General parallel optimization without a metric

Xuedong Shang XUEDONG.SHANG@INRIA.FR
SequeL team, INRIA Lille - Nord Europe, France
40 avenue Halley 59650, Villeneuve d'Ascq, France

Emilie Kaufmann EMILIE.KAUFMANN@UNIV-LILLE.FR
CNRS & ULille, UMR 9189 (CRIStAL), SequeL team, INRIA Lille - Nord Europe, France
40 avenue Halley 59650, Villeneuve d'Ascq, France

Michal Valko MICHAL.VALKO@INRIA.FR
SequeL team, INRIA Lille - Nord Europe, France
40 avenue Halley 59650, Villeneuve d'Ascq, France

Editors: Aurélien Garivier and Satyen Kale

Abstract

Hierarchical bandits are an approach for global optimization of *extremely* irregular functions. This paper provides new elements regarding POO, an adaptive meta-algorithm that does not require the knowledge of local smoothness of the target function. We first highlight the fact that the subroutine algorithm used in POO should have a small regret under the assumption of *local smoothness with respect to the chosen partitioning*, which is unknown if it is satisfied by the standard subroutine HOO. In this work, we establish such regret guarantee for HCT, which is another hierarchical optimistic optimization algorithm that needs to know the smoothness. This confirms the validity of POO. We show that POO can be used with HCT as a subroutine with a regret upper bound that matches the one of best-known algorithms using the knowledge of smoothness up to a $\sqrt{\log n}$ factor. On top of that, we propose a general wrapper, called GPO, that can cope with algorithms that only have simple regret guarantees. Finally, we complement our findings with experiments on difficult functions.

Keywords: continuously-armed bandits, global optimization, black-box optimization

1. Introduction

Global optimization (GO) has applications in several domains including hyper-parameter tuning (Jamieson and Talwalkar, 2016; Li et al., 2017; Samothrakis et al., 2013). GO usually consists of a data-driven optimization process over an expensive-to-evaluate function. It is also known as *black-box optimization* since the inner behavior of a function is often unknown.

In GO, we optimize an unknown and costly-to-evaluate function $f : \mathcal{X} \to \mathbb{R}$ based on n noisy evaluations, that can be sequentially selected. This setting is a generalization of *multi-armed bandits*, where the arm space $\mathcal{X}$ is some measurable space (Bubeck et al. 2011). Each *arm* $x \in \mathcal{X}$ gets its mean reward $f(x)$ through the reward function f, which is the function to be optimized. At each round t, the learner chooses an arm $x_t \in \mathcal{X}$ and receives a reward r_t. We study the noisy setting in which the obtained reward is a noisy evaluation of f: $r_t \triangleq f(x_t) + \varepsilon_t$, where ε_t is a bounded noise.

Treating the setting without any further assumption would be a *mission impossible*. However, the setting gets easier if we assume a global smoothness of the reward function (Agrawal, 1995;

Kleinberg, 2005; Kleinberg et al., 2008; Cope, 2009; Auer et al., 2007; Slivkins, 2011; Kleinberg et al., 2015). A weaker condition is some *local* smoothness where only neighborhoods around the maximum are required to be smooth. In fact, local smoothness is sufficient for achieving near-optimality (Valko et al., 2013; Azar et al., 2014; Grill et al., 2015; Bull, 2015). We base our work on optimistic tree-based optimization algorithms (Munos, 2011; Valko et al., 2013; Preux et al., 2014; Azar et al., 2014) that approach the problem with a hierarchical partitioning of the arm space and take the *optimistic principle*. This idea comes from *planning* in Markov decision processes (Kocsis and Szepesvári, 2006; Munos, 2014; Grill et al., 2016).

Our work is motivated by the **p**arallel **o**ptimistic **o**ptimization (POO) approach proposed by Grill et al. (2015), that *adapts to the smoothness* without the knowledge of it. POO is a *meta-algorithm* which can be used on top of any hierarchical optimization algorithm that *knows the smoothness*, that we call a subroutine. Not only does POO require only the mildest local regularity conditions, but it also gets rid of the unnecessary metric assumption that is often required. Local smoothness naturally covers a larger class of functions than global smoothness, yet still assures that the function does not decrease too fast around the maximum. We highlight that the analysis of POO is modular: Assuming the subroutine has a regret of order R_n *under a local smoothness assumption with respect to a fixed partitioning* (Grill et al. 2015, Assumption 1, formally introduced in Section 2), POO run with such subroutine has a regret bounded by $R_n\sqrt{\log n}$. POO was originally analyzed using HOO as a subroutine. However, unlike what Grill et al. (2015) hypothesize, it is non-trivial to provide a regret bound for HOO under Assumption 1. We elaborate on that in Section 3. In order to validate POO, there needs to exist a subroutine with a regret guarantee that is provable under Assumption 1. This is what we deliver.

In particular, we prove that HCT-iid[1] of Azar et al. (2014) satisfies the required regret guarantee, and is, therefore, a desirable subroutine to be plugged in POO. Similar to HOO, HCT is a hierarchical optimization algorithm based on confidence intervals. However, unlike HOO, these confidence intervals are obtained by repeatedly sampling a representative point of each cell in the partitioning before splitting the cell. This yields partition trees that have a *controlled depth*, which are easier to analyze under a local smoothness assumption with respect to the partitioning. Whether HOO has similar regret guarantees under the desired local metricless assumption remains an open question.

POO requires the subroutine to have a *cumulative* regret guarantee. In this paper, we also provide a more general wrapper for algorithms that only have guarantee for their simple regret, called GPO (for **g**eneral **p**arallel **o**ptimization). We show that with a cross-validation scheme instead of the original recommendation strategy, any hierarchical bandit algorithm with simple regret guarantee can be plugged into GPO with only a tiny increase in the resulting simple regret.

Paper outline We first formulate the sequential optimization problem and introduce some preliminary notions and assumptions in Section 2. Our main result is presented in Section 3, where we provide a regret upper bound for HCT under local smoothness with respect to the partitioning. In Section 4, we present the instantiation of POO studied in the paper that we call PCT, in which the underlying subroutine HOO is replaced by HCT. We show that PCT enjoys the same regret bound as HCT up to a $\sqrt{\log n}$ factor. The general wrapper and its simple regret analysis are presented in Section 5. We conclude by some numerical simulations in Section 6.

[1] Denoted by HCT in the rest of the paper since we do not consider the correlated feedback setting.

2. Smoothness assumptions for black-box optimization

Let $\mathcal{X}$ be a measurable space. Our goal is to find the maximum of an unknown noisy function $f : \mathcal{X} \to \mathbb{R}$ of which the cost of evaluation is high, given a total budget of n evaluations. At each round t, a learner selects a point $x_t \in \mathcal{X}$ and observes a reward $r_t \triangleq f(x_t) + \varepsilon_t$, bounded by $[0, 1]$, from the environment where the noise ε_t is assumed to be independent from previous observations and such that $\mathbb{E}[\varepsilon_t | x_t] = 0$. After n evaluations, the algorithm outputs a guess for the maximizer, denoted by $x(n)$. We assume that there exists at least one $x^\star \in \mathcal{X}$ s.t. $f(x^\star) \triangleq \sup_{x\in\mathcal{X}} f(x)$, denoted by $f^\star$ in the following. We measure the performance by the *simple regret*, also called the *optimization error*,

$$S_n \triangleq f^\star - f(x(n)).$$

Another related notion is the cumulative regret, defined as

$$R_n \triangleq nf^\star - \sum_{t=1}^{n} f(x_t).$$

As observed by Bubeck et al. (2009), a good cumulative regret naturally implies a good simple regret: If we recommend $x(n)$ according to the distribution of previous plays, we immediately get $\mathbb{E}[S_n] = \mathbb{E}[R_n]/n$.

2.1. Covering tree that guides the optimization

Hierarchical bandits rely on the existence of hierarchical partitioning $\mathcal{P} \triangleq \{\mathcal{P}_{h,i}\}_{h,i}$ defined recursively, where

$$\mathcal{P}_{0,1} = \mathcal{X}, \quad \mathcal{P}_{h,i} = \bigcup_{j=0}^{K-1} \mathcal{P}_{h+1,Ki-j}.$$

Such a partition can be naturally represented by a tree, where K denotes the maximum number of children of a node in that tree. Many of known algorithms depend on a metric/dissimilarity over the search space to define the regularity assumptions that link the partitioning to some near-optimality dimension, that is independent of the partitioning. However, this was shown to be artificial (Grill et al., 2015), since (i) the metric is not fully exploited by the algorithms and (ii) the notion of near-optimality dimension independent of partitioning is ill-defined. Hence, it is natural to make smoothness assumptions directly related only to the partitioning.

We now present *the only regularity assumption* on the target function f that is expressed in terms of the partitioning $\mathcal{P}$. We stress again that requiring only local smoothness assumptions is an improvement since (i) it covers a larger class of functions, (ii) it only constrains f along the optimal path of the covering tree which is a plausible property in an optimization scenario, and (iii) shows that the optimization is actually easier than it was previously believed.

Assumption 1 (local smoothness w.r.t. $\mathcal{P}$) *For $x^\star$ be a global maximizer, we denote by $i_h^\star$ be the index of the only cell at depth h that contains $x^\star$. Then, there exist a global maximizer $x^\star$ and two constants $\nu > 0$, $\rho \in (0, 1)$ s.t.,*

$$\forall h \geq 0, \forall x \in \mathcal{P}_{h,i_h^\star}, \;\; f(x) \geq f^\star - \nu\rho^h.$$

Note that this assumption is the same as the one of Grill et al. (2015). Multiple maximizers may exist, but this assumption needs to be satisfied only by one of them.

As first observed by Auer et al. (2007), the difficulty of a GO should depend on the size of near-optimal regions and on how fast they shrink. Auer et al. (2007) use a margin condition that quantifies this difficulty by the volume of near-optimal regions. In this work, we use a similar notion of near-optimality dimension instead. This notion is directly related to the partitioning.

Definition 1 (near-optimality dimension w.r.t. $\mathcal{P}$) *For any $\nu > 0$, $C > 1$, and $\rho \in (0,1)$, we define the near-optimality dimension of f with respect to $\mathcal{P}$ as*

$$d(\nu, C, \rho) \triangleq \inf\left\{d' \in \mathbb{R}^+ : \forall h \geq 0, \mathcal{N}_h(3\nu\rho^h) \leq C\rho^{-d'h}\right\},^{2}$$

where $\mathcal{N}_h(\varepsilon)$ is the number of cells $\mathcal{P}_{h,i}$ such that $\sup_{x\in\mathcal{P}_{h,i}} f(x) \geq f^\star - \varepsilon$.

$\mathcal{N}_h(3\nu\rho^h)$ can be thought as the number of cells that any algorithm needs to sample in order to find the maximum. A smaller $d(\nu, C, \rho)$ implies an easier optimization problem.

