A Proofs

We begin by introducing some notation. Let $\mathbf{x}_h^*$ denote the optimal node at level h. That is the cell of $\mathbf{x}_h^*$ contains the optimizer $\mathbf{x}^*$. Also let f^+ and $\mathbf{x}^+$ represent the best function value observed thus far and the associated node respectively.

A.1 Technical Lemmas

Lemma 1 (Lemma 5 of de Freitas et al. (2012)). *Given a set of points $\mathbf{x}_{1:T} := \{\mathbf{x}_1, \ldots, \mathbf{x}_T\} \in \mathcal{D}$ and a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$ with kernel κ the following bounds hold:*

1. *Any $f \in \mathcal{H}$ is Lipschitz continuous with constant $\|f\|_{\mathcal{H}} L$, where $\|\cdot\|_{\mathcal{H}}$ is the Hilbert space norm and L satisfies the following:*

$$L^2 \leq \sup_{\mathbf{x}\in\mathcal{D}} \partial_{\mathbf{x}} \partial_{\mathbf{x}'} \kappa(x, x')|_{\mathbf{x}=\mathbf{x}'}$$

and for $\kappa(\mathbf{x}, \mathbf{x}') = \widetilde{\kappa}(\mathbf{x} - \mathbf{x}')$ we have

$$L^2 \leq \partial_{\mathbf{x}}^2 \widetilde{\kappa}(\mathbf{x})|_{x=0}.$$

2. *The projection operator $P_{1:T}$ on the subspace $\operatorname{span}_{t=1:T}\{\kappa(x_t, \cdot)\} \subseteq \mathcal{H}$ is given by*

$$P_{1:T} f := \mathbf{k}^\top(\cdot) \mathbf{K}^{-1} \langle \mathbf{k}(\cdot), f \rangle$$

where $\mathbf{k}(\cdot) = \mathbf{k}_{1:T}(\cdot) := [\kappa(\mathbf{x}_1, \cdot) \cdots \kappa(\mathbf{x}_T, \cdot)]^\top$ and $\mathbf{K} := [\kappa(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1:T}$; moreover, we have that

$$\langle \mathbf{k}(\cdot), f \rangle := \begin{bmatrix} \langle \kappa(\mathbf{x}_1, \cdot), f \rangle \\ \vdots \\ \langle \kappa(\mathbf{x}_T, \cdot), f \rangle \end{bmatrix} = \begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_T) \end{bmatrix}.$$

Here $P_{1:T} P_{1:T} = P_{1:T}$ and $\|P_{1:T}\| \leq 1$ and $\|\mathbf{1} - P_{1:T}\| \leq 1$.

3. *Given tuples $(\mathbf{x}_i, f_i)$ with $f_i = f(\mathbf{x}_i)$, the minimum norm interpolation $\bar{f}$ with $\bar{f}(\mathbf{x}_i) = f(\mathbf{x}_i)$ is given by $\bar{f} = P_{1:T} f$. Consequently its residual $g := (\mathbf{1} - P_{1:T}) f$ satisfies $g(\mathbf{x}_i) = 0$ for all $\mathbf{x}_i \in \mathbf{x}_{1:T}$.*

Lemma 2 (Lemma 6 of de Freitas et al. (2012)). *Under the assumptions of Lemma 1 it follows that*

$$|f(\mathbf{x}) - P_{1:T} f(\mathbf{x})| \leq \|f\|_{\mathcal{H}} \sigma_T(\mathbf{x}),$$

where $\sigma_T^2(\mathbf{x}) = \kappa(\mathbf{x}, \mathbf{x}) - \mathbf{k}_{1:T}^\top(\mathbf{x}) \mathbf{K}^{-1} \mathbf{k}_{1:T}(\mathbf{x})$ and this bound is tight. Moreover, $\sigma_T^2(\mathbf{x})$ is the posterior predictive variance of a Gaussian process with the same kernel.

