Adaptive black-box optimization got easier: HCT only needs local smoothness

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-lille1.fr
CNRS, Univ. Lille (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

Abstract

Hierarchical bandits is 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.

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 problem without any further assumption is a *mission impossible*. A natural solution is to assume a global smoothness of the reward function (Agrawal, 1995; Kleinberg, 2005; Kleinberg et al., 2008; Cope, 2009; Auer et al., 2007; Kleinberg et al., 2015). A weaker condition is the *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). Optimistic tree-based optimization algorithms (Munos, 2011; Valko et al., 2013; Preux et al., 2014; Azar et al., 2014) 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 POO (Grill et al. 2015), which *adapts to the smoothness* without the knowledge of it. POO is a *meta-algorithm* 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, 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. 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 is 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] by 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. The question whether HOO have similar regret guarantees under the desired assumption remains open.

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 function 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 $\mathbb{E}[\varepsilon_t|x_t] \triangleq 0$. 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

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

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

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 $\mathbb{E}[S_n] \leq \mathbb{E}[R_n]/n$.

2.1 Covering tree that guides the optimization

Hierarchical bandits rely on a tree-formed hierarchical partitioning $\mathcal{P} \triangleq \{\mathcal{P}_{h,i}\}$ defined recursively,

$$\mathcal{P}_{0,1} \triangleq \mathcal{X}, \mathcal{P}_{h,i} \triangleq \mathcal{P}_{h+1,2i-1} \cup \mathcal{P}_{h+1,2i}.$$

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 (1) the metric is not fully exploited by the algorithms and (2) the notion of near-optimality dimension independent of partitioning is ill-defined. Hence, it is natural to make smoothness assumptions directly related 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 (1) it covers a larger class of functions, (2) it only constrains f along the optimal path of the covering tree which is a plausible property in an optimization scenario, and (3) shows that the optimization is actually easier than it was previously believed.

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

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

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 problem should depend on the size of near-optimal regions, and 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, that is directly related to the partitioning.

Definition 1 (near-optimality dimension w.r.t. $\mathcal{P}$) *We define the near-optimality dimension of f w.r.t. $\mathcal{P}$ as*

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

where $\mathcal{N}_h(3\nu\rho^h)$[2] is the number of cells $\mathcal{P}_{h,i}$ s.t. $\sup_{x\in\mathcal{P}_{h,i}} f(x) \geq f^\star - 3\nu\rho^h$.

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

[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.

3. HCT under local smoothness w.r.t. $\mathcal{P}$

Analyzing HOO under Assumption 1 is not trivial. A key lemma in the analysis of HOO (Lemma 3 by Bubeck et al. 2011) which 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,[3] while producing too many nodes that are neither near-optimal nor sub-optimal. The concept of near-optimal and sub-optimal nodes is here characterized by the *sub-optimality gap* of the 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 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 the present a new analysis of it.

3.1 Description of HCT

The pseudocode of HCT (Algorithm 1) and a detailed snippet (Algorithm 2) describes 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),

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

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

$$B_{h,i}(t) \triangleq \begin{cases} \min\left\{U_{h,i}(t), \max\left\{B_{h+1,2i-1}(t), B_{h+1,2i}(t)\right\}\right\} & \text{if } (h,i) \text{ is a leaf} \\ U_{h,i}(t) & \text{otherwise,} \end{cases}$$

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*[4] of the region $\mathcal{P}_{h,i}$.

At each round t, the algorithm traverses the current covering tree along an *optimistic path* P_t before choosing a point. 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. Otherwise, the node that is not sampled enough is re-sampled. More precisely, HCT samples one node a certain number of times $\tau_h(t)$ in order to sufficiently reduce the uncertainty

[3] This can, however, bring good *empirical* performance in hyper-parameter tuning.