3. HCT under local smoothness with respect to $\mathcal{P}$

Analyzing HOO under Assumption 1 is not trivial. A key lemma in the analysis of HOO (Lemma 3 by Bubeck et al. 2011) that controls the variance of near-optimal cells *is not true* under local smoothness assumptions as Assumption 1. Indeed, HOO could induce a very deep covering tree, while producing too many nodes that are neither near-optimal nor sub-optimal. The concept of near-optimal and sub-optimal nodes is then characterized by the *sub-optimality gap* of each node which measures the distance between the local maximum of the node and the global maximum. Intuitively, nodes that are neither near-optimal nor sub-optimal represent the nodes of whom the sub-optimality gap is neither too large nor too small. To control the regret due to these nodes, Bubeck et al. (2011) use global smoothness (weakly Lipschitz) assumption. Assumption 1 is weaker, only local, and does not offer such comfort. If we want to control the regret due to these nodes without Lemma 3 of Bubeck et al. (2011), one possible way is to control the depth of the covering tree to ensure that we do not have too many of them. In particular, another algorithm known as HCT (Azar et al., 2014) implies a controlled depth of the tree which allows it to be analyzed under Assumption 1 as opposed to HOO. We now give a brief description of HCT and present a new analysis of it.

3.1. Description of HCT

The pseudocode of HCT (Algorithm 1) and two detailed snippets (Algorithm 2 and Algorithm 3) describe the process of traversing the covering tree. The algorithm stores a finite subtree $\mathcal{T}_t$ at each round t which is initialized by $\mathcal{T}_0 = \{(0,1)\}$. Each cell is associated with a representative point $x_{h,i}$ and the algorithm keeps track of some statistics regarding this point. One of these statistics is the empirical mean reward $\widehat{\mu}_{h,i}(t)$ which is the average on the first $T_{h,i}(t)$ rewards received when querying $x_{h,i}$. The HCT algorithm also keeps track of an upper confidence bound U-value for the cell (h,i),

$$U_{h,i}(t) \triangleq \widehat{\mu}_{h,i}(t) + \nu\rho^h + c\sqrt{\frac{\log(1/\widetilde{\delta}(t^+))}{T_{h,i}(t)}},$$

[2] This definition is slightly different from the original POO paper, where a coefficient 3 is present instead of 2 due to a technical detail.

Algorithm 1: High confidence tree (HCT, Azar et al. 2014)

```
Input     : $K$, $\nu > 0$, $\rho \in (0,1)$, $c > 0$, tree partitioning $\{\mathcal{P}_{h,i}\}$, confidence $\delta$
Initialize: $\mathcal{T}_1 \leftarrow \{(0,1),(1,1),\dots,(1,K)\}$, $U_{1,1}(1) \leftarrow \cdots \leftarrow U_{1,K}(1) \leftarrow +\infty$
for $t \leftarrow 1$ to $n$ do
    if $t = t^+$ then
        for $(h,i) \in \mathcal{T}_t$ do
            $U_{h,i}(t) \leftarrow \widehat{\mu}_{h,i}(t) + \nu\rho^h + c\sqrt{\frac{\log\left(1/\widetilde{\delta}(t^+)\right)}{T_{h,i}(t)}}$
        end
        UpdateBackward($\mathcal{T}_t, t$)
    end
    $(h_t, i_t), P_t \leftarrow$ OptTraverse($\mathcal{T}_t, t$)
    Evaluate $x_{h_t,i_t}$ and obtain $r_t$
    $T_{h_t,i_t}(t) \leftarrow T_{h_t,i_t}(t) + 1$
    Update $\widehat{\mu}_{h_t,i_t}(t)$
    $U_{h_t,i_t}(t) \leftarrow \widehat{\mu}_{h_t,i_t}(t) + \nu\rho^{h_t} + c\sqrt{\frac{\log\left(1/\widetilde{\delta}(t^+)\right)}{T_{h_t,i_t}(t)}}$
    UpdateBackward($P_t, t$)
    $\tau_{h_t}(t) \leftarrow \left\lceil \frac{c^2 \log(1/\widetilde{\delta}(t^+))}{\nu^2}\rho^{-2h_t} \right\rceil$
    if $T_{h_t,i_t}(t) \geq \tau_{h_t}(t)$ and $(h_t, i_t)$ is a leaf then
        Expand($(h_t, i_t)$)
    end
end
```

Algorithm 2: OptTraverse

```
Input     : a tree $\mathcal{T}$, round $t$
Initialize: $(h,i) \leftarrow (0,1)$; $P \leftarrow \{(0,1)\}$; $T_{0,1}(t) = \tau_0(t) = 1$
while $(h,i)$ is not a leaf of $\mathcal{T}$ and $T_{h,i}(t) \geq \tau_h(t)$ do
    $j \leftarrow \operatorname*{arg\,max}_{j \in \{0,\dots,K-1\}} \{B_{h+1,Ki-j}(t)\}$
    $(h,i) \leftarrow (h+1, Ki-j)$
    $P \leftarrow P \cup \{(h,i)\}$
end
return $(h,i)$ and $P$
```

where $t^+ \triangleq 2^{\lceil \log_2(t) \rceil}$, $\widetilde{\delta}(t) \triangleq \min\{c_1\delta/t, 1/2\}$, and its corresponding B-value,

$$B_{h,i}(t) \triangleq \begin{cases} \min\left\{U_{h,i}(t), \max\limits_{j \in \{0,\dots,K-1\}} \{B_{h+1,Ki-j}(t)\}\right\} & \text{if } (h,i) \text{ is an internal node,} \\ U_{h,i}(t) & \text{otherwise,} \end{cases}$$

```
Algorithm 3: UpdateBackward
Input   : a tree $\mathcal{T}$, round $t$
          note that $P_t$ can also be considered as a tree, thus input of this function
for $(h, i) \in \mathcal{T}$ backward from each leaf of $\mathcal{T}$ do
    if $(h, i)$ is a leaf of $\mathcal{T}$ then
        $B_{h,i}(t) \leftarrow U_{h,i}(t)$
    else
        $B_{h,i}(t) \leftarrow \min\left\{U_{h,i}(t), \max_{j\in\{0,\ldots,K-1\}}\{B_{h+1,Ki-j}(t)\}\right\}$
    end
end
```

which is designed to be a tighter upper confidence bound than the U-value. Here, c and c_1 are two constants, and $\nu\rho^h$ represents the *resolution*[3] of the region $\mathcal{P}_{h,i}$. Observe that $U_{h,i}(t)$ and $B_{h,i}(t)$ are not updated at every round, but are constant on time intervals of the form $[2^k, 2^{k+1})$.

At each round t, the algorithm traverses the current covering tree along an *optimistic path* P_t before choosing a point (`OptTraverse` function). This optimistic path P_t is obtained by repeatedly selecting cells that have a larger B-value until a leaf or a node that is sampled less than a certain number of times is reached. If a leaf is reached, then this leaf is sampled and expanded (i.e., we split the leaf into K equal-sized regions and initialize their U-values to $+\infty$); otherwise, the node that is not sampled enough is re-sampled. All the B-values along the optimistic path are then updated backwardly from the current node to the root (`UpdateBackward` function). More precisely, `HCT` samples one node a certain number of times $\tau_h(t)$ in order to sufficiently reduce the uncertainty before expanding it. Hence, $\tau_h(t)$ is defined such that the uncertainty over the rewards in $\mathcal{P}_{h,i}$ is roughly equal to the resolution of the node,

$$\tau_h(t) \triangleq \left\lceil \frac{c^2 \log(1/\widetilde{\delta}(t^+))}{\nu^2}\rho^{-2h} \right\rceil.$$

3.2. Analysis of `HCT` under a local *metricless* assumption

We now state our main theorem. We prove that `HCT` achieves an expected regret bound under Assumption 1 which matches the regret bound given by Azar et al. (2014) up to constants. Moreover, compared to that result, the near-optimality dimension d featured in Theorem 2 is the one of Definition 1 that is defined with respect to the partitioning and not with respect to a metric. For a fixed budget n, we introduce the notation $\texttt{HCT}(\nu, \rho)$ to refer to the instantiation of `HCT` parameterized by ν, ρ, $c = 2\sqrt{1/(1-\rho)}$ and $\delta = 1/n$.

Theorem 2 *Assume that function f satisfies Assumption 1. Then, setting $\delta \triangleq 1/n$, the cumulative regret of $\texttt{HCT}(\nu, \rho)$ after n function evaluations is upper bounded as*

$$\mathbb{E}[R_n^{\texttt{HCT}(\nu,\rho)}] \leq \alpha C (\log n)^{1/(d(\nu,C,\rho)+2)} n^{(d(\nu,C,\rho)+1)/(d(\nu,C,\rho)+2)},$$

where α is a numerical constant and C is the constant associated to $d(\nu, C, \rho)$.

[3] The term *resolution* refers to the maximum variation in the cell. If it is too large, then we need to shrink the volume, thus increase the resolution.

As a consequence, by simply applying the recommendation strategy that follows the distribution of previous plays, we get the following simple-regret bound.

Corollary 3 *The simple regret of* HCT *after* n *function evaluations under Assumption 1 satisfies*

$$\mathbb{E}[S_n^{\texttt{HCT}(\nu,\rho)}] \leq \alpha C (\log n)^{1/(d(\nu,C,\rho)+2)} n^{-1/(d(\nu,C,\rho)+2)}.$$

We now sketch the proof. The full proof follows the analysis of Azar et al. (2014) and is detailed in Appendix A. As mentioned above, HCT has a controlled depth. Indeed, given the threshold $\tau_h(t)$ required at depth h, in Section A.2, we prove that the depth of the covering tree is bounded as stated in the following lemma.

Lemma 4 *The depth of the covering tree produced by* HCT after n function evaluations satisfies

$$H(n) \leq H_{\max}(n) \triangleq \left\lceil \frac{1}{2(1-\rho)} \log\left(\frac{n\nu^2}{c^2\rho^2}\right)\right\rceil.$$

Defining the mean reward $\mu_{h,i} \triangleq f(x_{h,i})$, we introduce a favorable event under which the mean reward of all expanded nodes is within a confidence interval,

$$\xi_t \triangleq \left\{\forall (h,i) \in \mathcal{L}_t, |\widehat{\mu}_{h,i}(t) - \mu_{h,i}| \leq c\sqrt{\log(1/\widetilde{\delta}(t))/T_{h,i}(t)}\right\},$$

where $\mathcal{L}_t$ is the set of all possible nodes in trees of maximum depth $H_{\max}(t)$.

We split the regret into two parts depending on whether ξ_t holds or not. In Appendix A.4, we prove that the failing confidence term is with high probability bounded by $\sqrt{n}$. In the case when ξ_t holds, we bound the regret in Appendix A.5 by treating separately the two parts, Δ_{h_t,i_t} and $\widehat{\Delta}_t$, of the instantaneous regret Δ_t,

$$\Delta_t \triangleq f^\star - r_t = f^\star - f(x_{h_t,i_t}) + f(x_{h_t,i_t}) - r_t = \Delta_{h_t,i_t} + \widehat{\Delta}_t.$$

Next, we bound $\widehat{\Delta}_t$ by Azuma-Hoeffding concentration inequality (Azuma, 1967). Then, we bound Δ_{h_t,i_t} with the help of the following lemma, which is the major difference compared to the original HCT analysis by Azar et al. (2014). In particular, the lemma states that if Assumption 1 is verified then $f^\star$ is upper-bounded by the U-value of any optimal node.