Lemma 3 (Adapted from Proposition 1 of de Freitas et al. (2012)). *Let $\kappa : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ be a kernel that is twice differentiable along the diagonal $\{(\mathbf{x}, \mathbf{x}) \,|\, \mathbf{x} \in \mathbb{R}^D\}$, with L defined as in Lemma 1.1, and f be an element of the RKHS with kernel κ. If f is evaluated at point $\mathbf{x}$, then for any other point $\mathbf{y}$ we have $\sigma_T(\mathbf{y}) \leq L\|\mathbf{x} - \mathbf{y}\|$.*

Proof. Let $\mathcal{H}$ be the RKHS corresponding to κ and $f \in \mathcal{H}$ an arbitrary element with $g := (\mathbf{1} - P_{1:T}) f$; the residual defined in lemma 1.3. Since $g \in \mathcal{H}$, we have by Lemma 1.1, g is Lipschitz. Thus we have that for any point $\mathbf{y}$:

$$|g(\mathbf{y})| \leq L\|g\|_{\mathcal{H}} \|\mathbf{y} - \mathbf{x}\| \leq L\|f\|_{\mathcal{H}} \|\mathbf{y} - \mathbf{x}\|, \tag{2}$$

where the second inequality is guaranteed by Lemma 1.2. On the other hand, by Lemma 2, we know that for all $\mathbf{y}$ we have the following tight bound:

$$|g(\mathbf{y})| \leq \|f\|_{\mathcal{H}} \sigma_T(\mathbf{y}) \tag{3}$$

Now, given the fact that both inequalities (2) and (3) are bounding the same quantity and that the latter is a tight estimate, we necessarily have that:

$$\|f\|_{\mathcal{H}} \sigma_T(\mathbf{y}) \leq L\|f\|_{\mathcal{H}} \|\mathbf{y} - \mathbf{x}\|.$$

Canceling $\|f\|_{\mathcal{H}}$ gives us the result. □

Lemma 4 (Adapted from Lemma 5.1 of Srinivas et al. (2010))**.** *Let f be a sample from a GP. Consider $\eta \in (0,1)$ and set $B_T = 2\log(\pi_T/\eta)$ where $\sum_{i=1}^{\infty} \pi_T^{-1} = 1$, $\pi_T > 0$. Then,*

$$|f(\mathbf{x}_T) - \mu_T(\mathbf{x}_T)| \leq B_T^{\frac{1}{2}} \sigma_T(\mathbf{x}_T)\ \forall T \geq 1$$

holds with probability at least $1-\eta$.

Proof. For $\mathbf{x}_T$ we have that $f(\mathbf{x}) \sim \mathcal{N}(\mu_T(\mathbf{x}_T), \sigma_T(\mathbf{x}_T))$ since f is a sample from the GP. Now, if $r \sim \mathcal{N}(0,1)$, then

$$\begin{aligned}\mathbb{P}(r > c) &= e^{-c^2/2}(2\pi)^{-1/2} \int e^{-(r-c)^2/2 - c(r-c)} dr \\ &< e^{-c^2/2}\mathbb{P}(r>0) = \frac{1}{2}e^{-c^2/2}.\end{aligned}$$

Thus we have that

$$\mathbb{P}\left(f(\mathbf{x}) - \mu_T(\mathbf{x}) > B_T^{1/2}\sigma_T(\mathbf{x})\right) = \mathbb{P}(r > B_T^{1/2}) < \frac{1}{2}e^{-B_T/2}.$$

By symmetry and the union bound, we have that $\mathbb{P}\left(|f(\mathbf{x}) - \mu_T(\mathbf{x})| > B_T^{1/2}\sigma_T(\mathbf{x})\right) < e^{-B_T/2}$. By applying the union bound again, we derive

$$\mathbb{P}\left(|f(\mathbf{x}) - \mu_T(\mathbf{x})| > B_T^{1/2}\sigma_T(\mathbf{x})\ \forall T \geq 1\right) < \sum_{T=1}^{\infty} e^{-B_T/2}.$$

