A SUPPLEMENTAL DOCUMENT

A.1 Proof of Lemma 1

Let X^t be the $(t-1) \mod K+1$th selected context in round $\lceil t/K \rceil$, i.e., $X^t = x^{\lceil t/K \rceil}_{s^{\lceil t/K \rceil}_{(t-1) \mod K+1}}$ and let Y^t be its outcome, i.e., $Y^t = r(x^{\lceil t/K \rceil}_{s^{\lceil t/K \rceil}_{(t-1) \mod K+1}})$. Assume that the sequence of random variables $X^1, Y^1, X^2, Y^2, \ldots$ are defined over some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consistent with the assumptions made in the paper. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and

$$\mathcal{F}_t = \sigma(X^1, \ldots, X^{t+K}, Y^1, \ldots, Y^t), t > 0.$$

$\mathbb{F} = (\mathcal{F}_t)_{t=0}^{\infty}$ is a filtration of $\mathcal{F}$. We have that $\mathbb{E}[Y^t | \mathcal{F}_{t-1}] = \mu(X^t)$.

Fix a node $x_{h,i}$. Let θ^t be the indicator variable that represents the event that X^t is in the cell associated with node $x_{h,i}$ after $x_{h,i}$ was created (this also counts the times when X^t is in $X_{h,i}$ after $x_{h,i}$ is expanded into children nodes). When $\theta^t = 1$, we say that $x_{h,i}$ is *observed*. Let $Z_0 = 0$ and

$$Z_t = \sum_{j=1}^{t} (\mu(X^j) - Y^j)\theta^j .$$

Note that Z_t is $\mathcal{F}_t$-measurable. Moreover, X^t and θ^t are all $\mathcal{F}_{t-1}$-measurable while Y^t is not. Thus, we have

$$\begin{aligned}
\mathbb{E}[Z_t | \mathcal{F}_{t-1}] &= \mathbb{E}\left[(\mu(X^t) - Y^t)\theta^t \middle| \mathcal{F}_{t-1}\right] + \mathbb{E}\left[\sum_{j=1}^{t-1} (\mu(X^j) - Y^j)\theta^j \middle| \mathcal{F}_{t-1}\right] \\
&= (\mu(X^t) - \mathbb{E}[Y^t | \mathcal{F}_{t-1}])\theta^t + \sum_{j=1}^{t-1} (\mu(X^j) - Y^j)\theta^j \\
&= (\mu(X^t) - \mu(X^t))\theta^t + Z_{t-1} \\
&= Z_{t-1} .
\end{aligned}$$

This shows that $(Z_t)_{t=0}^{\infty}$ is an $\mathbb{F}$-adapted martingale.

Let

$$\tilde{C}^t(x_{h,i}) = \sum_{j=1}^{t} \theta^j .$$

We define a sequence of random stopping times when the arms associated with nodes $x_{h,i}$ were *observed*: $T_j = \min\{l : \tilde{C}^l(x_{h,i}) = j\}$. Note that $\{T_j = t\}$ is $\mathcal{F}_{t-1}$-measurable for all t and (T_j) is an increasing sequence of stopping times, i.e., $1 \leq T_1 < T_2 < \ldots$, hence it holds that $T_j \geq j$. Our next aim is to apply Doob's optional skipping theorem ([Doob, 1953], Theorem 2.3). As its application requires $T_1 < T_2 < \ldots < \infty$, we assume that after T there are infinitely many rounds in which all contexts in everyone of them arrive in $x_{h,i}$ (this ensures that arms with contexts in $x_{h,i}$ are observed infinitely often). Note that this technical assumption has no effect on our model as it happens after our horizon of interest. Let $\tilde{Z}_j = Z_{T_j}$. Then, by Doob's optional skipping theorem $(\tilde{Z}_j)$ is a martingale w.r.t. $(\mathcal{F}_{T_j})$.