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

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

```
Input     : $\nu > 0$ $\rho \in (0, 1)$, $c > 0$, tree partition $\{\mathcal{P}_{h,i}\}$, confidence $\delta$
Initialize: $\mathcal{T}_1 \leftarrow \{(0, 1), (1, 1), (1, 2)\}$, $H(1) \leftarrow 1$, $U_{1,1}(1) \leftarrow U_{1,2}(1) \leftarrow +\infty$
for $t \leftarrow 1$ to $n$ do
    if $t = t^+$ then
        for $(h, i) \in \mathcal{T}_t$ do
            $U_{h,i}(t) \leftarrow \widehat{\mu}_{h,i}(t) + \nu\rho^h + c\sqrt{\frac{\log(1/\widetilde{\delta}(t^+))}{T_{h,i}(t)}}$
        end
        for $(h, i) \in \mathcal{T}_t$ backward from $H(t)$ do
            if $(h, i)$ is a leaf then
                $B_{h,i}(t) \leftarrow U_{h,i}(t)$
            else
                $B_{h,i}(t) \leftarrow \min\{U_{h,i}(t), \max\{B_{h+1,2i-1}(t), B_{h+1,2i}(t)\}\}$
            end
        end
    end
    $(h_t, i_t), P_t \leftarrow$ OptTraverse($\mathcal{T}_t$)
    Pull $x_{h_t,i_t}$ and obtain $r_t$
    $t \leftarrow t + 1$
    $T_{h_t,i_t}(t) \leftarrow T_{h_t,i_t}(t) + 1$
    Update $\widehat{\mu}_{h_t,i_t}(t)$
    $U_{h_t,i_t}(t) \leftarrow \widehat{\mu}_{h_t,i_t}(t) + \nu\rho^{h_t} + c\sqrt{\frac{\log(1/\widetilde{\delta}(t^+))}{T_{h_t,i_t}(t)}}$
    UpdateB($\mathcal{T}_t, P_t, (h_t, i_t)$)
    $\tau_{h_t}(t) \leftarrow \left\lceil \frac{c^2 \log(1/\widetilde{\delta}(t^+))}{\nu^2} \rho^{-2h_t} \right\rceil$
    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
```

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

Algorithm 2: `OptTraverse`

```
Input      : $\mathcal{T}_t$
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 and $T_{h,i}(t) \geq \tau_h(t)$ do
    if $B_{h+1,2i-1}(t) \geq B_{h+1,2i}(t)$ then
        $(h, i) \leftarrow (h + 1, 2i - 1)$
    else
        $(h, i) \leftarrow (h + 1, 2i)$
    end
    $P \leftarrow P \cup \{(h, i)\}$
end
return $(h, i)$ and $P$
```

Theorem 2 is the one of Definition 1 that is defined with respect to the partitioning and not with respect to a metric.

Theorem 2 *Assume that function f satisfies Assumption 1. Then, setting $\delta \triangleq 1/n$, the cumulative regret of* `HCT` *after n rounds is bounded as*

$$\mathbb{E}[R_n] \leq \mathcal{O}\left((\log n)^{1/(d+2)} n^{(d+1)/(d+2)}\right).$$

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

Corollary 3 *The simple regret of* `HCT` *is bounded as*

$$\mathbb{E}[S_n] \leq \mathcal{O}\left((\log n)^{1/(d+2)} n^{-1/(d+2)}\right).$$

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 Appendix 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 rounds is bounded as*

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

Next, we introduce a favorable event under which the mean reward of all expanded nodes is within a confidence interval,

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

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 bounded by $\sqrt{n}$ with high probability. 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

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

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, we 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 of `HCT` with probability $1-\delta$,

$$R_n \leq \mathcal{O}\left((\log(n/\delta))^{1/(d+2)} n^{(d+1)/(d+2)}\right).$$

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

4. POO (HCT) that does not need to know the smoothness

In this section, we present a new instantiation of POO, more precisely POO (HCT), which is POO using HCT as a subroutine. POO (parallel optimistic optimization) is a meta-algorithm that uses any hierarchical optimization algorithm that knows the smoothness as a subroutine. In POO (HCT), several instances of HCT are run in parallel. Each instance of HCT is run with a different pair of parameters (ν, ρ) in a well-chosen grid denoted by $\mathcal{G}$. In the end, POO (HCT) chooses the instance that performs the best. In particular, POO (HCT) first selects the instance that has the largest empirical mean reward and then returns one of the points evaluated by this instance, chosen uniformly at random.