Lemma 5 *Under Assumption 1 and under event* ξ_t*, we have that for any optimal node* $(h^\star, i^\star)$*,* $U_{h^\star,i^\star}(t)$ *is an upper bound on* $f^\star$*.*

Proof Since $t^+ \geq t$, we have

$$U_{h^\star,i^\star}(t) \triangleq \widehat{\mu}_{h^\star,i^\star}(t) + \nu\rho^{h^\star} + c\sqrt{\frac{\log(1/\widetilde{\delta}(t^+))}{T_{h^\star,i^\star}(t)}} \geq \widehat{\mu}_{h^\star,i^\star}(t) + \nu\rho^{h^\star} + c\sqrt{\frac{\log(1/\widetilde{\delta}(t))}{T_{h^\star,i^\star}(t)}}.$$

Moreover, as we are under event ξ_t, we also have

$$\widehat{\mu}_{h^\star,i^\star}(t) + c\sqrt{\frac{\log(1/\widetilde{\delta}(t))}{T_{h^\star,i^\star}(t)}} \geq f(x_{h^\star,i^\star}).$$

Therefore, $U_{h^\star,i^\star}(t) \geq f(x_{h^\star,i^\star}) + \nu\rho^{h^\star} \geq f^\star$. ■

With the help of Lemma 5 (see Step 2 in Appendix A.5), we can then upper bound Δ_{h_t,i_t} as

$$\Delta_{h_t,i_t} \leq 3c\sqrt{\frac{\log(2/\widetilde{\delta}(t))}{T_{h_t,i_t}(t)}}.$$

To bound the total regret of the all nodes selected, we divide them into two categories, depending on whether their depth is smaller or equal than $\overline{H}$ (to be optimized later) or not.

For the nodes in depths $h \leq \overline{H}$, we use Lemma 5 again, now to show that `OptTraverse` only selects nodes that have a parent which is $(3\nu\rho^{h_t-1})$-optimal. For the nodes for which $h > \overline{H}$, we bound the regret using the selection rule of `HCT`.

The sums of the regrets from the two categories are proportional and inversely proportional to an increasing function of $\overline{H}$. By finding the value of $\overline{H}$ for which the sum of the two terms reaches its minimum and adding the regret coming from the situations where the favorable event does not hold, gives us the following cumulative regret for `HCT`: With probability $1-\delta$,

$$R_n^{\texttt{HCT}(\nu,\rho)} \leq \mathcal{O}\Big((\log(n/\delta))^{1/(d(\nu,C,\rho)+2)} n^{(d(\nu,C,\rho)+1)/(d(\nu,C,\rho)+2)}\Big).$$

However, the analysis of `POO` requires a bound on the expected regret of the underlying subroutine. For that purpose, we simply set $\delta \triangleq 1/n$ and that gives us the statement of Theorem 2, and consequently Corollary 3.

4. `PCT` that does not need to know the smoothness

In this section, we formally introduce a generic `POO`$(\mathcal{A})$ algorithm, taking as input *any* algorithm $\mathcal{A} = \mathcal{A}(\nu,\rho)$ requiring the smoothness parameters. We provide a simple regret upper bound for a particular instantiation called `PCT`, for `POO(HCT)`.

4.1. Generic parallel optimistic optimization

`POO`$(\mathcal{A})$ (**p**arallel **o**ptimistic **o**ptimization) is a meta-algorithm that uses any hierarchical optimization algorithm $\mathcal{A}$ that knows the smoothness as a subroutine, originally proposed by Grill et al. (2015) for $\mathcal{A} =$ `HOO`. In this algorithm, several instances of $\mathcal{A}$ are run in parallel, each one using a different pair of parameters (ν,ρ) in a well-chosen grid $\mathcal{G}$ (defined in Line 4 of Algorithm 4). In the end, `POO`$(\mathcal{A})$ chooses the instance that has the largest empirical mean reward and returns one of the points evaluated by this instance, chosen uniformly at random.

The pseudocode of `POO`$(\mathcal{A})$ is shown in Algorithm 4. Additionally to the base algorithm itself, it requires two parameters $\rho_{\max}$ and $\nu_{\max}$ that determine the range of $\mathcal{A}(\nu,\rho)$ instances that we can compete with. However, these parameters can be set as a function of the number of evaluations as explained in details in Appendix C of Grill et al. (2015), hence not mandatory in practice. An important remark is that given a budget n of function evaluations, the number of $\mathcal{A}$ instances N run by `POO`$(\mathcal{A})$ depends on n, and each instance is run for $\lfloor n/N(n)\rfloor$ times. Due to the doubling scheme used in Lines 2-10, note however that `POO`$(\mathcal{A})$ does not need to know this total number of function evaluations. Hence, if the base algorithm $\mathcal{A}$ is anytime, so is `POO`$(\mathcal{A})$.

Algorithm 4: POO($\mathcal{A}$) - parallel optimistic optimization with base algorithm $\mathcal{A}$

```
Input          : base algorithm $\mathcal{A}$, $\nu_{\max}$,$\rho_{\max}$, branching factor of the partitioning $K$
Initialization: $D_{\max} \leftarrow \ln K / \ln(1/\rho_{\max})$, number of function evaluations $n \leftarrow 0$, current
                 number of instances of $\mathcal{A}$: $N \leftarrow 1$, $\mathcal{S} \leftarrow \{(\nu_{\max}, \rho_{\max})\}$
while budget still available do
    while $N \leq \frac{1}{2} D_{\max} \log(n/(\log n))$ do
        for $i \leftarrow 1, \dots, N$ do
            $s \leftarrow \left(\nu_{\max}, {\rho_{\max}}^{2N/(2i+1)}\right)$
            Initialize $\mathcal{A}(s)$ (if not already done before).
            Continue running $\mathcal{A}(s)$ until it has given $\frac{n}{N}$ rewards $r_{s,1}, \dots, r_{s,n/N}$.
            Compute the average reward $\widehat{\mu}[s] = \frac{N}{n} \sum_{i=1}^{n/N} r_{s,i}$.
        end
        $n \leftarrow 2n$, $N \leftarrow 2N$
    end
    Perform each $\mathcal{A}(s)$ once and update the average reward $\widehat{\mu}[s]$.
    $n \leftarrow n + N$
end
$s^\star \leftarrow \mathrm{argmax}_{s \in \mathcal{S}} \widehat{\mu}[s]$ return A point sampled u.a.r. from the points evaluated by $\mathcal{A}(s^\star)$
```

4.2. Upper bound on the simple regret of PCT

Building on our new analysis of the HCT algorithm, we are able to provide theoretical guarantees for the resulting POO(HCT) algorithm, that we refer to as PCT (**p**arallel **c**onfidence **t**ree). More precisely we define PCT(δ) as POO run on top of HCT using confidence parameter δ.

Letting $(\nu^\star, C^\star, \rho^\star)$ be a triple of parameters for which Assumption 1 is true, we prove that PCT achieves a regret that is comparable to the one obtained by c.

Theorem 6 *Assume that the target function f satisfies Assumption 1 and $\nu^\star \leq \nu_{\max}$ and $\rho^\star \leq \rho_{\max}$. For $\delta = N(n)/n$ with $N(n) = \lceil (1/2) D_{\max} \log(n/\log n) \rceil$, the simple regret of* PCT(δ) *after n function evaluations is bounded as*

$$\mathbb{E}[S_n^{\texttt{PCT}(\delta)}] \leq \beta D_{\max} (\nu_{\max}/\nu^\star)^{D_{\max}} \left(((\log^2 n)/n)^{1/(d(\nu^\star, C^\star, \rho^\star)+2)} \right),$$

where β is a constant independent of $\nu_{\max}$ and $\rho_{\max}$.[4]

By Corollary 3, we know that the simple regret of HCT after n function evaluations run with $(\nu^\star, C^\star, \rho^\star)$ is of order $\mathcal{O}\left((\log n/n)^{1/(d(\nu^\star, C^\star, \rho^\star)+2)}\right)$. As a consequence, the performance of PCT is at most a $\sqrt{\log n}$ factor away from that of the best HCT instance.

Theorem 6 follows from Corollary 3 and Proposition 7 below. This wrapper result highlights how cumulative regret guarantees for *any* base algorithm translate into simple regret guarantees for the corresponding POO($\mathcal{A}$) algorithm. Its proof almost replicates the analysis of POO(HOO) by Grill et al. (2015) and we provide it in Appendix B for the sake of completeness.

[4] More generally, Theorem 6 holds for any $\nu \leq \nu_{\max}$ and $\rho \leq \rho_{\max}$.

Proposition 7 *If for all* (ν, ρ) *the* $\mathcal{A}(\nu, \rho)$ *algorithm has its* cumulative regret *bounded as*

$$\mathbb{E}\left[R_n^{\mathcal{A}(\nu,\rho)}\right] \le \alpha C (\log n)^{1/(d(\nu,C,\rho)+2)} n^{(d(\nu,C,\rho)+1)/(c+2)}, \tag{1}$$

for any function f *satisfying Assumption 1 with parameters* (ν, C, ρ)*, then there exists a constant* β *that is independent of* $\nu_{\max}$ *and* $\rho_{\max}$ *such that*

$$\mathbb{E}\left[S_n^{\texttt{POO}(\mathcal{A})}\right] \le \beta D_{\max}(\nu_{\max}/\nu^\star)^{D_{\max}}\left((\log^2 n)/n\right)^{1/(d(\nu^\star,C^\star,\rho^\star)+2)},$$

for any function f *satisfying Assumption 1 with parameters* $\nu^\star \le \nu_{\max}$ *and* $\rho^\star \le \rho_{\max}$.

5. General parallel optimization

The analysis of POO($\mathcal{A}$) proposed in Proposition 7 heavily relies on the fact that we control the *cumulative regret* of algorithm $\mathcal{A}$. POO indeed exploits this property when selecting $s^\star$ as the instance with largest empirical cumulative rewards. In this section, we propose a simple modification of POO($\mathcal{A}$) that allows using as base algorithms any hierarchical optimization algorithms that would only have *simple regret* guarantees.

The GPO($\mathcal{A}$) algorithm (**g**eneral **p**arallel **o**ptimization), whose pseudocode is shown in Algorithm 5, mostly needs to modify the model selection strategy of POO. There are two natural candidates: (i) Lepski's method which is a nested aggregation scheme (Lepski, 1992; Lepski and Spokoiny, 1997; Locatelli et al., 2017; Locatelli and Carpentier, 2018) that requires a single optimum, thus not directly applicable to our case, and (ii) a cross-validation scheme that we use and detail in the next. Given a total budget of n function evaluations, GPO($\mathcal{A}$) runs several instances of $\mathcal{A}$ in parallel with parameters chosen in the same grid as that used by POO, each using the same number of evaluations to output a recommendation $\widetilde{x}_i$. One half of the budget is then dedicated to estimating the function values at those points, and the one with the highest estimated value is kept.

