample satisfying $f(X_{(1)}) < f(X_{(2)}) < \ldots < f(X_{(n+1)})$ and let $\mathcal{R}_{\mathcal{P},N}$ be the set of polynomial rankings of degree N.

($\Rightarrow$) Assume that there exists $r \in \mathcal{R}_{\mathcal{P},N}$ such that $L_{n+1}(r) = 0$. Since r is induced by a polynom of degree N, there exists a polynom $f'_r = \langle h_r, \phi_N(x) \rangle + c_r$ where $h_r \in \mathbb{R}^{D(d,N)}$ and $c_r \in \mathbb{R}$ such that $r(x, x') = \text{sgn}(f'_r(x) - f'_r(x')) =$

$\text{sgn}(\langle h_r, \phi_N(x)\rangle + c_r - \langle h_r, \phi_N(x')\rangle - c_r) = \text{sgn}(\langle h_r, \phi_N(x) - \phi_N(x')\rangle)$. Combining the previous statement with Lemma 8 gives the first part of the equivalence.

($\Leftarrow$) Assume that there exists $h \in \mathbb{R}^{D(d,N)}$ such that $\langle h, \phi(X_{(i+1)}) - \phi(X_{(i)})\rangle > 0$ for all $i \in \{1 \dots n\}$. Define the (multivariate) polynom $f'(x) = \langle h, \phi_N(x)\rangle + c$ where $c \in \mathbb{R}$ is any constant. The polynomial ranking induced by f' is given by $r_{f'}(x, x') = \text{sgn}(f'(x) - f'(x')) = \text{sgn}(\langle h, \phi_N(x) - \phi_N(x')\rangle) \in \mathcal{R}_N$. Then, for any $i \in \{1 \dots n\}$,

$$r_{f'}(X_{(i+1)}, X_{(i)}) = \text{sgn}(\langle h, \phi_N(x) - \phi_N(x')\rangle) = 1.$$

Using Lemma 8 directly gives that $L_{n+1}(r_h) = 0$, which is the second part of the equivalence. □

Proof of Corollary 2. Let $\{(X_i, f(X_i))\}_{i=1}^n$ be any sample satisfying $f(X_{(1)}) < f(X_{(2)}) < \dots < f(X_{(n+1)})$ and let $\mathcal{R}_{\mathcal{P},N}$ be the set of polynomial rankings of degree at most N. Applying Proposition 6 and Lemma 7 gives us that,

$$\min_{r \in \mathcal{R}_{\mathcal{P},N}} L_{n+1}(r) = 0 \Leftrightarrow \exists h \in \mathbb{R}^d \text{ s.t. } \langle h, \phi_N(X_{(i+1)}) - \phi_N(X_{(i)})\rangle > 0, \forall i \in \{1 \dots n\} \quad \text{(Proposition 6)}$$

$$\Leftrightarrow \vec{0} \notin \text{CH}\{(\phi_N(X_{(i+1)}) - \phi_N(X_{(i)}))\}_{i=1}^n. \quad \text{(Lemma 7)}$$

Finally, using the definition of the convex hull, we know that $\vec{0} \in \text{CH}\{(\phi_N(X_{(i+1)}) - \phi_N(X_{(i)}))\}_{i=1}^n$ if and only if there exists $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}_+^n$ such that $\sum_{i=1}^n \lambda_i(\phi_N(X_{(i+1)}) - \phi_N(X_{(i)})) = \vec{0}$ and $\sum_{i=1}^n \lambda_i = 1$. Putting those constraints into matricial form gives the last equivalence,

$$\min_{r \in \mathcal{R}_N} L_{n+1}(r) = 0 \Leftrightarrow \{\lambda \in \mathbb{R}^n : \text{M}_N \lambda^T = \vec{0},\ \langle \vec{1}, \lambda\rangle = 1,\ \lambda \succeq \vec{0}\} \text{ is empty,}$$

where M_N is the $(D(d,n), n)$-matrix where its i-th column is equal to $(\phi_N(X_{(i+1)}) - \phi_N(X_{(i)}))^T$ and $\succeq$ stands for the inequality $\geq$ component-wise (i.e. $x \succeq x' \Leftrightarrow x_i \geq x_i', \forall i \in \{1 \dots d\}$). □

Proof of Proposition 7. Fix any $N \in \mathbb{N}^\star$, let $\mathcal{R}_{\mathcal{C},N}$ be the set of convex rankings of degree N, let $d = 1$ and let $\{(X_i, f(X_i))\}_{i=1}^n$ be any sample satisfying $f(X_{(1)}) < f(X_{(2)}) < \dots < f(X_{(n+1)})$.