Modularization of POO (HCT) Let $(\nu^\star, \rho^\star)$ be the optimal pair of parameters for POO, a pair for which Assumption 1 is true, and let $\overline{\rho} > \rho^\star$ and $\overline{\nu} > \nu^\star$ be the instance in $\mathcal{G}$ such that $d(\overline{\nu}, \overline{\rho})$ is the closest to $d(\nu^\star, \rho^\star)$—see its formal definition in Appendix B.1 of Grill et al. (2015). Since the regret bound obtained in the previous section matches the one of HOO up to constants, we can directly plug the simple regret bound into the second step of POO's analysis (Appendix B.2 of Grill et al. 2015) and obtain the following bound for the simple regret $S_n^{\overline{\nu}, \overline{\rho}}$ of the HCT instance run with $\overline{\rho}$,

$$S_n^{\overline{\nu}, \overline{\rho}} \leq \alpha((\log^2 n)/n)^{1/(d(\overline{\nu}, \overline{\rho})+2)},$$

where α is a constant depending on $\nu^\star$. This bound matches the regret bound of HCT up to a $\sqrt{\log n}$ factor. The empirically best instance that is picked by POO may not always be the instance that is run with this $\overline{\rho}$. However, as shown by Grill et al. (2015), the error is negligible w.r.t. the simple regret. We stress that the last statement is true because the underlying subroutine has a cumulative regret guarantee and POO chooses its best instance according to the empirical mean reward. Fortunately, this happens to be the case for HCT.

5. Discussion

We studied POO (HCT), 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. However, whether it is possible to weaken the assumptions of HOO in the same way while keeping similar regret guarantees remains open.

Acknowledgments

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 École 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. $\mathcal{X}$-armed bandits. *Journal of Machine Learning Research*, 12:1587–1627, 2011.

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.

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.

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.

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

Appendix A. Detailed regret analysis

A.1 Preliminaries

Before starting the proof, we first fix some constants and introduce some additional notation required for the analysis in Section 3.

- $c_1 \triangleq (\rho/(3\nu))^{1/8}$, $c \triangleq 2\sqrt{1/(1-\rho)}$
- $\forall 1 \leq h \leq H(t)$ and $t > 0$, $\mathcal{I}_h(t)$ denotes the set of all nodes created by `HCT` at level h up to step t
- $\forall 1 \leq h \leq H(t)$ 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.
- $\mathcal{C}_{h,i} \triangleq \{t = 1, \cdots, n : (h_t, i_t) = (h, i)\}$
- $\mathcal{C}_{h,i}^+ \triangleq \mathcal{C}_{h+1,2i-1} \cup \mathcal{C}_{h+1,2i}$
- $\bar{t}_{h,i} \triangleq \max_{t \in \mathcal{C}_{h,i}} t$ 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
- $t_{h,i} \triangleq \min\{t : T_{h,i}(t) \geq \tau_h(t)\}$ 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,

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

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 $t \leq t^+ \leq 2t$. In addition, log is an increasing function, thus we have

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

where the first inequality follows from the fact that $0 < \widetilde{\delta}(t) \leq 1/2$. 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 *rounds is bounded as*

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

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 $|\mathcal{I}_h(n)| = 2$ for all $h < H(n)$. Therefore, we have

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

By solving this expression, we obtain

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

■

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 $H_{\max}(t)$,

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

and we introduce the favorable event

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

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

Lemma 6 *With c_1 and c defined in Section A.1, for any fixed round t,*

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

Proof We upper-bound the probability of the complementary event ξ_t^c as

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

■

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

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

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

Lemma 7 *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*

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

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 $|\Delta_t| \le 1$. Next, we show that the probability that the second term is non zero is bounded by $\delta/(5n^2)$.

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

■

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* HCT *after n rounds is bounded as*

$$\mathbb{E}[R_n] \le \mathcal{O}\left((\log n)^{1/(d+2)} n^{(d+1)/(d+2)}\right).$$

We study the regret under events $\{\xi_t\}_t$ and prove that

$$R_n \le 2\sqrt{2n \log\left(\frac{4n^2}{\delta}\right)} + 3\left(\frac{2^{2d+6}\nu^d C \rho^d}{(1-\rho)^{1-d/2}}\right)^{\frac{1}{d+2}} \left(\log\left(\frac{2n}{\delta}\sqrt[8]{\frac{3\nu}{\rho}}\right)\right)^{\frac{1}{d+2}} n^{\frac{d+1}{d+2}}$$