Let (H^t, I^t) represent the level and the index of the leaf node that contains X^t. We denote by $\tilde{X}^j = X^{T_j}$ the jth arm pulled in the region corresponding to $x_{h,i}$ and $\tilde{Y}^j = Y^{T_j}$. Let $n = C^t(x_{h,i})$. Note that when $x_{h,i} \in \mathcal{L}^t$, we have $\tilde{C}^{Kt}(x_{h,i}) = C^t(x_{h,i})$ and $\theta^j = \mathbb{I}((H^j, I^j) = (h,i))$. We have

$$\begin{aligned}
&\mathbb{P}\Big\{ \Big| \sum_{j=1}^{t} \sum_{k=1}^{K} (\mu(x^j_{s^j_k}) - r(x^j_{s^j_k})) \mathbb{I}((H^j_k, I^j_k) = (h,i)) \Big| \geq \sqrt{2n \log T}, x_{h,i} \in \mathcal{L}^t \Big\} \\
&= \sum_{l=1}^{Kt} \mathbb{P}\Big\{ \Big| \sum_{j=1}^{t} \sum_{k=1}^{K} (\mu(x^j_{s^j_k}) - r(x^j_{s^j_k})) \mathbb{I}((H^j_k, I^j_k) = (h,i)) \Big| \geq \sqrt{2n \log T}, x_{h,i} \in \mathcal{L}^t, n = l \Big\} \qquad (3) \\
&= \sum_{l=1}^{Kt} \mathbb{P}\Big\{ \Big| \sum_{j=1}^{Kt} (\mu(X^j) - Y^j) \mathbb{I}((H^j, I^j) = (h,i)) \Big| \geq \sqrt{2n \log T}, x_{h,i} \in \mathcal{L}^t, \tilde{C}^{Kt}(x_{h,i}) = l \Big\}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{l=1}^{Kt} \mathbb{P}\Big\{ \Big| \sum_{j=1}^{l} (\mu(\tilde{X}^j) - \tilde{Y}^j) \Big| \geq \sqrt{2l \log T}, x_{h,i} \in \mathcal{L}^t, \tilde{C}^{Kt}(x_{h,i}) = l \Big\} \\
&= \sum_{l=1}^{Kt} \mathbb{P}\Big\{ |\tilde{Z}_l| \geq \sqrt{2l \log T}, x_{h,i} \in \mathcal{L}^t, \tilde{C}^{Kt}(x_{h,i}) = l \Big\} \\
&\leq \sum_{l=1}^{Kt} \mathbb{P}\Big\{ \Big|\tilde{Z}_l\Big| \geq \sqrt{2l \log T} \ \Big\} \\
&\leq \sum_{l=1}^{Kt} 2 \exp\left(-\frac{2(\sqrt{2l \log T})^2}{l} \right) \qquad (4) \\
&\leq \sum_{l=1}^{KT} 2 \exp\left(-\frac{2(\sqrt{2l \log T})^2}{l} \right) \\
&= 2KT \cdot T^{-4} = 2KT^{-3} \qquad (5)
\end{aligned}
$$

where (3) follows from the law of total probability and for (4) we use the Azuma-Hoeffding inequality for martingale differences noting that the outcomes lie in the unit interval. Therefore, with probability at least $1 - 2KT^{-3}$ we have

$$
\sum_{j=1}^{t} \sum_{k=1}^{K} (r(x^j_{s^j_k}) - \mu(x^j_{s^j_k})) \mathbb{I}((H^j_k, I^j_k) = (h, i)) \leq \sqrt{2n \log T} \ .
$$

Next, by Definition 1 and Assumption 1, when $\mathbb{I}((H^j_k, I^j_k) = (h,i)) = 1$, we have that $\mu(x^j_{s^j_k}) - \mu(x_{h,i}) \leq v_1 \rho^h$ since the context lives in $X_{h,i}$. Consequently, by using the triangle inequality, it follows that

$$
\begin{aligned}
&\Big| \sum_{j=1}^{t} \sum_{k=1}^{K} (r(x^j_{s^j_k}) - \mu(x_{h,i})) \mathbb{I}((H^j_k, I^j_k) = (h,i)) \Big| \\
&\leq \Big| \sum_{j=1}^{t} \sum_{k=1}^{K} (r(x^j_{s^j_k}) - \mu(x^j_{s^j_k})) \mathbb{I}((H^j_k, I^j_k) = (h,i)) \Big| + \Big| \sum_{j=1}^{t} \sum_{k=1}^{K} (\mu(x^j_{s^j_k}) - \mu(x_{h,i})) \mathbb{I}((H^j_k, I^j_k) = (h,i)) \Big| \\
&\leq \sqrt{2n \log T} + n v_1 \rho^h \ .
\end{aligned}
$$