By substituting $B_T = 2\log(\pi_T/\eta)$, we obtain the result. As in Srinivas et al. (2010), we can set $\pi_T = \pi^2 T^2/6$. □

Since each node's UCB and LCB are only evaluated at most once, we give the following shorthands in notation. Let $N(\mathbf{x})$ be the number of evaluations of confidence bounds by the time the UCB of $\mathbf{x}$ is evaluated (line 12 of Algorithm 3) and let $T(\mathbf{x}) = |\mathcal{D}_t|$ be the time the UCB of $\mathbf{x}$ is evaluated. Define $\mathcal{U}(\mathbf{x}) = \mathcal{U}_{N(\mathbf{x})}(\mathbf{x}|\mathcal{D}_{T(\mathbf{x})}) = \mu(\mathbf{x}|\mathcal{D}_{T(\mathbf{x})}) + B_{N(\mathbf{x})}\sigma(\mathbf{x}|\mathcal{D}_{T(\mathbf{x})})$ and $\mathcal{L}(\mathbf{x}) = \mathcal{L}_{N(\mathbf{x})}(\mathbf{x}|\mathcal{D}_{T(\mathbf{x})}) = \mu(\mathbf{x}|\mathcal{D}_{T(\mathbf{x})}) - B_{N(\mathbf{x})}\sigma(\mathbf{x}|\mathcal{D}_{T(\mathbf{x})})$.

Lemma 5. *Consider $\mathcal{B}(\mathbf{x}^*, \rho)$ and $\gamma in(0,1)$ as in Assumptions 2 and 3. Suppose $\mathcal{L}(\mathbf{x}_h^*) \leq f(\mathbf{x}_h^*) \leq \mathcal{U}(\mathbf{x}_h^*)$. If $\mathbf{x}_h^* \in \mathcal{B}(\mathbf{x}^*, \rho)$ and $\delta(h) < \epsilon_0$ then there exists a constant $\bar{c}$ such that $\mathcal{L}(\mathbf{x}_h^*) \geq f^* - \bar{c}B_{N(\mathbf{x}_h^*)}\gamma^{\frac{h}{2}}$.*

Proof. If $\mathbf{x}_h^*$ is not evaluated then $f(\mathbf{x}^+) \geq \mathcal{U}_T(\mathbf{x}_h^*) \geq f^* - \delta(h) \geq f^* - \epsilon_0$ which implies that $\mathbf{x}^+ \in \mathcal{B}(\mathbf{x}^*, \rho)$. Therefore, $f^* - c_2\|\mathbf{x}^+ - \mathbf{x}^*\|^2 \geq f(\mathbf{x}^+) \geq \mathcal{U}_T(\mathbf{x}_h^*) \geq f^* - \delta(h)$ which in turn implies that $\|\mathbf{x}^+ - \mathbf{x}^*\| \leq \sqrt{\frac{\delta(h)}{c_2}}$. Similarly $f^* - c_2\|\mathbf{x}_h^* - \mathbf{x}^*\|^2 \geq f(\mathbf{x}_h^*) \geq f^* - \delta(h)$. Therefore $\|\mathbf{x}_h^* - \mathbf{x}^*\| \leq \sqrt{\frac{\delta(h)}{c_2}}$. By the triangle inequality, we have

$$\|\mathbf{x}^+ - \mathbf{x}_h^*\| \leq \|\mathbf{x}^+ - \mathbf{x}^*\| + \|\mathbf{x}_h^* - \mathbf{x}^*\| \leq 2\sqrt{\frac{\delta(h)}{c_2}}.$$