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

$$R_n^\xi = \sum_{t=1}^n \Delta_{h_t,i_t}\mathbf{1}_{\xi_t} + \sum_{t=1}^n \widehat{\Delta}_t \mathbf{1}_{\xi_t} \le \sum_{t=1}^n \Delta_{h_t,i_t}\mathbf{1}_{\xi_t} + \sum_{t=1}^n \widehat{\Delta}_t = \widetilde{R}_n^\xi + \widehat{R}_n^\xi. \tag{2}$$

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

$$\widehat{R}_n^\xi \le \sqrt{2n\log\left(\frac{4n^2}{\delta}\right)} \tag{3}$$

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

$$B_{h',i'}(t) \le \max(B_{h'+1,2i'-1}(t), B_{h'+1,2i'}(t)) = B_{h'',i''}(t), \tag{4}$$

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

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

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

$$B_{h^\star,i^\star}(t) \le U_{h_t,i_t}(t), \tag{5}$$
$$B_{h^\star,i^\star}(t) \le U_{h_t^p,i_t^p}(t). \tag{6}$$

We now expand Eq. 5 on both sides under ξ_t. First, we have

$$U_{h_t,i_t}(t) \triangleq \widehat{\mu}_{h_t,i_t}(t) + \nu\rho^{h_t} + c\sqrt{\frac{\log(1/\widetilde{\delta}(t^+))}{T_{h_t,i_t}(t)}} \le f(x_{h_t,i_t}) + \nu\rho^{h_t} + 2c\sqrt{\frac{\log(1/\widetilde{\delta}(t^+))}{T_{h_t,i_t}(t)}} \tag{7}$$

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 Eq. 4, 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

$$B_{h^\star,i^\star}(t) \geq f^\star. \tag{8}$$

We combine Eq. 8 with Eq. 5, and Eq. 7 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

$$\Delta_{h_t,i_t} \leq 3c\sqrt{\frac{\log(2/\widetilde{\delta}(t))}{T_{h_t,i_t}(t)}}. \tag{9}$$

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

$$\begin{aligned}
\widetilde{R}_n^\xi &\triangleq \sum_{t=1}^{n} \Delta_{h_t,i_t} \mathbf{1}_{\xi_t} \\
&\le \sum_{h=0}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} \sum_{t=1}^{n} \Delta_{h,i} \mathbf{1}_{(h_t,i_t)=(h,i)} \mathbf{1}_{\xi_t} \\
&\le \sum_{h=0}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} \sum_{t=1}^{n} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(t))}{T_{h,i}(t)}} \mathbf{1}_{(h_t,i_t)=(h,i)} && \text{Eq. 9} \\
&= \sum_{h=0}^{\overline{H}} \sum_{i \in \mathcal{I}_h(n)} \sum_{t=1}^{n} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(t))}{T_{h,i}(t)}} \mathbf{1}_{(h_t,i_t)=(h,i)} + \sum_{h=\overline{H}+1}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} \sum_{t=1}^{n} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(t))}{T_{h,i}(t)}} \mathbf{1}_{(h_t,i_t)=(h,i)} \\
&\le \sum_{h=0}^{\overline{H}} \sum_{i \in \mathcal{I}_h(n)} \sum_{s=1}^{\tau_h(\bar{t}_{h,i})} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(\bar{t}_{h,i}))}{s}} + \sum_{h=\overline{H}+1}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} \sum_{s=1}^{T_{h,i}(n)} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(\bar{t}_{h,i}))}{s}} \\
&\le \sum_{h=0}^{\overline{H}} \sum_{i \in \mathcal{I}_h(n)} \int_{1}^{\tau_h(\bar{t}_{h,i})} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(\bar{t}_{h,i}))}{s}}\,\mathrm{d}s + \sum_{h=\overline{H}+1}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} \int_{1}^{T_{h,i}(n)} 3c\sqrt{\frac{\log(2/\widetilde{\delta}(\bar{t}_{h,i}))}{s}}\,\mathrm{d}s \\
&\le \sum_{h=0}^{\overline{H}} \sum_{i \in \mathcal{I}_h(n)} 6c\sqrt{\tau_h(\bar{t}_{h,i})\log(2/\widetilde{\delta}(\bar{t}_{h,i}))} + \sum_{h=\overline{H}+1}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} 6c\sqrt{T_{h,i}(n)\log(2/\widetilde{\delta}(\bar{t}_{h,i}))} \\
&= 6c\left(\underbrace{\sum_{h=0}^{\overline{H}} \sum_{i \in \mathcal{I}_h(n)} \sqrt{\tau_h(\bar{t}_{h,i})\log(2/\widetilde{\delta}(\bar{t}_{h,i}))}}_{\text{(a)}} + \underbrace{\sum_{h=\overline{H}+1}^{H(n)} \sum_{i \in \mathcal{I}_h(n)} \sqrt{T_{h,i}(n)\log(2/\widetilde{\delta}(\bar{t}_{h,i}))}}_{\text{(b)}} \right).
\end{aligned}$$