Algorithm 5: General parallel optimization (GPO)

```
Input   : base algorithm $\mathcal{A}$, budget $n$, $\rho_{\max}$, $\nu_{\max}$, $K$
Compute $N = \lceil (1/2) D_{\max} \ln((n/2)/\ln(n/2)) \rceil$ the number of instances
for $i \leftarrow 1, \dots, N$ do
    $s \leftarrow \left(\nu_{\max}, {\rho_{\max}}^{2N/(2i+1)}\right)$
    Run $\mathcal{A}(s)$ for $\lfloor n/(2N) \rfloor$ time steps and output a recommendation $\widetilde{x}_s$
    Get $\lfloor n/(2N) \rfloor$ noisy evaluations of $f(\widetilde{x}_s)$ and compute their average $V[s]$
end
$s^\star \leftarrow \arg\max_s V[s]$
return $\widetilde{x}_{s^\star}$
```

In Theorem 8, we provide a general analysis of the GPO algorithm, showing that it attains an (order)-optimal simple regret without knowing the parameter triple $(\nu^\star, C^\star, \rho^\star)$ provided that its base algorithm does. As a consequence GPO(HCT) is an alternative to PCT with similar simple regret guarantees.

Theorem 8 *If for all* (ν, ρ) *the* $\mathcal{A}(\nu, \rho)$ *algorithm has its* simple regret *bounded as*

$$\mathbb{E}\left[S_n^{\mathcal{A}(\nu,\rho)}\right] \le \alpha C\left((\log n/n)^{1/(d(\nu,C,\rho)+2)}\right), \tag{2}$$

for any function f satisfying Assumption 1 with parameters (ν, ρ), then there exists a constant β that is independent of $\nu_{\max}$ and $\rho_{\max}$ such that

$$\mathbb{E}\left[S_n^{\texttt{GPO}(\mathcal{A})}\right] \le \beta D_{\max}(\nu_{\max}/\nu^\star)^{D_{\max}}\left((\log^2 n)/n\right)^{1/(d(\nu^\star,C^\star,\rho^\star)+2)},$$

for any function f satisfying Assumption 1 with parameters $\nu^\star \le \nu_{\max}$ and $\rho^\star \le \rho_{\max}$.

Proof We start by fixing some notation. Recall that N (that depends on n) is the number of instances run in parallel. For $j \in \{1, \dots, N\}$, we let $\widetilde{x}_j$ denote the point recommended by the instance $\mathcal{A}(\nu_{\max}, \rho_j)$ with $\rho_j = \rho_{\max}^{2N/(2j+1)}$. Let $(r_{i,j})_{1\le i\le n^+}$ be the i.i.d. evaluations of $f(\widetilde{x}_j)$ used during the validation phase, with $n^+ \triangleq \lfloor n/(2N)\rfloor$ and $\widehat{\mu}_{n^+,j} = \frac{1}{n^+}\sum_{i=1}^{n^+} r_{i,j}$ be the estimated value of $f(\widetilde{x}_j)$ computed by the algorithm. We let

$$\widehat{\jmath} = \arg\max_j\ \widehat{\mu}_{n^+,j} \quad\text{and}\quad \widetilde{\jmath} = \arg\max_j\ f(\widetilde{x}_j)$$

be the index of the empirical best and true best among the recommended point. We notice that for any j, $\{r_{i,j} - f(\widetilde{x}_j)\}_{i=1}^{n^+}$ is a bounded i.i.d. sequence with zero mean (conditionally to $\widetilde{x}_j$) thus using Hoeffding's inequality one can show that for all $\Delta > 0$,

$$\mathbb{P}\left[\left|\widehat{\mu}_{n^+,j} - f(\widetilde{x}_j)\right| > \Delta\right] \le 2\exp\left(-2n^+\Delta^2\right).$$

By integrating over $\Delta \in [0, 1]$, we get

$$\forall j \in \{1, \dots, N\},\ \mathbb{E}\left[\left|\widehat{\mu}_{n^+,j} - f(\widetilde{x}_j)\right|\right] \le \frac{\sqrt{\pi/2}}{\sqrt{n^+}}. \tag{3}$$

As in the analysis of `POO`, the instance $\overline{\jmath}$ defined as

$$\overline{\jmath} \triangleq \mathop{\arg\min}_{j\le N:\rho_j\ge\rho^\star} \left[d(\nu_{\max}, C^\star, \rho_j) - d(\nu^\star, C^\star, \rho^\star)\right]$$

shall play a crucial role. Indeed, inequality (2) is exactly what is needed in Appendix B.2 and Appendix B.3 of Grill et al. (2015) to control the simple regret of that instance in terms of $(\nu^\star, C^\star, \rho^\star)$. Following the exact same steps, we can show that for some constant α,

$$\mathbb{E}\left[S_{(n/2N)}^{\mathcal{A}(\nu_{\max},\rho_{\overline{\jmath}})}\right] \le \alpha D_{\max}(\nu_{\max}/\nu^\star)^{D_{\max}}\left((\log^2 n)/n\right)^{1/(d(\nu^\star,C^\star,\rho^\star)+2)}. \tag{4}$$

We now turn our attention to the simple regret of `GPO`$(\mathcal{A})$ after n function evaluations.

$$\mathbb{E}\left[S_n^{\texttt{GPO}}\right] = \mathbb{E}\left[f^\star - f(\widetilde{x}_{\widehat{\jmath}})\right] = \mathbb{E}[f^\star - f(\widetilde{x}_{\overline{\jmath}})] + \mathbb{E}\left[f(\widetilde{x}_{\overline{\jmath}}) - f(\widetilde{x}_{\widetilde{\jmath}})\right] + \mathbb{E}\left[f(\widetilde{x}_{\widetilde{\jmath}}) - f(\widetilde{x}_{\widehat{\jmath}})\right]. \tag{5}$$

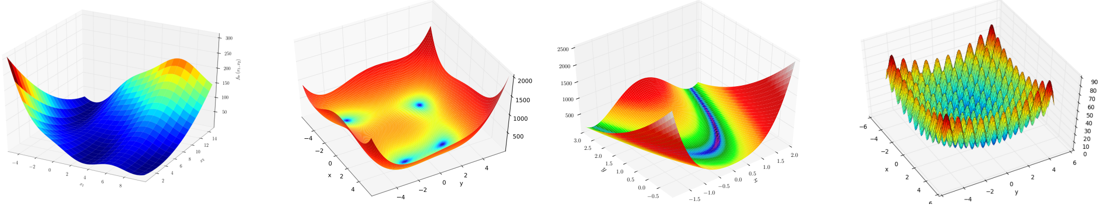
The first term in (5) is equal to the simple regret of the instance $\overline{\jmath}$ that uses n/N samples, which is upper bounded in (4). The second term in (5) is always negative by definition of $\widetilde{\jmath}$ and the third term can be rewritten as

$$\mathbb{E}\left[f(\widetilde{x}_{\widetilde{\jmath}}) - f(\widetilde{x}_{\widehat{\jmath}})\right] = \mathbb{E}\left[f(\widetilde{x}_{\widetilde{\jmath}}) - \widehat{\mu}_{n^+,\widetilde{\jmath}}\right] + \mathbb{E}\left[\widehat{\mu}_{n^+,\widetilde{\jmath}} - \widehat{\mu}_{n^+,\widehat{\jmath}}\right] + \mathbb{E}\left[\widehat{\mu}_{n^+,\widehat{\jmath}} - f(\widetilde{x}_{\widehat{\jmath}})\right]. \tag{6}$$

where the first and the third term of (6) are both upper bounded by $(\sqrt{\pi/2})/\sqrt{n^+}$ using (3), and the second term is always negative by definition of $\widehat{\jmath}$. Putting things together yields

$$\mathbb{E}\left[S_n^{\texttt{GPO}}\right] \le \alpha D_{\max}(\nu_{\max}/\nu^\star)^{D_{\max}}\left((\log^2 n)/n\right)^{1/(d(\nu^\star,C^\star,\rho^\star)+2)} + O\left(\frac{\sqrt{N}}{\sqrt{n}}\right).$$

The conclusion follows by observing that the second term in the right-hand side is negligible with respect to the first. ■



(a) Branin (b) Himmelblau (c) Rosenbrock (d) Rastrigin in 2D

Figure 1: Benchmark functions[5]

6. Experimental illustrations

We run experiments on several test functions comparing the original POO along with several instances of HOO and our new instantiation PCT along with HCT instances of different ρ values. In these experiments, we set $\rho_{\max} = 0.9$, and we add Gaussian noise to the function evaluations with a relatively small variance ($\sigma = 0.1$).

Artificial landscapes We test the algorithms on some functions from the *artificial landscapes*, including (i) two functions with many local minima: Himmelblau function and Rastrigin function, (ii) one valley-shaped function: Rosenbrock function, and (iii) Branin function (see Figure 1). Note that the Rastrigin function shown is its 2D version. In our experiments, we use a Rastrigin function in 5D.

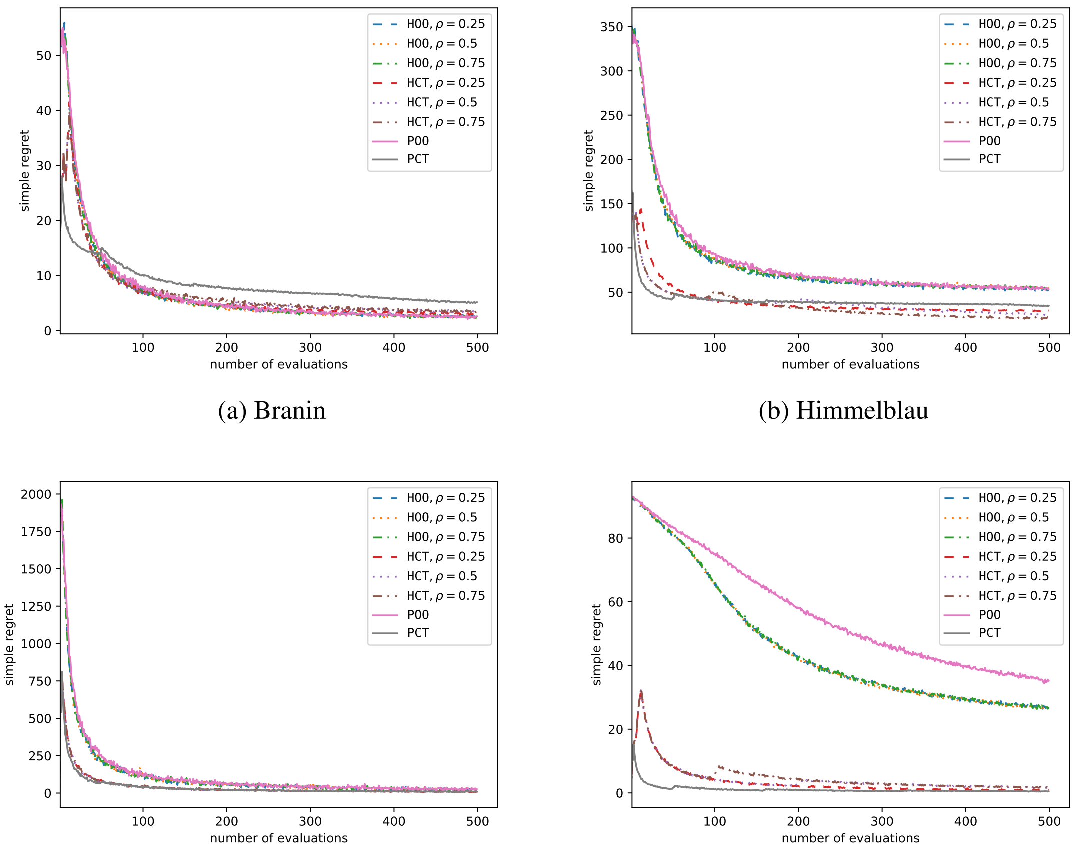
In Figure 2, we plot the simple regret of the algorithms as a function of the number of evaluations. All the results are averaged over 5000 runs and we plot the simple regret after 500 function evaluations. Each instance of HOO or HCT would recommend a point picked uniformly at random among those evaluated so that we have the same recommendation strategy as POO and PCT.