By Lemma 3, we have that $\sigma_{T(\mathbf{x}_h^*)}(\mathbf{x}_h^*) \leq 2L\sqrt{\frac{\delta(h)}{c_2}}$. By the definition of $\mathcal{L}_T$, we can argue that

$$\begin{aligned}\mathcal{L}(\mathbf{x}_h^*) &\geq \mathcal{U}(\mathbf{x}_h^*) - 4B_{N(\mathbf{x}_h^*)}L\sqrt{\frac{\delta(h)}{c_2}} \\ &\geq f^* - \delta(h) - 4B_{N(\mathbf{x}_h^*)}L\sqrt{\frac{\delta(h)}{c_2}} \\ &= f^* - c\gamma^h - 4B_{N(\mathbf{x}_h^*)}L\sqrt{\frac{c\gamma^h}{c_2}}.\end{aligned}$$

Note that since $\gamma \in (0,1)$, $\gamma < \gamma^{1/2}$. Assume that $B_1 = b$. Let $\bar{c} = c/b + 4L\sqrt{\frac{c}{c_2}}$. Since $B_N > B_1\ \forall N > 1$, we have the statement.

If $\mathbf{x}_h^*$ is evaluated then the statement is trivially true. □

Definition 1. *Let* $\bar{\gamma} := \gamma^{\frac{1}{2}}$, $\bar{\delta}_h := \bar{c}B_{N(\mathbf{x}_h^*)}\bar{\gamma}^h$, *and* $I_h^\epsilon = \{(h,i) : f(\mathbf{x}_{h,i}) + \epsilon \geq f^*\}$.

Lemma 6. *Assume that* $h_{\max} = n^\epsilon$. *For a node* $\mathbf{x}_{h,i}$ *at level* h, $B_{N(\mathbf{x}_{h,i})} = \mathcal{O}(\sqrt{h})$.

Proof. Assume that there are n_i nodes expanded at the end of iteration i of the outer loop (the while loop). In the $i+1^{th}$ iteration of the outer loop, there can be at most $h_{\max}(n_i)$ additional expansions added. Thus the total number of expansions at the end of iteration i is at most $n_{i-1} + h_{\max}(n_{i-1})$. We can prove by induction that $n_i \leq i^{\frac{1}{1-\epsilon}}$. Since any node at level h would be expanded after at most 2^h iterations, at the time of expansion of any node at level h, we have that $n < (2^h)^{\frac{1}{1-\epsilon}} = 2^{\frac{h}{1-\epsilon}}$ where n is the total number of expansions. Thus, there would be at most $2 \times 2^{\frac{h}{1-\epsilon}}$ evaluations. Hence,

$$B_{N(\mathbf{x}_{h,i})} \leq \sqrt{2\log(\pi^2 2^{\frac{2h}{1-\epsilon}+2}/6\eta)} \leq \sqrt{2\log(2^{\frac{2h}{1-\epsilon}+2}) + 2\log(\pi^2/6\eta)} = \mathcal{O}(\sqrt{h}).$$

□

Lemma 7. *After a finite number of node expansions, an optimal node* $\mathbf{x}_{h_0}^* \in \mathcal{B}(\mathbf{x}^*, \rho)$ *is expanded such that* $\bar{c}B_{N(\mathbf{x}_{h_0}^*)}\bar{\gamma}_0^h \leq \epsilon_0$. *Also* $\forall h > h_0$, *we have that* $\bar{c}B_{N(\mathbf{x}_h^*)}\bar{\gamma}^h \leq \epsilon_0$ *and* $\mathbf{x}_h^* \in \mathcal{B}(\mathbf{x}^*, \rho)$.