Dividing both sides by n, we obtain

$$
|\hat{\mu}^t(x_{h,i}) - \mu(x_{h,i})| \leq c^t(x_{h,i}) + v_1 \rho^h \ .
$$

Note that until round T, the maximal number of possible new node activations is KT and therefore for the above event to hold simultaneously for all nodes at all times, we take the union bound over all nodes and over all times. From this, we obtain

$$
\mathbb{P}\Big\{ \forall t \leq T, \, \forall x_{h,i} \in \mathcal{L}^t : \, |\hat{\mu}^t(x_{h,i}) - \mu(x_{h,i})| \leq c^t(x_{h,i}) + v_1 \rho^h \Big\} \geq 1 - 2K^2 T^{-1} \ .
$$

A.2 Proof of Lemma 2

The following inequalities hold:

$$
\begin{aligned}
g^t(x_{h,i}) &= b^t(x_{h,i}) + v_1 \rho^h && (6) \\
&\leq \hat{\mu}^{t-1}(p(x_{h,i})) + c^{t-1}(p(x_{h,i})) + v_1 \rho^{(h-1)} + v_1 \rho^h && (7) \\
&\leq \mu(p(x_{h,i})) + 2c^{t-1}(p(x_{h,i})) + 2v_1 \rho^{(h-1)} + v_1 \rho^h && (8) \\
&\leq (\mu(p(x_{h,i})) - v_1 \rho^{(h-1)}) + 5v_1 \rho^{(h-1)} + v_1 \rho^h && (9) \\
&\leq \mu(x_{h,i}) + 5v_1 \rho^{(h-1)} + v_1 \rho^h && (10) \\
&\leq \mu(x_{h,i}) + (5N v_1 / v_2 + 1) v_1 \rho^h && (11)
\end{aligned}
$$

where the inequality in (7) follows from the definition of $b^t(x_{h,i})$, while (8) follows from Lemma 1. The inequality in (9) follows from the fact that $p(x_{h,i})$ must have been expanded, thus we must have had $c^{t-1}(p(x_{h,i})) \leq v_1\rho^{(h-1)}$. For the inequality in (10) we use Assumption 1 for the means of nodes and the fact that $x_{h,i}$ and $p(x_{h,i})$ lie in the cell associated with $p(x_{h,i})$. Lastly, the inequality in (11) follows from the triangle inequality. By Definition 1, the cell $X_{h-1,i}$ is expanded into N disjoint cells of diameter less than $v_1\rho^h$, for any $1 \leq i \leq N^{h-1}$, therefore $v_2\rho^{h-1} \leq Nv_1\rho^h$.

Finally, solving $c^T(x_{h,i}) = v_1\rho^h$ gives $C^T(x_{h,i}) = 2\log T/(v_1\rho^h)^2$. Thus, we must have

$$C^T(x_{h,i}) \leq \left\lceil \frac{2\log T}{(v_1\rho^h)^2} \right\rceil \leq \frac{2\log T}{(v_1\rho^h)^2} + 1 \leq \frac{2\log T + v_1^2}{(v_1\rho^h)^2} = \frac{2\log T + 2\log(\sqrt{e^{v_1^2}})}{(v_1\rho^h)^2} = \frac{2\log(Tv_3)}{(v_1\rho^h)^2}$$

where $v_3 = \sqrt{e^{v_1^2}}$. ■

A.3 Proof of Lemma 3

First, note that we have $N(v_1/v_2) \geq 1$ (since $N \geq 2$ and $v_2 \leq 1 \leq v_1$). Now, for a context $x_m^t \in X_{h,i}$, the index of its base arm is

$$g^t(x_m^t) = g^t(x_{h,i}) + N(v_1/v_2)v_1\rho^h \ .$$

We now have on event $\mathcal{F}$

$$\begin{aligned} g^t(x_m^t) &= g^t(x_{h,i}) + N(v_1/v_2)v_1\rho^h \\ &= b^t(x_{h,i}) + v_1\rho^h + N(v_1/v_2)v_1\rho^h \\ &= \min\{\hat{\mu}^{t-1}(x_{h,i}) + c^{t-1}(x_{h,i}), \hat{\mu}^{t-1}(p(x_{h,i})) + c^{t-1}(p(x_{h,i})) + v_1\rho^{(h-1)}\} + v_1\rho^h + N(v_1/v_2)v_1\rho^h \ . \end{aligned}$$