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

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

Notice that the covering tree is binary and therefore $|\mathcal{I}_h(n)| \le 2|\mathcal{I}_{h-1}(n)|$. 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,

$$|\mathcal{I}_h(n)| \le |2\mathcal{I}_{h-1}(n)| \le 2C\rho^{-d(h-1)},$$

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

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

Consequently, we bound (a) as

$$\text{(a)} \leq 2C\rho^{d}\frac{c\log\left(2/\widetilde{\delta}(n)\right)}{\nu}\frac{\rho^{-\overline{H}(d+1)}}{1-\rho}. \tag{10}$$

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

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

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,

$$\begin{aligned}
n &= \sum_{h=0}^{H(n)}\sum_{i\in\mathcal{I}_h(n)} T_{h,i}(n) \\
&\geq \sum_{h=0}^{H(n)-1}\sum_{i\in\mathcal{I}_h^+(n)} T_{h,i}(n) \\
&\geq \sum_{h=0}^{H(n)-1}\sum_{i\in\mathcal{I}_h^+(n)} T_{h,i}(\widetilde{t}_{h,i}) && \widetilde{t}_{h,i}\text{ well defined for } i\in\mathcal{I}_h^+(n) \\
&\geq \sum_{h=0}^{H(n)-1}\sum_{i\in\mathcal{I}_h^+(n)} \tau_h(\widetilde{t}_{h,i}) \\
&\geq \sum_{h=\overline{H}}^{H(n)-1}\sum_{i\in\mathcal{I}_h^+(n)} \tau_h(\widetilde{t}_{h,i}) \\
&= \sum_{h=\overline{H}}^{H(n)-1}\sum_{i\in\mathcal{I}_h^+(n)} \frac{c^2\log(1/\widetilde{\delta}(\widetilde{t}_{h,i}^+)))}{\nu^2}\rho^{-2h}
\end{aligned}$$

$$
\begin{aligned}
&\sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \frac{c^2\log(1/\widetilde{\delta}(\widetilde{t}^+_{h,i})))}{\nu^2}\rho^{-2h}\\
&\geq \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \frac{c^2\log(1/\widetilde{\delta}(\widetilde{t}^+_{h,i})))}{\nu^2}\rho^{-2\overline{H}}\\
&= \frac{c^2\rho^{-2\overline{H}}}{\nu^2} \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \log(1/\widetilde{\delta}(\widetilde{t}^+_{h,i})))\\
&= \frac{c^2\rho^{-2\overline{H}}}{\nu^2} \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \log(1/\widetilde{\delta}(\big[\max(\overline{t}_{h+1,2i-1},\overline{t}_{h+1,2i})\big]^+)) && \text{since } \widetilde{t}_{h,i} = \max(\overline{t}_{h+1,2i-1},\overline{t}_{h+1,2i})\\
&= \frac{c^2\rho^{-2\overline{H}}}{\nu^2} \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \log(1/\widetilde{\delta}(\max(\overline{t}^+_{h+1,2i-1},\overline{t}^+_{h+1,2i}))) && \forall t_1,t_2, [\max(t_1,t_2)]^+ = \max(t_1^+,t_2^+)\\
&= \frac{c^2\rho^{-2\overline{H}}}{\nu^2} \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \max(\log(1/\widetilde{\delta}(\overline{t}^+_{h+1,2i-1})),\log(1/\widetilde{\delta}(\overline{t}^+_{h+1,2i})))\\
&\quad\geq \frac{c^2\rho^{-2\overline{H}}}{\nu^2} \sum_{h=\overline{H}}^{H(n)-1} \sum_{i\in\mathcal{I}_h^+(n)} \frac{\log(1/\widetilde{\delta}(\overline{t}^+_{h+1,2i-1}))+\log(1/\widetilde{\delta}(\overline{t}^+_{h+1,2i}))}{2}\\
&= \frac{c^2\rho^{-2\overline{H}}}{\nu^2} \sum_{h=\overline{H}+1}^{H(n)} \sum_{i\in\mathcal{I}_{h-1}^+(n)} \frac{\log(1/\widetilde{\delta}(\overline{t}^+_{h,2i-1}))+\log(1/\widetilde{\delta}(\overline{t}^+_{h,2i}))}{2} && \text{change of variables}\\
&= \frac{c^2\rho^{-2\overline{H}}}{2\nu^2} \sum_{h=\overline{H}+1}^{H(n)} \sum_{i\in\mathcal{I}_h^+(n)} \log(1/\widetilde{\delta}(\overline{t}^+_{h,i})).
\end{aligned}
$$