Proof. Since it is clear that BaMSOO would expand every node after a finite number of node expansions, we only have to show that there exists an h_0 that satisfies the conditions. By Lemma 6, we have that $\forall h$ $B_{N(\mathbf{x}_h^*)} = \mathcal{O}(\sqrt{h})$. Since $\bar{\gamma} < 1$, there exists an h_0 such that $\bar{c}B_{N(\mathbf{x}_h^*)}\bar{\gamma}^h \leq \epsilon_0$ $\forall h > h_0$. Since $f(\mathbf{x}_h^*) > f^* - \delta(h) > f^* - \bar{c}B_{N(\mathbf{x}_h^*)}\bar{\gamma}^h \geq f^* - \epsilon_0$, we have by Assumption 2 that, $\mathbf{x}_h^* \in \mathcal{B}(\mathbf{x}^*, \rho)$. □

Lemma 8. $\sum_{h=0}^{H} |I_h^{\bar{\delta}(H)}| \leq C\left(B_{N(\mathbf{x}_H^*)}\right)^{D/2} \gamma^{(D/4-D/\alpha)H}$ *for some constant* C *for all* $H > h_0$.

Proof. By Lemma 7, we know that $\bar{\delta}(H) = \bar{c}B_{N(\mathbf{x}_H^*)}\bar{\gamma}^H < \epsilon_0$ if $H > h_0$. Therefore, by Assumption 2, we have that $\chi_{\bar{\delta}(H)} = \{\mathbf{x} \in \chi : f(\mathbf{x}) \geq f^* - \bar{\delta}(H)\} \subseteq \mathcal{B}(x^*, \rho)$. Again by Assumption 2, we have that

$$f^* - \bar{\delta}(H) \leq f(\mathbf{x}) \leq f^* - c_2\|\mathbf{x} - \mathbf{x}^*\|_2^2 \; \forall \mathbf{x} \in \chi_{\bar{\delta}(H)}.$$

Thus $\chi_{\bar{\delta}(H)} \subseteq \mathcal{B}\left(x^*, \sqrt{\frac{\bar{\delta}(H)}{c_2}}\right) = \mathcal{B}\left(x^*, \sqrt{\frac{\bar{c}B_{N(\mathbf{x}_H^*)}\gamma^{H/2}}{c_2}}\right)$.

Since each cell (h,i) contains a ℓ-ball of radius $\nu\delta(h)$ centered at $\mathbf{x}_{h,i}$ we have that each cell contains a ball $\mathcal{B}(\mathbf{x}_{h,i}, (\nu\delta(h))^{1/\alpha}) = \mathcal{B}(\mathbf{x}_{h,i}, (\frac{\nu c}{c_1})^{1/\alpha}\gamma^{h/\alpha})$. By the argument of volume, we have that $|I_h^{\bar{\delta}(H)}| \leq C_1\left(B_{N(\mathbf{x}_H^*)}\right)^{D/2}\gamma^{HD/4-hD/\alpha}$ for some constant C_1. Finally,

$$\begin{aligned}
\sum_{h=0}^{H} |I_h^{\bar{\delta}(H)}| &\leq C_1 \sum_{h=0}^{H}\left(B_{N(\mathbf{x}_H^*)}\right)^{D/2}\gamma^{HD/4-hD/\alpha} \\
&= C_1\left(B_{N(\mathbf{x}_H^*)}\right)^{D/2}\gamma^{HD/4}\sum_{h=0}^{H}\gamma^{-hD/\alpha} \\
&= C_1\left(B_{N(\mathbf{x}_H^*)}\right)^{D/2}\gamma^{HD/4}\sum_{h=0}^{H}\left(\gamma^{D/\alpha}\right)^{h-H} \\
&\leq C_1\left(B_{N(\mathbf{x}_H^*)}\right)^{D/2}\gamma^{HD/4}\sum_{h=0}^{\infty}\left(\gamma^{D/\alpha}\right)^{h-H} \\
&= C_1\left(B_{N(\mathbf{x}_H^*)}\right)^{D/2}\gamma^{HD/4}\frac{\gamma^{-DH/\alpha}}{1-\gamma^{D/\alpha}} \\
&= \frac{C_1}{1-\gamma^{D/\alpha}}\left(B_{N(\mathbf{x}_H^*)}\right)^{D/2}\gamma^{HD/4-DH/\alpha} \\
&= \frac{C_1}{1-\gamma^{D/\alpha}}\left(B_{N(\mathbf{x}_H^*)}\right)^{D/2}\gamma^{(D/4-D/\alpha)H}.
\end{aligned}$$