We have two cases. If the minimum is $\hat{\mu}^{t-1}(x_{h,i}) + c^{t-1}(x_{h,i})$, then we have

$$\begin{aligned} g^t(x_m^t) &= \hat{\mu}^{t-1}(x_{h,i}) + c^{t-1}(x_{h,i}) + v_1\rho^h + N(v_1/v_2)v_1\rho^h \\ &\geq \hat{\mu}^{t-1}(x_{h,i}) + c^{t-1}(x_{h,i}) + 2v_1\rho^h \\ &\geq \mu(x_{h,i}) + v_1\rho^h \\ &\geq \mu(x_m^t) \end{aligned}$$

since event $\mathcal{F}$ holds, $N(v_1/v_2) \geq 1$, and also by Assumption 1 and Definition 1.

If the minimum is $\hat{\mu}^{t-1}(p(x_{h,i})) + c^{t-1}(p(x_{h,i})) + v_1\rho^{(h-1)}$, then we have

$$\begin{aligned} g^t(x_m^t) &= \hat{\mu}^{t-1}(p(x_{h,i})) + c^{t-1}(p(x_{h,i})) + v_1\rho^{(h-1)} + N(v_1/v_2)v_1\rho^h + v_1\rho^h \\ &\geq \mu(p(x_{h,i})) + N(v_1/v_2)v_1\rho^h + v_1\rho^h \\ &\geq \mu(p(x_{h,i})) + v_1\rho^{h-1} + v_1\rho^h \\ &\geq \mu(x_{h,i}) + v_1\rho^h \\ &\geq \mu(x_m^t) \end{aligned}$$

since $N(v_1/v_2)v_1\rho^h \geq v_1\rho^{h-1}$ and $|\mu(p(x_{h,i})) - \mu(x_{h,i})| \leq v_1\rho^{h-1}$. ■

A.4 Proof of Lemma 4

Let $g(\boldsymbol{x}_{S^t}^t) = [g^t(x_m^t)]_{m\in S^t}$ be the vector of indices of base arms in S^t. Then we have:

$$u(g(\boldsymbol{x}_{S^t}^t)) \geq \alpha \cdot u(g(\boldsymbol{x}_{G^t}^t)) \geq \alpha \cdot u(g(\boldsymbol{x}_{S^{*t}}^t)) \geq \alpha \cdot u(\mu(\boldsymbol{x}_{S^{*t}}^t))$$

where $G^t := \text{argmax}_{S\in\mathcal{S}^t} u(g(\boldsymbol{x}_S^t))$. The first inequality holds because we choose S^t by the α-approximation oracle; the second inequality follows from the definition of G^t; the third inequality follows from Lemma 3 and Assumption 3, since $g(x_m^t) \geq \mu(x_m^t)$, for all $m \in S^{*t}$. Letting $\Delta(S^t)$ be the approximate optimality gap of S^t, we now have

$$
\begin{aligned}
\Delta(S^t) &= \alpha \cdot \mathrm{opt}(\mu^t) - u(\mu(\boldsymbol{x}^t_{S^t})) \\
&\le u(g(\boldsymbol{x}^t_{S^t})) - u(\mu(\boldsymbol{x}^t_{S^t})) && (12)\\
&\le B \cdot \sum_{m \in S^t} |g^t(x^t_m) - \mu(x^t_m)| && (13)\\
&\le B \cdot \sum_{m \in S^t} |g^t(x_{\tilde{h}^t_m, \tilde{i}^t_m}) + N(v_1/v_2)v_1\rho^{\tilde{h}^t_m} - \mu(x_{\tilde{h}^t_m, \tilde{i}^t_m})| + |\mu(x_{\tilde{h}^t_m, \tilde{i}^t_m}) - \mu(x^t_m)| && (14)\\
&\le B \cdot \sum_{m \in S^t} |g^t(x_{\tilde{h}^t_m, \tilde{i}^t_m}) - \mu(x_{\tilde{h}^t_m, \tilde{i}^t_m})| + v_1\rho^{\tilde{h}^t_m} + N(v_1/v_2)v_1\rho^{\tilde{h}^t_m} && (15)\\
&\le B \cdot \sum_{m \in S^t} (6Nv_1/v_2 + 2)v_1\rho^{\tilde{h}^t_m} && (16)\\
&\le BK(6Nv_1/v_2 + 2)v_1\rho^{h(t)} && (17)
\end{aligned}
$$