The first observation is that PCT does match the performance of some single HCT instances as expected. We also notice that PCT has comparable performance w.r.t. POO in these plots, which justifies the choice of using HCT as a subroutine for the POO meta-algorithm.

7. Discussion

We studied PCT, a new instantiation of POO on top of HCT. We proved that HCT is a plausible subroutine for POO by adapting the analysis of HCT under a new assumption w.r.t. a fixed partitioning. We also proposed GPO, a general framework for making any hierarchical bandit algorithm that only has a simple regret guarantee adaptive to unknown smoothness. However, whether it is possible to weaken the assumptions of HOO in the same way as HCT while keeping similar regret guarantees remains open.

[5] Source: https://en.wikipedia.org/wiki/Test_functions_for_optimization



(a) Branin

(b) Himmelblau

(c) Rosenbrock

(d) Rastrigin in 5D

Figure 2: Simple regret of POO and PCT run for different ρ values.

Acknowledgements

The research presented was supported by European CHIST-ERA project DELTA, French Ministry of Higher Education and Research, Nord-Pas-de-Calais Regional Council, a doctoral grant of Ecole Normale Supérieure in Rennes, Inria and Otto-von-Guericke-Universität Magdeburg associated-team north-European project Allocate, and French National Research Agency projects ExTra-Learn (n.ANR-14-CE24-0010-01), BADASS (n.ANR-16-CE40-0002), and BoB (n.ANR-16-CE23-0003).

References

Rajeev Agrawal. The continuum-armed bandit problem. *SIAM Journal on Control and Optimization*, 33:1926–1951, 1995.

Peter Auer, Ronald Ortner, and Csaba Szepesvári. Improved rates for the stochastic continuum-armed bandit problem. In *Conference on Learning Theory*, 2007.

Mohammad Gheshlaghi Azar, Alessandro Lazaric, and Emma Brunskill. Online stochastic optimization under correlated bandit feedback. In *International Conference on Machine Learning*, 2014.

Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357–367, 1967.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *Algorithmic Learning Theory*, 2009.

Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. X-armed bandits. *Journal of Machine Learning Research*, 12:1587–1627, 2011.

Adam D. Bull. Adaptive-treed bandits. *Bernoulli*, 21(4):2289–2307, 2015.

Eric W Cope. Regret and convergence bounds for immediate-reward reinforcement learning with continuous action spaces. *IEEE Transactions on Automatic Control*, 54(6):1243–1253, 2009.

Jean-Bastien Grill, Michal Valko, and Rémi Munos. Black-box optimization of noisy functions with unknown smoothness. In *Neural Information Processing Systems*, 2015.

Jean-Bastien Grill, Michal Valko, and Rémi Munos. Blazing the trails before beating the path: Sample-efficient Monte-Carlo planning. In *Neural Information Processing Systems*, 2016.

Kevin Jamieson and Ameet Talwalkar. Non-stochastic best arm identification and hyperparameter optimization. In *International Conference on Artificial Intelligence and Statistics*, 2016.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-armed bandit problems in metric spaces. In *Symposium on Theory Of Computing*, 2008.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Bandits and experts in metric spaces. *Journal of ACM*, 2015.

Robert D Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Neural Information Processing Systems*, 2005.

Levente Kocsis and Csaba Szepesvári. Bandit-based Monte-Carlo planning. In *European Conference on Machine Learning*, 2006.

O. V. Lepski and V. G. Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *The Annals of Statistics*, 25(6):2512–2546, 1997.

Oleg V. Lepski. Asymptotically minimax adaptive estimation. I: Upper bounds. optimally adaptive estimates. *Theory of Probability & Its Applications*, 36(4):682–697, 1992.

Lisha Li, Kevin Jamieson, Giulia DeSalvo, and Afshin Rostamizadeh Ameet Talwalkar. Hyperband: Bandit-based configuration evaluation for hyperparameter optimization. In *International Conference on Learning Representations*, 2017.

Andrea Locatelli and Alexandra Carpentier. Adaptivity to Smoothness in X-armed bandits. In *Conference on Learning Theory*, 2018.

Andrea Locatelli, Alexandra Carpentier, and Samory Kpotufe. Adaptivity to noise parameters in nonparametric active learning. In *Conference on Learning Theory*, 2017.

Rémi Munos. Optimistic optimization of deterministic functions without the knowledge of its smoothness. In *Neural Information Processing Systems*, 2011.

Rémi Munos. From bandits to Monte-Carlo tree search: The optimistic principle applied to optimization and planning. *Foundations and Trends in Machine Learning*, 7(1):1–130, 2014.

Philippe Preux, Rémi Munos, and Michal Valko. Bandits attack function optimization. In *Congress on Evolutionary Computation*, 2014.

Spyridon Samothrakis, Diego Perez, and Simon Lucas. Training gradient boosting machines using curve-fitting and information-theoretic features for causal direction detection. In *NIPS Workshop on Causality*, 2013.

Aleksandrs Slivkins. Multi-armed bandits on implicit metric spaces. In *Neural Information Processing Systems*, 2011.

Michal Valko, Alexandra Carpentier, and Rémi Munos. Stochastic simultaneous optimistic optimization. In *International Conference on Machine Learning*, 2013.

Appendix A. Detailed regret analysis for HCT under Assumption 1

A.1. Preliminaries

We first fix some constants and introduce some additional notation that are needed for the proof of Theorem 2.

- $c_1 \triangleq (\rho/(3\nu))^{1/8}$, $c \triangleq 2\sqrt{1/(1-\rho)}$
- $\forall 1 \le h \le H(t)$ and $t > 0$, $\mathcal{I}_h(t)$ denotes the set of all nodes created by HCT at level h up to step t
- $\forall 1 \le h \le H(t)$ and $t > 0$, $\mathcal{I}_h^+(t)$ denotes the subset of $\mathcal{I}_h(t)$ which contains only the internal nodes
- At each step t, (h_t, i_t) denotes the node selected by the algorithm.
- $\mathcal{C}_{h,i} \triangleq \{t = 1, \cdots, n : (h_t, i_t) = (h, i)\}$
- $\mathcal{C}_{h,i}^+ \triangleq \bigcup\limits_{j \in \{0,\ldots,K-1\}} \mathcal{C}_{h+1,Ki-j}$
- $\bar{t}_{h,i} \triangleq \max_{t \in \mathcal{C}_{h,i}} t$ denotes the last time (h,i) has been selected
- $\widetilde{t}_{h,i} \triangleq \max_{t \in \mathcal{C}_{h,i}^+} t$ denotes the last time when one of its children has been selected
- $t_{h,i} \triangleq \min\{t : T_{h,i}(t) \ge \tau_h(t)\}$ is the time when (h,i) is expanded
- For any t, let $y_t \triangleq (r_t, x_t)$ be a random variable, we define the filtration $\mathcal{F}_t$ as a σ-algebra generated by $(y_1, \ldots, y_t)$.

Another important notion in HCT is the threshold τ_h on the number of pulls needed before a node at level h can be expanded. The threshold τ_h is chosen such that the two confidence terms in $U_{h,i}$ are roughly equivalent, that is,

$$\nu\rho^h \simeq c\sqrt{\frac{\log(1/\widetilde{\delta}(t^+))}{\tau_h(t)}}.$$

Therefore, we choose

$$\tau_h(t) \triangleq \left\lceil \frac{c^2 \log(1/\widetilde{\delta}(t^+))}{\nu^2} \rho^{-2h} \right\rceil.$$

Since t^+ is defined as $2^{\lceil \log(t) \rceil}$, we have $t \le t^+ \le 2t$. In addition, $\log$ is an increasing function, thus we have

$$\frac{c^2}{\nu^2}\rho^{-2h} \le \frac{c^2 \log(1/\widetilde{\delta}(t))}{\nu^2}\rho^{-2h} \le \tau_h(t) \le \frac{c^2 \log(2/\widetilde{\delta}(t))}{\nu^2}\rho^{-2h}, \tag{7}$$

where the first inequality follows from the fact that $0 < \widetilde{\delta}(t) \le 1/2$. We begin our analysis by bounding the maximum depth of the trees constructed by HCT.

A.2. Maximum depth of the tree (proof of Lemma 4)

Lemma 4 *The depth of the covering tree produced by* `HCT` *after* n function evaluations satisfies

$$H(n) \leq H_{\max}(n) \triangleq \left\lceil \frac{1}{2(1-\rho)} \log\left(\frac{n\nu^2}{c^2\rho^2}\right) \right\rceil.$$

Proof The deepest tree that can be constructed by `HCT` is a linear one, where at each level one unique node is expanded. In such case, $|\mathcal{I}_h^+(n)| = 1$ and $|\mathcal{I}_h(n)| = K$ for all $h < H(n)$. Therefore, we have

$$\begin{aligned}
n &= \sum_{h=0}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} T_{h,i}(n) \\
&\geq \sum_{h=0}^{H(n)-1} \sum_{i \in \mathcal{I}_h^+(n)} T_{h,i}(n) \\
&\geq \sum_{h=0}^{H(n)-1} \sum_{i \in \mathcal{I}_h^+(n)} T_{h,i}(t_{h,i}) \\
&\geq \sum_{h=0}^{H(n)-1} \sum_{i \in \mathcal{I}_h^+(n)} \tau_h(t_{h,i}) && \text{definition of } t_{h,i} \\
&\geq \sum_{h=0}^{H(n)-1} \frac{c^2}{\nu^2}\rho^{-2h} && \text{ineq. (7)} \\
&\geq \frac{(c\rho)^2}{\nu^2}\rho^{-2H(n)} H(n) && \text{since } h \leq H(n)-1 \\
&\geq \frac{(c\rho)^2}{\nu^2}\rho^{-2H(n)}.
\end{aligned}$$

By solving this expression, we obtain

$$\begin{aligned}
H(n) &\leq \frac{1}{2}\log\left(\frac{n\nu^2}{c^2\rho^2}\right) / \log(1/\rho) \\
&\leq \frac{1}{2(1-\rho)} \log\left(\frac{n\nu^2}{c^2\rho^2}\right) && \text{follows from } \log(1/\rho) \geq 1-\rho \\
&\leq \left\lceil \frac{1}{2(1-\rho)} \log\left(\frac{n\nu^2}{c^2\rho^2}\right) \right\rceil \\
&\triangleq H_{\max}(n).
\end{aligned}$$

■

A.3. High-probability event

In Section 3.2, we described the favorable event ξ_t. We now define it precisely. We first define a set $\mathcal{L}_t$ that contains all possible nodes in trees of maximum depth $H_{\max}(t)$,

$$\mathcal{L}_t \triangleq \bigcup_{\mathcal{T}:\mathrm{depth}(\mathcal{T})\leq H_{\max}(t)} \mathrm{Nodes}(\mathcal{T})$$

and we recall the definition of the favorable event

$$\xi_t \triangleq \left\{\forall (h,i)\in\mathcal{L}_t, |\widehat{\mu}_{h,i}(t)-\mu_{h,i}|\leq c\sqrt{\frac{\log(1/\widetilde{\delta}(t))}{T_{h,i}(t)}}\right\}.$$

Next, we prove that our favorable event holds with high probability.