Setting $C = \frac{C_1}{1-\gamma^{D/\alpha}}$ gives us the desired result. □

Lemma 9. *Suppose $\mathcal{L}(\mathbf{x}_h^*) \le f(\mathbf{x}_h^*) \le \mathcal{U}(\mathbf{x}_h^*)$. If $\mathbf{x}_h^*$ is not evaluated (that is $\mathcal{U}(\mathbf{x}_h^*) < f^+$) then f^+ is $\delta(h)$-optimal.*

Proof. $f^+ > \mathcal{U}(\mathbf{x}_h^*) \ge f(\mathbf{x}_h^*) > f^* - \delta(h)$. □

A.2 Main Results

A.2.1 Simple Regret

Let h_n^* be the deepest level of an expanded optimal node with n node expansions. This following lemma is adapted from Lemma 2 of Munos (2011).

Lemma 10. *Suppose $\mathcal{L}(\mathbf{x}) \le f(\mathbf{x}) \le \mathcal{U}(\mathbf{x})$ for all $\mathbf{x}$ whose confidence region are evaluated. Whenever $h \le h_{\max}(n)$ and $n \ge C h_{\max}(n) \sum_{i=h_0}^{h} \left(B_{N(\mathbf{x}_i^*)}\right)^{D/2} \gamma^{(D/4-D/\alpha)i} + n_0$ for some constant C, we have $h_n^* \ge h$.*

Proof. We prove the statement by induction. By Lemma 7, we have that after n_0 node expansions, a node $\mathbf{x}_{h_0}^* \in \mathcal{B}(\mathbf{x}^*, \rho)$ is expanded. Also $\forall h > h_0$, we have that $\bar{c} B_{N(\mathbf{x}_h^*)} \bar{\gamma}^h \le \epsilon_0$ and $\mathbf{x}_h^* \in \mathcal{B}(\mathbf{x}^*, \rho)$. For $h = h_0$, the statement is trivially satisfied. Thus assume that the statement is true for h. Let n be such that $n \ge C h_{\max}(n) \sum_{i=h_0}^{h+1} \left(B_{N(\mathbf{x}_i^*)}\right)^{D/2} \gamma^{(D/4-D/\alpha)i} + n_0$. By the inductive hypothesis we have that $h_n^* \ge h$. Assume $h_n^* = h$ since otherwise the proof is finished. As long as the optimal node at level $h+1$ is not expanded, all nodes expanded at the level are $\bar{\delta}(h+1)$-optimal by Lemma 5. By Lemma 8, we know that after $C h_{\max}(n) \left(B_{N(\mathbf{x}_{h+1}^*)}\right)^{D/2} \gamma^{(D/4-D/\alpha)(h+1)}$ node expansions, the optimal node at level $h+1$ will be expanded since there are at most $\sum_{i=0}^{h+1} \left| I_i^{\bar{\delta}(h+1)} \right|$ $\bar{\delta}(h+1)$-optimal nodes at or beneath level $h+1$. Thus $h_n^* \ge h+1$. □

Theorem 1. *Suppose $\mathcal{L}(\mathbf{x}) \le f(\mathbf{x}) \le \mathcal{U}(\mathbf{x})$ for all $\mathbf{x}$ whose confidence region is evaluated. Let us write $h(n)$ to be the smallest integer $h \ge h_0$ such that*

$$C h_{\max}(n) \sum_{i=h_0}^{h} \left(B_{N(\mathbf{x}_i^*)}\right)^{D/2} \gamma^{(D/4-D/\alpha)i} + n_0 \ge n.$$

Then the loss is bounded as

$$r_n \le \delta(\min\{h(n), h_{\max}(n) + 1\})$$

and $h_n^* \ge \min\{h(n) - 1, h_{\max}(n)\}$.