where (12) follows from the above argument; (13) follows from the Lipschitz continuity of the expected reward function; in (14) we have added and subtracted the true mean of the context $x_{\tilde{h}^t_m, \tilde{i}^t_m}$ (which represents a node), and than used the triangle inequality; in (15) we have used the Lipschitz continuity for the expected outcomes and Definition 1, in the sense that any two contexts in the same cell are at most a diameter away from each other; (16) follows from Lemma 2 and (17) follows from the way we have defined $h(t)$. ■

A.5 Proof of Lemma 5

By Definition 2 we have that

$$\limsup_{r \to 0} \frac{\log(M(\mathcal{X}^\kappa_{cr}, \|\cdot\|_2, r))}{\log(r^{-1})} \le \bar{D}$$

therefore, since $D_1 > \bar{D}$, there exists $r(D_1)$ such that for all $r \le r(D_1)$ we have

$$\frac{\log(M(\mathcal{X}^\kappa_{cr}, \|\cdot\|_2, r))}{\log(r^{-1})} \le D_1$$

and thus, for all $r \le r(D_1)$, we have

$$M(\mathcal{X}^\kappa_{cr}, \|\cdot\|_2, r) \le r^{-D_1} .$$

If $r(D_1) \ge v_2$, then we let $Q = 1$ and we are done. If $r(D_1) < v_2$, then we have to show that the result holds for all $r(D_1) \le r \le v_2$. First, by the definition of the r-packing number, we have that

$$M(\mathcal{X}^\kappa_{cr}, \|\cdot\|_2, r) \le M(\mathcal{X}, \|\cdot\|_2, r(D_1))$$

since $|\mathcal{X}^\kappa_{cr}| \le |\mathcal{X}|$ and $r \ge r(D_1)$. Thus, for all $r \in [r(D_1), v_2]$, we can say that

$$M(\mathcal{X}^\kappa_{cr}, \|\cdot\|_2, r) \le \frac{v_2^{D_1}}{r^{D_1}} M(\mathcal{X}, \|\cdot\|_2, r(D_1)) \le Qr^{-D_1}$$

for $Q = \max\{1, M(\mathcal{X}, \|\cdot\|_2, r(D_1)) \cdot v_2^{D_1}\}$. ■

A.6 Proof of Theorem 1

By Lemma 4, at any given round t, the algorithm only selects contexts from $(BK(6Nv_1/v_2+2)v_1\rho^{h(t)}, u, \alpha u^*_{\min})$-optimal sets. Note that for all h, we have $|\mathcal{X}_h| = N^h$ by Definition 1. Furthermore, we select only contexts from sets of the form $\mathcal{X}^{\alpha u^*_{\min}}_{BK(6Nv_1/v_2+2)v_1\rho^{h(t)}}$, for all $t \ge 1$. From Lemma 4, we have that

$$R_\alpha(T) \le \sum_{t \le T} \sum_{k=1}^{K} B(6Nv_1/v_2 + 2)v_1\rho^{H^t_k} \le \sum_{t \le T} BK(6Nv_1/v_2 + 2)v_1\rho^{h(t)} .$$

Note that in each round, we consider the contribution that comes from the highest leaf node in the tree (i.e., the leaf node with the smallest h) associated with the selected super arm. In order to obtain sublinear regret, we express the summation in terms of the levels of the tree. In order to do that, we observe that we may have $h(t) = h(t')$ for $t \neq t'$. That is for several reasons. First, we have up to $N^{h(t)}$ nodes in a certain level, but the learner does not select contexts associated to all those nodes, only the $BK(6Nv_1/v_2+2)v_1\rho^{h(t)}$-optimal ones. But since we want to define a notion of regret that captures volatility of the base arms and structure of the expected reward function, we may think that the learner selects contexts from the set $\mathcal{X}^{\alpha u^*_{\min}}_{BK(6Nv_1/v_2+2)v_1\rho^{h(t)}}$, as given in Definition 2. Second, the learner may select a node several times, up to q_h times by Lemma 2, until it is expanded. Using this approach, we will consider splitting the regret into two sums.