Lemma 9 *With c_1 and c defined in Section A.1, for any fixed round t,*

$$\mathbb{P}[\xi_t]\geq 1-\frac{4\delta}{3t^6}.$$

Proof Letting $\widehat{\mu}_{h,i,s}$ denote the empirical mean reward of the first s noisy evaluations of f in $x_{h,i}$, we upper-bound the probability of the complementary event ξ_t^c as

$$\begin{aligned}
\mathbb{P}[\xi_t^c] &\leq \sum_{(h,i)\in\mathcal{L}_t}\sum_{s=1}^{t}\mathbb{P}\left[|\widehat{\mu}_{h,i,s}-\mu_{h,i}|\geq c\sqrt{\frac{\log(1/\widetilde{\delta}(t))}{s}}\right] && \text{union bound}\\
&\leq \sum_{(h,i)\in\mathcal{L}_t}\sum_{s=1}^{t} 2\exp\Big(-2c^2\log(1/\widetilde{\delta}(t))\Big) && \text{Chernoff-Hoeffding inequality}\\
&= 2\exp\Big(-2c^2\log(1/\widetilde{\delta}(t))\Big)t|\mathcal{L}_t| \\
&= 2(\widetilde{\delta}(t))^{2c^2}t|\mathcal{L}_t| \\
&\leq 2(\widetilde{\delta}(t))^{2c^2}t2^{H_{max}(t)+1} \\
&= 2(\widetilde{\delta}(t))^{2c^2}t2^{\left\lceil\frac{1}{2(1-\rho)}\log\left(\frac{n\nu^2}{c^2\rho^2}\right)\right\rceil+1} && \text{Lemma 4}\\
&\leq 8t(\widetilde{\delta}(t))^{2c^2}\left(\frac{t\nu^2}{c^2\rho^2}\right)^{\frac{1}{2(1-\rho)}} \\
&\leq 8t\left(\frac{\delta}{t}(\rho/(3\nu))^{1/8})^{\frac{8}{1-\rho}}\right)\left(\frac{t\nu^2(1-\rho)}{4\rho^2}\right)^{\frac{1}{2(1-\rho)}} && \text{plugging in values of } c \text{ and } c_1\\
&= 8t\left(\frac{\delta}{t}\right)^{\frac{8}{1-\rho}}\left(\frac{\rho}{3\nu}\right)^{\frac{1}{1-\rho}}t^{\frac{1}{2(1-\rho)}}\left(\frac{\nu\sqrt{1-\rho}}{2\rho}\right)^{\frac{1}{1-\rho}} \\
&\leq \frac{4}{3}\delta t^{\frac{-2\rho-13}{2(1-\rho)}} \\
&\leq \frac{4\delta}{3t^6}.
\end{aligned}$$

■

A.4. Failing confidence bound

We decompose the regret of HCT into two terms depending on whether ξ_t holds. Let us define $\Delta_t \triangleq f^\star - r_t$. Then, we decompose the regret as

$$R_n^{\texttt{HCT}} = \sum_{t=1}^{n} \Delta_t = \sum_{t=1}^{n} \Delta_t \mathbf{1}_{\xi_t} + \sum_{t=1}^{n} \Delta_t \mathbf{1}_{\xi_t^c} = R_n^{\xi} + R_n^{\xi^c}.$$

The failing confidence term $R_n^{\xi^c}$ is bounded by the following lemma.

Lemma 10 *With c_1 and c defined in Section A.1, when the favorable event does not hold, the regret of HCT is with probability $1 - \delta/(5n^2)$ bounded as*

$$R_n^{\xi^c} \leq \sqrt{n}.$$

Proof We split the term into rounds from 1 to $\sqrt{n}$ and the rest,

$$R_n^{\xi^c} = \sum_{t=1}^{n} \Delta_t \mathbf{1}_{\xi_t^c} = \sum_{t=1}^{\sqrt{n}} \Delta_t \mathbf{1}_{\xi_t^c} + \sum_{t=\sqrt{n}+1}^{n} \Delta_t \mathbf{1}_{\xi_t^c}.$$

The first term can be bounded trivially by $\sqrt{n}$ since $|\Delta_t| \leq 1$. Next, we show that the probability that the second term is non zero is bounded by $\delta/(5n^2)$.

$$\begin{aligned}
\mathbb{P}\left[\sum_{t=\sqrt{n}+1}^{n} \Delta_t \mathbf{1}_{\xi_t^c} > 0\right] &= \mathbb{P}\left[\bigcup_{t=\sqrt{n}+1}^{n} \xi_t^c\right] \\
&\leq \sum_{t=\sqrt{n}+1}^{n} \mathbb{P}[\xi_t^c] && \text{union bound} \\
&\leq \sum_{t=\sqrt{n}+1}^{n} \frac{\delta}{t^6} && \text{Lemma 9} \\
&\leq \int_{\sqrt{n}}^{\infty} \frac{\delta}{t^6}\,\mathrm{d}t \\
&= \frac{\delta}{5n^{5/2}} \\
&\leq \frac{\delta}{5n^2}.
\end{aligned}$$

■

A.5. Proof of Theorem 2

Theorem 2 *Assume that function f satisfies Assumption 1. Then, setting $\delta \triangleq 1/n$, the cumulative regret of $\texttt{HCT}(\nu, \rho)$ after n function evaluations is upper bounded as*

$$\mathbb{E}[R_n^{\texttt{HCT}(\nu,\rho)}] \leq \alpha C (\log n)^{1/(d(\nu,C,\rho)+2)} n^{(d(\nu,C,\rho)+1)/(d(\nu,C,\rho)+2)},$$

where α is a numerical constant and C is the constant associated to $d(\nu, C, \rho)$.

For the sake of simplicity, we denote $d(\nu, C, \rho)$ as d in the rest of this section. We study the regret under events $\{\xi_t\}_t$ and prove that

$$R_n^{\texttt{HCT}(\nu,\rho)} \leq 2\sqrt{2n\log(\frac{4n^2}{\delta})} + 3\left(\frac{2^{3d+7}\nu^d KC\rho^d}{(1-\rho)^2}\right)^{\frac{1}{d+2}}\left(\log\left(\frac{2n}{\delta}\sqrt[8]{\frac{3\nu}{\rho}}\right)\right)^{\frac{1}{d+2}} n^{\frac{d+1}{d+2}}$$

holds with probability $1-\delta$. We decompose the proof into 3 steps.

Step 1: Decomposition of the regret. We start by further decomposing the instantaneous regret into two terms,

$$\Delta_t = f^\star - r_t = f^\star - f(x_{h_t,i_t}) + f(x_{h_t,i_t}) - r_t = \Delta_{h_t,i_t} + \widehat{\Delta}_t.$$

The regret of HCT when confidence intervals hold can thus be rewritten as

$$R_n^\xi = \sum_{t=1}^n \Delta_{h_t,i_t}\mathbf{1}_{\xi_t} + \sum_{t=1}^n \widehat{\Delta}_t\mathbf{1}_{\xi_t} \leq \sum_{t=1}^n \Delta_{h_t,i_t}\mathbf{1}_{\xi_t} + \sum_{t=1}^n \widehat{\Delta}_t = \widetilde{R}_n^\xi + \widehat{R}_n^\xi. \tag{8}$$

We notice that $\{\widehat{\Delta}_t\}_{t=1}^n$ is a bounded martingale difference sequence since $\mathbb{E}\left[\widehat{\Delta}_t|\mathcal{F}_{t-1}\right] = 0$ and $|\widehat{\Delta}_t| \leq 1$. Thus, we apply the Azuma's inequality on this sequence and obtain

$$\widehat{R}_n^\xi \leq \sqrt{2n\log\left(\frac{4n^2}{\delta}\right)} \tag{9}$$

with probability $1-\delta/(4n^2)$.

Step 2: Preliminary bound on the regret of selected nodes and their parents. Now we proceed with the bound of the first term $\widetilde{R}_n^\xi$. Recall that P_t is the optimistic path traversed by HCT at round t. Let $(h', i') \in P_t$ and (h'', i'') be the node which immediately follows (h', i') in P_t. By definition of B-values and U-values, we have

$$B_{h',i'}(t) \leq \max_{j\in\{0,\ldots,K-1\}}\left\{B_{h'+1,Ki'-j}(t)\right\} = B_{h'',i''}(t), \tag{10}$$

where the last equality follows from the fact that the subroutine OptTraverse selects the node with the largest B-value. By iterating the previous inequality along the path P_t until the selected node (h_t, i_t) and its parent (h_t^p, i_t^p), we obtain

$$\forall (h', i') \in P_t, B_{h',i'}(t) \leq B_{h_t,i_t}(t) \leq U_{h_t,i_t}(t),$$

$$\forall (h', i') \in P_t \setminus \{(h_t, i_t)\}, B_{h',i'}(t) \leq B_{h_t^p,i_t^p}(t) \leq U_{h_t^p,i_t^p}(t).$$

Since the root, which is an optimal node, is in P_t, there exists at least one optimal node $(h^\star, i^\star)$ in path P_t. As a result, we have

$$B_{h^\star,i^\star}(t) \leq U_{h_t,i_t}(t), \tag{11}$$

$$B_{h^\star,i^\star}(t) \leq U_{h_t^p,i_t^p}(t). \tag{12}$$

We now expand (11) on both sides under ξ_t. First, we have

$$U_{h_t,i_t}(t) \triangleq \widehat{\mu}_{h_t,i_t}(t) + \nu\rho^{h_t} + c\sqrt{\frac{\log(1/\widetilde{\delta}(t^+))}{T_{h_t,i_t}(t)}} \leq f(x_{h_t,i_t}) + \nu\rho^{h_t} + 2c\sqrt{\frac{\log(1/\widetilde{\delta}(t^+))}{T_{h_t,i_t}(t)}} \tag{13}$$

and the same holds for the parent of the selected node,

$$U_{h_t^p,i_t^p}(t) \leq f(x_{h_t^p,i_t^p}) + \nu\rho^{h_t^p} + 2c\sqrt{\frac{\log(1/\widetilde{\delta}(t^+))}{T_{h_t^p,i_t^p}(t)}}.$$