($\Rightarrow$) Assume that $\min_{r \in \mathcal{R}_{\mathcal{C},N}} L_{n+1}(r) = 0$ and let $r \in \mathcal{R}_{\mathcal{C},N}$ be any convex ranking that perfectly ranks the sample. Applying Lemma 8 gives us that $r(X_{(i+1)}, X_{(i)}) = 1$ for all $i \in \{1 \dots n\}$. Now, let $\{h_i\}_{i=1}^{n+1}$ be the sequence of classifiers defined by $h_i(x) = \mathbb{1}\{r(x, X_{(i)}) \geq 0\}$. Since $r \in \mathcal{R}_{\mathcal{C},N}$ is a convex ranking of degree N, all the classifiers are of the form $h_i(x) = \sum_{k=1}^N \mathbb{1}\{l_{i,k} \leq x \leq u_{i,k}\}$ and since r is a continuous ranking and so transitive (see Lemma 8), $h_1 \geq h_2 \geq \dots \geq h_{n+1}$.

($\Leftarrow$)Assume that there exists a sequence of classifiers $\{h_i\}_{i=1}^{n+1}$ of the form $h_i(x) = \sum_{k=1}^N \mathbb{1}\{l_{i,k} \leq x \leq u_{i,k}\}$ satisfying: (i) $h_1 \geq h_2 \geq \dots \geq h_{n+1}$ and (ii) $h_i(X_{(j)}) = \mathbb{1}\{(j) \geq i\}$ for all $(i,j) \in \{1 \dots n+1\}^2$. Let $f'_\epsilon = \sum_{i=1}^{n+1} \sum_{k=1}^N \phi_\epsilon(x, l_{i,k}, u_{i,k})$ be an approximation of the function $f(x) = \sum_{i=1}^{n+1} h_i(x)$, for which $L_{n+1}(r_f) = 0$, where

$$\phi_\epsilon(x, l, u) = \left(1 - \frac{l - x}{\epsilon}\right) \cdot \mathbb{1}\{l - \epsilon \leq x < l[\} + \mathbb{1}\{l \leq x \leq u\} + \mathbb{1}\{u < x \leq u + \epsilon\} \cdot \left(1 - \frac{x - u}{\epsilon}\right)$$

is an approximation of the function $x \mapsto \mathbb{1}\{l \leq x \leq u\}$. Since the functions $\phi_\epsilon(\cdot, l, u)$ are continuous, it also implies that their sum, f'_ϵ, is continuous. Now, note that $f'_\varepsilon(x) = f(x) + er_\epsilon(x)$ where $er_\epsilon(x) = \sum_{i=1}^{n+1} \sum_{k=1}^N (1 - (l_{i,k} - x)/\epsilon)\mathbb{1}\{x \in [l_{i,k} - \epsilon, l_{i,k}[\} + (1 - (x - u_{i,k})/\epsilon)\mathbb{1}\{x \in]u_{i,k}, u_{i,k} + \epsilon]\}$. For any $\epsilon < \epsilon_1 = \min\{|x_1 - x_2| : x_1 \neq x_2 \in \{X_{(i)}\}_{i=1}^{n+1} \cup \{l_{i,k}\}_{i=1 \dots n+1}^{k=1 \dots N} \cup \{u_{i,k}\}_{i=1 \dots n+1}^{k=1 \dots N}\}$ small enough and any $i \in \{1 \dots n+1\}$, we have that $er_\epsilon(X_{(i)}) = 0$. We deduce that $L_{n+1}(r_{f'_\epsilon}) = L_{n+1}(r_f) = 0$ whenever $\epsilon < \epsilon_1$. Using the same decomposition, one can easily show that for any $\epsilon < \epsilon_2 = \min\{|x_1 - x_2| : x_1 \neq x_2 \in \{l_{i,k}\}_{i=1 \dots n+1}^{k=1 \dots N} \cup \{u_{i,k}\}_{i=1 \dots n+1}^{k=1 \dots N}\}/2$ all the level sets of f_ϵ are a union of at most N segements (convex sets). Putting previous statements altogether and denoting $\epsilon^\star = \min(\epsilon_1, \epsilon_2)/2$, gives us that $r_{f'_{\epsilon^\star}} \in \mathcal{R}_{\mathcal{C},N}$ and $L_{n+1}(r_{f'_{\epsilon^\star}}) = 0$. □

Proof of Proposition 8. Fix any $d \in \mathbb{N}^\star$, let $N = 1$, and let $\{(X_i, f(X_i))\}_{i=1}^n$ be any sample satisfying $f(X_{(1)}) < f(X_{(2)}) < \dots < f(X_{(n+1)})$.