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

$$
\sum_{h=\overline{H}+1}^{H(n)} \sum_{i\in\mathcal{I}_h^+(n)} \log(1/\widetilde{\delta}(\overline{t}^+_{h,i})) \leq \frac{2\nu^2\rho^{2\overline{H}}n}{c^2}\cdot \tag{11}
$$

On the other hand, we have

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

By plugging Eq. 11 into above expression, we get

$$\text{(b)} \leq \frac{2\nu\rho^{\overline{H}}n}{c}. \tag{12}$$

Now if we combine Eq. 12 with Eq. 10, we bound $\widetilde{R}_n^\xi$ as

$$\widetilde{R}_n^\xi \leq 12\nu \left[C\rho^d \frac{c^2 \log(2/\widetilde{\delta}(n))}{\nu^2} \frac{\rho^{-\overline{H}(d+1)}}{1-\rho} + \rho^{\overline{H}} n \right]. \tag{13}$$

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

$$\rho^{\overline{H}} = \left(\frac{C\rho^d c^2 \log(2/\widetilde{\delta}(n))}{n(1-\rho)\nu^2} \right)^{\frac{1}{d+2}}, \tag{14}$$

which after being plugged into Eq. 13 gives

$$\widetilde{R}_n^\xi \leq \frac{24\nu}{c} \left(\frac{C\rho^d c^2 \log(2/\widetilde{\delta}(n))}{(1-\rho)\nu^2} \right)^{\frac{1}{d+2}} n^{\frac{d+1}{d+2}}. \tag{15}$$

Finally, combining Eq. 15, Eq. 3, and Lemma 7, we obtain

$$\begin{aligned} R_n &\leq \sqrt{n} + \sqrt{2n \log(\frac{4n^2}{\delta})} + \frac{12\nu}{\sqrt{1/(1-\rho)}} \left(\frac{4C\rho^d}{(1-\rho)^2\nu^2} \right)^{\frac{1}{d+2}} \left(\log\left(\frac{2n}{\delta} \sqrt[8]{\frac{3\nu}{\rho}} \right) \right)^{\frac{1}{d+2}} n^{\frac{d+1}{d+2}} \\ &= \sqrt{n} + \sqrt{2n \log(\frac{4n^2}{\delta})} + 3 \left(\frac{2^{2d+6}\nu^d C\rho^d}{(1-\rho)^{1-d/2}} \right)^{\frac{1}{d+2}} \left(\log\left(\frac{2n}{\delta} \sqrt[8]{\frac{3\nu}{\rho}} \right) \right)^{\frac{1}{d+2}} n^{\frac{d+1}{d+2}} \\ &\leq 2\sqrt{2n \log(\frac{4n^2}{\delta})} + 3 \left(\frac{2^{2d+6}\nu^d C\rho^d}{(1-\rho)^{1-d/2}} \right)^{\frac{1}{d+2}} \left(\log\left(\frac{2n}{\delta} \sqrt[8]{\frac{3\nu}{\rho}} \right) \right)^{\frac{1}{d+2}} n^{\frac{d+1}{d+2}} \end{aligned}$$

with probability $1-\delta$.