By Lemma 5, we know that $U_{h^\star,i^\star}(t)$ is a valid upper bound on $f^\star$. If an optimal node $(h^\star, i^\star)$ is a leaf, then $B_{h^\star,i^\star}(t) = U_{h^\star,i^\star}(t)$ is also a valid upper bound on $f^\star$. Otherwise, there always exists a leaf which contains the maximum for which $(h^\star, i^\star)$ is its ancestor. Now, if we propagate the bound backward from this leaf to $(h^\star, i^\star)$ through (10), we have that $B_{h^\star,i^\star}(t)$ is still a valid upper bound on $f^\star$. Thus for any optimal node $(h^\star, i^\star)$, at round t under ξ_t, we have

$$B_{h^\star,i^\star}(t) \geq f^\star. \tag{14}$$

We combine (14) with (11) and (13) to obtain

$$\Delta_{h_t,i_t} \triangleq f^\star - f(x_{h_t,i_t}) \leq \nu\rho^{h_t} + 2c\sqrt{\frac{\log(1/\widetilde{\delta}(t^+))}{T_{h_t,i_t}(t)}}.$$

The same result holds for its parent,

$$\Delta_{h_t^p,i_t^p} \triangleq f^\star - f(x_{h_t^p,i_t^p}) \leq \nu\rho^{h_t^p} + 2c\sqrt{\frac{\log(1/\widetilde{\delta}(t^+))}{T_{h_t^p,i_t^p}(t)}}.$$

We now refine the two above expressions. The subroutine `OptTraverse` tells us that `HCT` only selects a node when $T_{h,i}(t) < \tau_h(t)$. Therefore, by definition of $\tau_{h_t}(t)$, we have

$$\Delta_{h_t,i_t} \leq 3c\sqrt{\frac{\log(2/\widetilde{\delta}(t))}{T_{h_t,i_t}(t)}}. \tag{15}$$

On the other hand, `OptTraverse` tells us that $T_{h_t^p,i_t^p}(t) \geq \tau_{h_t^p}(t)$, thus

$$\Delta_{h_t^p,i_t^p} \leq 3\nu\rho^{h_t^p},$$

which means that every selected node has a parent which is $(3\nu\rho^{h_t-1})$-optimal.

Step 3: Bound on the cumulative regret. We return to term $\widetilde{R}_n^\xi$ and split it into different depths. Let $1 \le \overline{H} \le H(n)$ be a constant that we fix later. We have

$$\begin{aligned}
\widetilde{R}_n^\xi &\triangleq \sum_{t=1}^{n} \Delta_{h_t,i_t} \mathbf{1}_{\xi_t} \\
&\le \sum_{h=0}^{H(n)} \sum_{i\in\mathcal{I}_h(n)} \sum_{t=1}^{n} \Delta_{h,i} \mathbf{1}_{(h_t,i_t)=(h,i)} \mathbf{1}_{\xi_t} \\
&\le \sum_{h=0}^{H(n)} \sum_{i\in\mathcal{I}_h(n)} \sum_{t=1}^{n} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(t))}{T_{h,i}(t)}} \mathbf{1}_{(h_t,i_t)=(h,i)} \\
&= \sum_{h=0}^{\overline{H}} \sum_{i\in\mathcal{I}_h(n)} \sum_{t=1}^{n} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(t))}{T_{h,i}(t)}} \mathbf{1}_{(h_t,i_t)=(h,i)} \\
&\quad + \sum_{h=\overline{H}+1}^{H(n)} \sum_{i\in\mathcal{I}_h(n)} \sum_{t=1}^{n} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(t))}{T_{h,i}(t)}} \mathbf{1}_{(h_t,i_t)=(h,i)} \\
&\le \sum_{h=0}^{\overline{H}} \sum_{i\in\mathcal{I}_h(n)} \sum_{s=1}^{\tau_h(\bar{t}_{h,i})} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(\bar{t}_{h,i}))}{s}} \\
&\quad + \sum_{h=\overline{H}+1}^{H(n)} \sum_{i\in\mathcal{I}_h(n)} \sum_{s=1}^{T_{h,i}(n)} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(\bar{t}_{h,i}))}{s}} \\
&\le \sum_{h=0}^{\overline{H}} \sum_{i\in\mathcal{I}_h(n)} \int_1^{\tau_h(\bar{t}_{h,i})} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(\bar{t}_{h,i}))}{s}}\,\mathrm{d}s \\
&\quad + \sum_{h=\overline{H}+1}^{H(n)} \sum_{i\in\mathcal{I}_h(n)} \int_1^{T_{h,i}(n)} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(\bar{t}_{h,i}))}{s}}\,\mathrm{d}s \\
&\le \sum_{h=0}^{\overline{H}} \sum_{i\in\mathcal{I}_h(n)} 6c\sqrt{\tau_h(\bar{t}_{h,i})\log(2/\widetilde{\delta}(\bar{t}_{h,i}))} \\
&\quad + \sum_{h=\overline{H}+1}^{H(n)} \sum_{i\in\mathcal{I}_h(n)} 6c\sqrt{T_{h,i}(n)\log(2/\widetilde{\delta}(\bar{t}_{h,i}))} \\
&= 6c\left(\underbrace{\sum_{h=0}^{\overline{H}} \sum_{i\in\mathcal{I}_h(n)} \sqrt{\tau_h(\bar{t}_{h,i})\log(2/\widetilde{\delta}(\bar{t}_{h,i}))}}_{\text{(a)}} + \underbrace{\sum_{h=\overline{H}+1}^{H(n)} \sum_{i\in\mathcal{I}_h(n)} \sqrt{T_{h,i}(n)\log(2/\widetilde{\delta}(\bar{t}_{h,i}))}}_{\text{(b)}}\right).
\end{aligned}$$

We bound separately the terms (a) and (b). Since $\bar{t}_{h,i} \leq n$, we have

$$\text{(a)} \leq \sum_{h=0}^{\overline{H}} \sum_{i \in \mathcal{I}_h(n)} \sqrt{\tau_h(n) \log(2/\widetilde{\delta}(n))} \leq \sum_{h=0}^{\overline{H}} |\mathcal{I}_h(n)| \sqrt{\tau_h(n) \log(2/\widetilde{\delta}(n))}.$$

Notice that the covering tree is K-ary and therefore $|\mathcal{I}_h(n)| \leq K|\mathcal{I}_{h-1}(n)|$. Recall that `HCT` only selects a node (h_t, i_t) when its parent is $3\nu\rho^{h_t-1}$-optimal. Therefore, by definition of the near-optimality dimension,

$$|\mathcal{I}_h(n)| \leq |K\mathcal{I}_{h-1}(n)| \leq KC\rho^{-d(h-1)},$$

where d is the near-optimality dimension. As a result, for term (a), we obtain that

$$\begin{aligned}
\text{(a)} &\leq \sum_{h=0}^{\overline{H}} KC\rho^{-d(h-1)} \sqrt{\tau_h(n) \log(2/\widetilde{\delta}(n))} \\
&= \sum_{h=0}^{\overline{H}} KC\rho^{-d(h-1)} \sqrt{\frac{c^2 \log(2/\widetilde{\delta}(n))}{\nu^2} \rho^{-2h} \log(2/\widetilde{\delta}(n))} && \text{ineq. (7)} \\
&= KC\rho^d \frac{c \log(2/\widetilde{\delta}(n))}{\nu} \sum_{h=0}^{\overline{H}} \rho^{-h(d+1)}.
\end{aligned}$$

Consequently, we bound (a) as

$$\text{(a)} \leq KC\rho^d \frac{c \log\left(2/\widetilde{\delta}(n)\right)}{\nu} \frac{\rho^{-\overline{H}(d+1)}}{1-\rho}. \tag{16}$$

We proceed to bound the second term (b). By the Cauchy-Schwarz inequality,

$$\begin{aligned}
\text{(b)} &\leq \sqrt{\sum_{h=\overline{H}+1}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} \log\left(2/\widetilde{\delta}(\bar{t}_{h,i})\right)} \sqrt{\sum_{h=\overline{H}+1}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} T_{h,i}(n)} \\
&\leq \sqrt{n \sum_{h=\overline{H}+1}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} \log\left(2/\widetilde{\delta}(\bar{t}_{h,i})\right)},
\end{aligned}$$

where we trivially bound the second square-root factor by the total number of pulls. Now consider the first square-root factor. Recall that the `HCT` algorithm only selects a node when $T_{h,i}(t) \geq \tau_h(t)$

for its parent. We therefore have $T_{h,i}(\widetilde{t}_{h,i}) \geq \tau_h(\widetilde{t}_{h,i})$ and the following sequence of inequalities,

$$\begin{aligned}
n &= \sum_{h=0}^{H(n)} \sum_{i\in\mathcal{I}_h(n)} T_{h,i}(n) \\
&\geq \sum_{h=0}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} T_{h,i}(n) \\
&\geq \sum_{h=0}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} T_{h,i}(\widetilde{t}_{h,i}) \\
&\geq \sum_{h=0}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \tau_h(\widetilde{t}_{h,i}) \\
&\geq \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \tau_h(\widetilde{t}_{h,i}) \\
&= \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \frac{c^2\log(1/\widetilde{\delta}(\widetilde{t}^+_{h,i})))}{\nu^2}\rho^{-2h} \\
&\geq \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \frac{c^2\log(1/\widetilde{\delta}(\widetilde{t}^+_{h,i})))}{\nu^2}\rho^{-2\overline{H}} \\
&= \frac{c^2\rho^{-2\overline{H}}}{\nu^2} \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \log(1/\widetilde{\delta}(\widetilde{t}^+_{h,i}))) \\
&= \frac{c^2\rho^{-2\overline{H}}}{\nu^2} \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \log(1/\widetilde{\delta}(\big[\max(\bar{t}_{h+1,2i-1},\bar{t}_{h+1,2i})\big]^+)) \\
&= \frac{c^2\rho^{-2\overline{H}}}{\nu^2} \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \log(1/\widetilde{\delta}(\max(\bar{t}^+_{h+1,2i-1},\bar{t}^+_{h+1,2i}))) \\
&= \frac{c^2\rho^{-2\overline{H}}}{\nu^2} \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \max(\log(1/\widetilde{\delta}(\bar{t}^+_{h+1,2i-1})),\log(1/\widetilde{\delta}(\bar{t}^+_{h+1,2i}))) \\
&\geq \frac{c^2\rho^{-2\overline{H}}}{\nu^2} \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \frac{\log(1/\widetilde{\delta}(\bar{t}^+_{h+1,2i-1}))+\log(1/\widetilde{\delta}(\bar{t}^+_{h+1,2i}))}{2} \\
&= \frac{c^2\rho^{-2\overline{H}}}{\nu^2} \sum_{h=\overline{H}+1}^{H(n)} \sum_{i\in\mathcal{I}_{h-1}^+(n)} \frac{\log(1/\widetilde{\delta}(\bar{t}^+_{h,2i-1}))+\log(1/\widetilde{\delta}(\bar{t}^+_{h,2i}))}{2} && \text{change of variables} \\
&= \frac{c^2\rho^{-2\overline{H}}}{2\nu^2} \sum_{h=\overline{H}+1}^{H(n)} \sum_{i\in\mathcal{I}_h^+(n)} \log(1/\widetilde{\delta}(\bar{t}^+_{h,i})).
\end{aligned}$$