($\Rightarrow$) Assume that $\min_{r\in\mathcal{R}_{\mathcal{C},1}} L_{n+1}(r) = 0$ and let $r \in \mathcal{R}_{\mathcal{C},1}$ be any ranking that perfectly ranks the sample. Since r perfectly ranks the sample, for any $k \in \{1 \dots n+1\}$, $r(X_{(j)}, X_{(k)}) = 2\mathbb{1}\{(j) > (k)\} - 1 \,\forall j \neq k$ and since r is a convex ranking, the set $\{x \in \mathcal{X} : r(x, X_{(k)})\}$ is a convex set. Now, fix any $k \in \{1 \dots n\}$. Using the previous statements, we get that $\{X_{(i)}\}_{i=k+1}^{n+1} \in \{x \in \mathcal{X} : r(x, X_{(k+1)}) \geq 0\}$ and $X_{(k)} \notin \{x \in \mathcal{X} : r(x, X_{(k+1)}) \geq 0\}$. Since CH$\{X_{(i)}\}_{i=k+1}^{n+1}$ is the smallest convex set that contains $\{X_{(i)}\}_{i=k+1}^{n+1}$ and since $\{x \in \mathcal{X} : r(x, X_{(k+1)}) \geq 0\}$ is also a convex set that contains $\{X_{(i)}\}_{i=k}^{n+1}$, necessarily CH$\{X_{(i)}\}_{i=k+1}^{n+1} \subseteq \{x \in \mathcal{X} : r(x, X_{(k+1)}) \geq 0\}$. Combining the previous statements gives us that $X_{(k)} \notin$ CH$\{X_{(i)}\}_{i=k+1}^{n+1}$. Therefore, using the definition of convex hull, we know that there does not exist any $(\lambda_1 \dots \lambda_k) \in \mathbb{R}^k$ such that $\sum_{i=1}^{k} \lambda_i X_{(n+2-i)} = X_{(n+1-k)}$, $\sum_{i=1}^{k} \lambda_i = 1$ and $\lambda_i \geq 0$. Rewriting this constraint in a matricial form gives us that the polyhedron

$$\{\lambda \in \mathbb{R}^k : \mathrm{M}_k \lambda^T = X_{(n+1-k)}^T,\ \langle \vec{1}, \lambda \rangle = 1,\ \lambda \succeq \vec{0}\}$$

where M_k is the (d, k)-matrix where its i-th column is equal to $X_{(n+2-i)}^T$, is empty.

($\Leftarrow$) Assume that the cascade of polyhedrons is empty. Reproducing the (inverse) same steps as in the first equivalence of the proof, we get that $X_{(k)} \notin$ CH$\{X_{(i)}\}_{i=k+1}^{n+1}$ for any $k \in \{1 \dots n\}$. Therefore the convex hulls form a nested sequence CH$\{X_{(n+1)}\} \subset$ CH$\{X_{(i)}\}_{i=n}^{n+1} \subset \dots \subset$ CH$\{X_{(i)}\}_{i=1}^{n+1}$. Now, let $f'_\epsilon(x) = \sum_{i=1}^{n+1} \phi_{\epsilon,i}(x)$ be an approximation of the function $f(x) = \sum_{i=1}^{n+1} \mathbb{1}\{x \in \mathrm{CH}\{X_{(j)}\}_{j=i}^{n+1}\}$, for which $L_{n+1}(r) = 0$, where for any $i \in \{1 \dots n+1\}$

$$\phi_{\epsilon,i}(x) = \begin{cases} 1 - d(x, B(\mathrm{CH}\{X_{(j)}\}_{j=i}^{n+1}, 2(n+1-i)\epsilon))/\epsilon & \text{if } \ d(x, B(\mathrm{CH}\{X_{(j)}\}_{j=i}^{n+1}, 2(n+1-i)\epsilon) \leq \epsilon \\ 0 & \text{otherwise.} \end{cases}$$

is an approximation of the function $x \mapsto \mathbb{1}\{x \in CH\{X_{(j)}\}_{j=i}\}^n$. For any convex set $\mathcal{X}$, the function $x \mapsto d(x, \mathcal{X})$ is continuous. we deduce that the functions $\{\phi_{\epsilon,i}\}_{i=1}^{n+1}$ are continuous and so is their sum f'_ϵ. First, using the same decomposition as in the proof of Proposition 7, one can show that $L_{n+1}(r_{f'_\epsilon}) = L_{n+1}(r_f) = 0$ for any $\epsilon < \epsilon^\star = \min_{i=1\dots n} d(X_{(i)}, \mathrm{CH}\{X_{(j)}\}_{j=i+1}^{n+1})/(2n+2)$ small enough. Secondly, using the fact that for any convex set $\mathcal{X} \subset \mathbb{R}^d$, its ϵ-Ball $B(\mathcal{X}, \epsilon)$ is also a convex set (see Lemma 6), for any $\epsilon < \epsilon^\star$ and any $x' \in \mathcal{X}$, we have that the level set $\{x \in \mathcal{X} : f_\epsilon(x) \geq f_\epsilon(x')\}$ is a convex set. Putting the previous statements altogether gives us that $r_{f'_\epsilon} \in \mathcal{R}_{\mathcal{C},1}$ and $L_{n+1}(r_{f'_\epsilon}) = 0$. □

References

Boucheron, Stéphane, Lugosi, Gábor, and Massart, Pascal. *Concentration inequalities: A nonasymptotic theory of independence*. OUP Oxford, 2013.

Clémençon, Stéphan, Lugosi, Gabor, and Vayatis, Nicolas. Ranking and empirical minimization of u-statistics. *The Annals of Statistics*, pp. 844–874, 2010.

Zabinsky, Zelda B and Smith, Robert L. Pure adaptive search in global optimization. *Mathematical Programming*, 53 (1-3):323–338, 1992.