Proof. From Lemma 8, and the definition of $h(n)$ we have that

$$C h_{\max}(n) \sum_{i=h_0}^{h(n)-1} \left(B_{N(\mathbf{x}_i^*)}\right)^{D/2} \gamma^{(D/4-D/\alpha)i} + n_0 < n.$$

By Lemma 10, we have that $h_n^* \ge h(n) - 1$ if $h(n) - 1 \le h_{\max}(n)$ and $h_n^* \ge h_{\max}(n)$ otherwise. Therefore $h_n^* \ge \min\{h(n) - 1, h_{\max}(n)\}$.

By Lemma 9, we know that if $\mathbf{x}_{h_n^*+1}^*$ is not evaluated then f^+ is $\delta(h_n^*+1)$-optimal. If $\mathbf{x}_{h_n^*+1}^*$ is evaluated, then $f\left(\mathbf{x}_{h_n^*+1}^*\right)$ is $\delta(h_n^*+1)$-optimal. Thus $r_n \le \delta(\min\{h(n), h_{\max}(n) + 1\})$. □

Proof of Corollary 1. Suppose $\mathcal{L}(\mathbf{x}) \le f(\mathbf{x}) \le \mathcal{U}(\mathbf{x})$ for all $\mathbf{x}$ whose confidence region is evaluated. By Lemma 4, we know that this holds with probability at least $1 - \eta$.

By the definition of $h(n)$ we have that

$$\begin{aligned} n &\leq Ch_{\max}(n) \sum_{i=h_0}^{h(n)} \left(B_{N\left(\mathbf{x}_i^*\right)}\right)^{D/2} \gamma^{(D/4-D/\alpha)i} + n_0 \\ &\leq Ch_{\max}(n) \left(B_{N\left(\mathbf{x}_{h(n)}^*\right)}\right)^{D/2} \sum_{i=h_0}^{h(n)} \gamma^{-di} + n_0 \\ &\leq Ch_{\max}(n) \left(B_{N\left(\mathbf{x}_{h(n)}^*\right)}\right)^{D/2} \gamma^{-dh_0} \frac{\gamma^{-dh(n)}-1}{\gamma^{-d}-1} + n_0 \end{aligned} \tag{4}$$

If $h(n) \leq h_{\max}(n)+1$, then by Theorem 1, we have that $h_n^* \geq h(n)-1$. After n expansions, the optimal node $\mathbf{x}_{h(n)-1}^*$ has been expanded which suggests that its children's confidence bounds have been evaluated. Hence, $N\left(\mathbf{x}_{h(n)}^*\right) < 2n$ since there have only been n expansions. Therefore,

$$(4) \leq Kn^{\epsilon} \left(B_{2n}\right)^{D/2} \gamma^{-dh(n)}$$

for some constant K which implies that

$$\gamma^{h(n)} \leq K^{1/d} B_{2n}^{\frac{2\alpha}{4-\alpha}} n^{-\frac{1-\epsilon}{d}} = K^{1/d} \left[2\log(4\pi^2 n^2/6\eta)\right]^{\frac{\alpha}{4-\alpha}} n^{-\frac{1-\epsilon}{d}}.$$

By Theorem 1, we have that

$$r_n \leq c \min\left\{K^{1/d} \left[2\log(4\pi^2 n^2/6\eta)\right]^{\frac{\alpha}{4-\alpha}} n^{-\frac{1-\epsilon}{d}}, \gamma^{(n+1)^{\epsilon}}\right\} = \mathcal{O}\left(n^{-\frac{1-\epsilon}{d}} \log^{\frac{\alpha}{4-\alpha}}(n^2/\eta)\right).$$

□