In the above derivation, the last equality relies on the fact that for any $h > 0$, $\mathcal{I}_h^+(n)$ covers all the internal nodes at level h and therefore its children cover $\mathcal{I}_{h+1}(n)$. We thus obtain

$$\sum_{h=\overline{H}+1}^{H(n)} \sum_{i \in \mathcal{I}_h^+(n)} \log(1/\widetilde{\delta}(\overline{t}_{h,i}^+)) \leq \frac{2\nu^2 \rho^{2\overline{H}} n}{c^2}. \tag{17}$$

On the other hand, we have

$$\begin{aligned}
\text{(b)} &\leq \sqrt{n \sum_{h=\overline{H}+1}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} \log(2/\widetilde{\delta}(\overline{t}_{h,i}))} \leq \sqrt{n \sum_{h=\overline{H}+1}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} 2\log(1/\widetilde{\delta}(\overline{t}_{h,i}))} \\
&\leq \sqrt{n \sum_{h=\overline{H}+1}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} 2\log(1/\widetilde{\delta}(\overline{t}_{h,i}^+))}
\end{aligned}$$

since $\overline{t}_{h,i} \leq \overline{t}_{h,i}^+$. By plugging (17) into above expression, we get

$$\text{(b)} \leq \frac{2\nu \rho^{\overline{H}} n}{c}. \tag{18}$$

Now if we combine (18) with (16), we bound $\widetilde{R}_n^\xi$ as

$$\widetilde{R}_n^\xi \leq 6\nu \left[KC\rho^d \frac{c^2 \log(2/\widetilde{\delta}(n))}{\nu^2} \frac{\rho^{-\overline{H}(d+1)}}{1-\rho} + 2\rho^{\overline{H}} n \right]. \tag{19}$$

We choose $\overline{H}$ to minimize the above bound by equalizing the two terms in the sum and obtain

$$\rho^{\overline{H}} = \left(\frac{KC\rho^d c^2 \log(2/\widetilde{\delta}(n))}{2n(1-\rho)\nu^2} \right)^{\frac{1}{d+2}}, \tag{20}$$

which after being plugged into (19) gives

$$\widetilde{R}_n^\xi \leq 24\nu \left(\frac{KC\rho^d c^2 \log(2/\widetilde{\delta}(n))}{2(1-\rho)\nu^2} \right)^{\frac{1}{d+2}} n^{\frac{d+1}{d+2}}. \tag{21}$$

Finally, combining (21), (9), and Lemma 10, we obtain

$$\begin{aligned}
R_n^{\mathtt{HCT}} &\leq \sqrt{n} + \sqrt{2n \log(\frac{4n^2}{\delta})} + 24\nu \left(\frac{2KC\rho^d}{(1-\rho)^2 \nu^2} \right)^{\frac{1}{d+2}} \left(\log\left(\frac{2n}{\delta} \sqrt[8]{\frac{3\nu}{\rho}} \right) \right)^{\frac{1}{d+2}} n^{\frac{d+1}{d+2}} \\
&= \sqrt{n} + \sqrt{2n \log(\frac{4n^2}{\delta})} + 3 \left(\frac{2^{3d+7} \nu^d KC\rho^d}{(1-\rho)^2} \right)^{\frac{1}{d+2}} \left(\log\left(\frac{2n}{\delta} \sqrt[8]{\frac{3\nu}{\rho}} \right) \right)^{\frac{1}{d+2}} n^{\frac{d+1}{d+2}} \\
&\leq 2\sqrt{2n \log(\frac{4n^2}{\delta})} + 3 \left(\frac{2^{3d+7} \nu^d KC\rho^d}{(1-\rho)^2} \right)^{\frac{1}{d+2}} \left(\log\left(\frac{2n}{\delta} \sqrt[8]{\frac{3\nu}{\rho}} \right) \right)^{\frac{1}{d+2}} n^{\frac{d+1}{d+2}}
\end{aligned}$$

with probability $1 - \delta$.

Appendix B. General analysis of POO

We prove Proposition 7 in this section. The analysis of POO originally proposed by Grill et al. (2015) consists in two main parts, that can be adapted to any base algorithm satisfying assumption (1) on its cumulative regret. In the following, we assume that $\nu^\star \le \nu_{\max}$ and $\rho^\star \le \rho_{\max}$.

The first part of the analysis consists in proving that there exists a parameter $\overline{\rho}$ such that $(\nu_{\max}, \overline{\rho}) \in \mathcal{G}$ and the instance $\mathcal{A}(\nu_{\max}, \overline{\rho})$ has its simple regret bounded in terms of the *true parameters* $(\nu^\star, \rho^\star)$. One important ingredient is the following lemma, which upper bounds the difference between the near-optimality dimension $d(\nu_{\max}, C, \rho)$ and $d(\nu^\star, C^\star, \rho^\star)$ for $\rho > \rho^\star$.

Lemma 11 (Appendix B.1 of Grill et al. 2015) *Under Assumption 1, for any choice of $\rho^\star$ and ρ s.t. $0 < \rho^\star < \rho < 1$, we have*

$$d(\nu_{\max}, C, \rho) - d(\nu^\star, C^\star, \rho^\star) \le \log K \left(\frac{1}{\log(1/\rho)} - \frac{1}{\log(1/\rho^\star)} \right).$$

Lemma 11 endorses the choice of grid $\mathcal{G} = \{(\nu_{\max}, \rho_{\max}^{2N/(2i+1)})_i\}$, which ensures that

$$\overline{\rho} \triangleq \underset{\rho_i \ge \rho^\star}{\arg\min} [d(\nu_{\max}, C_i, \rho_i) - d(\nu^\star, C^\star, \rho^\star)].$$

satisfies $d(\nu_{\max}, \overline{C}, \overline{\rho}) - d(\nu^\star, C^\star, \rho^\star) \le D_{\max}/N$, where $\overline{C}$ is associated to $\overline{\rho}$. A close examination of Appendix B.2 and B.3 of Grill et al. (2015) shows that under the assumption

$$\log \mathbb{E}\left[S_t^{\mathcal{A}(\nu_{\max}, \overline{\rho})}\right] \le \log \alpha + \frac{\log C(\nu_{\max}, \overline{\rho})}{d(\nu_{\max}, \overline{C}, \overline{\rho}) + 2} - \frac{\log(t/\log t)}{d(\nu_{\max}, \overline{C}, \overline{\rho}) + 2}, \tag{22}$$

the simple regret of $\mathcal{A}(\nu_{\max}, \overline{\rho})$ can also be related to $(\nu^\star, C^\star, \rho^\star)$: for some constant α',

$$\mathbb{E}\left[S_t^{\mathcal{A}(\nu_{\max}, \overline{\rho})}\right] \le \alpha' D_{\max} (\nu_{\max}/\nu^\star)^{D_{\max}} \left((\log^2 t)/t)^{1/(d(\nu^\star, C^\star, \rho^\star)+2)} \right) \tag{23}$$

under assumption described by (1) on the cumulative regret of the base algorithms. Note that (22) holds as the recommendation rule ensures that $\mathbb{E}[S_t] = \mathbb{E}[R_t]/t$.

The second part of the analysis controls the simple regret of $\texttt{POO}(\mathcal{A})$ by showing that the error made when choosing $s^\star \neq (\nu_{\max}, \overline{\rho})$ is negligible. We highlight that for this part, having cumulative regret guarantees is crucial. Denoting by $(x_{i,j})_{1 \le i \le n/N}$ the successive points selected by algorithm j and $(r_{i,j})_{1 \le i \le n/N}$ the reward observed, the final output of $\texttt{POO}(\mathcal{A})$ can be written

$$\widehat{x} = x_{I, \widehat{\jmath}} \text{ where } I \sim \mathcal{U}(\{1, \dots, n/N\}) \text{ and } \widehat{\jmath} = \arg\max_j \widehat{\mu}_j$$

with $\widehat{\mu}_j = (N/n) \sum_{i=1}^{n/N} r_{i,j}$. One can also define $\widetilde{\jmath} = \arg\max_j \mu_j$ with $\mu_j = \frac{N}{n} \sum_{i=1}^{n/N} f(x_{i,j})$ and $\overline{\jmath}$ to be the index of the instance such that $\rho_{\overline{\jmath}} = \overline{\rho}$. First, some concentration results (see Appendix B.4 of Grill et al. 2015) show that for all j, $\mathbb{E}[|\widehat{\mu}_j - \mu_j|] \le C/\sqrt{n/N}$. The simple regret can then be upper bounded as

$$\begin{aligned}
\mathbb{E}\left[S_n^{\texttt{POO}(\mathcal{A})}\right] &= \mathbb{E}[f^\star - f(\widehat{x})] = \mathbb{E}\left[f^\star - \frac{N}{n} \sum_{i=1}^{n/N} f(x_{i,\widehat{\jmath}}) \right] = \mathbb{E}\left[f^\star - \mu_{\widehat{\jmath}}\right] \\
&= \mathbb{E}[f^\star - \mu_{\overline{\jmath}}] + \mathbb{E}\left[\mu_{\overline{\jmath}} - \mu_{\widetilde{\jmath}}\right] + \mathbb{E}\left[\mu_{\widetilde{\jmath}} - \widehat{\mu}_{\widetilde{\jmath}}\right] + \mathbb{E}\left[\widehat{\mu}_{\widetilde{\jmath}} - \widehat{\mu}_{\widehat{\jmath}}\right] + \mathbb{E}\left[\widehat{\mu}_{\widehat{\jmath}} - \mu_{\widehat{\jmath}}\right]
\end{aligned}$$

The second and fourth terms in this sum are negative by definition of $\widetilde{\jmath}$ and $\widehat{\jmath}$ respectively, while the third and last terms are $O(\sqrt{N/n})$ using the concentration result mentioned above. As for the first term, one has

$$\mathbb{E}[f^\star - \mu_{\bar{\jmath}}] = \frac{N}{n}\mathbb{E}\left[\sum_{t=1}^{T}(f^\star - r_{i,\bar{\jmath}})\right] = \frac{N}{n}\mathbb{E}\Big[R_{n/N}^{\mathcal{A}(\nu_{\max},\overline{\rho})}\Big] = \mathbb{E}\Big[S_{n/N}^{\mathcal{A}(\nu_{\max},\overline{\rho})}\Big],$$

where again the recommendation rule matters. Using the upper bound (23) obtained in the first part of the analysis permits to conclude by noting that the first term is actually the leading term.