We fix a positive integer H. We will first bound the regret coming from rounds t such that $h(t) < H$, which we will denote by $R^1_\alpha(T)$, and the rounds t such that $h(t) \geq H$ by $R^2_\alpha(T)$. We have

$$R_\alpha(T) \leq R^1_\alpha(T) + R^2_\alpha(T) \ .$$

We bound $R^1_\alpha(T)$ first. Letting $\bar{D} = D^u(f, \alpha u^*_{\min})$ be the $(f, u, \alpha u^*_{\min})$-optimality dimension, where $f(r) = BK(6Nv_1/v_2+2)(v_1/v_2)r$, then for any $D_1 > \bar{D}$, there exists $Q > 0$ (by Lemma 5), for which we have

$$\begin{aligned}
R^1_\alpha(T) &:= \sum_{\substack{t\leq T:\\ h(t)<H}} BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^{h(t)} \\
&= \sum_{h=0}^{H-1}\sum_{t\leq T} \mathbb{I}(h(t)=h) BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^h && (18)\\
&\leq \sum_{h=0}^{H-1} |\mathcal{X}_h \cap \mathcal{X}^{\alpha u^*_{\min}}_{BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^h}| \cdot BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^h \cdot q_h && (19)\\
&\leq \sum_{h=0}^{H-1} M(\mathcal{X}^{\alpha u^*_{\min}}_{BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^h}, \|\cdot\|_2, v_2\rho^h) \cdot BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^h \cdot q_h && (20)\\
&\leq \sum_{h=0}^{H-1} Q\cdot(v_2\rho^h)^{-D_1} BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^h \cdot \frac{2\log(Tv_3)}{(v_1\rho^h)^2} && (21)\\
&= \sum_{h=0}^{H-1} QBK(6Nv_1/v_2+2)\frac{(v_2)^{-D_1}}{v_1}\cdot\frac{2\log(Tv_3)}{\rho^{h(D_1+1)}} \\
&= 2QBK(6Nv_1/v_2+2)\frac{{v_2}^{-D_1}}{v_1}\log(Tv_3)\sum_{h=0}^{H-1}\frac{1}{\rho^{h(D_1+1)}} \\
&\leq 2QBK(6Nv_1/v_2+2)\frac{{v_2}^{-D_1}}{v_1}\frac{1-(\rho^{-1})^{H(D_1+1)}}{(1-\rho^{-1})}\cdot\log(Tv_3) \\
&\leq 2QBK(6Nv_1/v_2+2)\frac{{v_2}^{-D_1}}{v_1(\rho^{-1}-1)}\rho^{-H(D_1+1)}\log(Tv_3) \ . && (22)
\end{aligned}$$

For (19) we argue as follows. We have expressed the summation in terms of levels of the tree. For a certain level h, in T rounds, the learner may have selected up to $|\mathcal{X}_h \cap \mathcal{X}^{\alpha u^*_{\min}}_{BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^h}|$ nodes (following an argument similar to the proof of Lemma 4, it can be verified that the active leaf nodes associated with the cells that contain the selected contexts, i.e., ϕ^t_m, $m \in S^t$ are also in $(BK(6Nv_1/v_2+2)v_1\rho^{h(t)}, u, \alpha u^*_{\min})$-optimal sets), and any of them up to q_h times. Any of these nodes contributes to the regret with a maximum amount of $BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^h$. Summing over all levels, we are sure that the expected cumulative regret cannot exceed this number. (20) follows from the definition of $v_2\rho^h$-packing number and the fact that any two nodes in $\mathcal{X}_h$ are at least $v_2\rho^h$ apart; in (21) we use Lemma 5, since $v_2\rho^h \leq v_2$, for any $h \geq 0$, and the fact that $q_h \leq 2\log(Tv_3)/(v_1\rho^h)^2$ in ACC-UCB.

Next, we bound $R^2_\alpha(T)$.

$$R^2_\alpha(T) := \sum_{\substack{t\leq T:\\ h(t)\geq H}} BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^{h(t)}$$

$$\leq TKB(6Nv_1/v_2 + 2)v_1\rho^H \ . \tag{23}$$

From (22) and (23), we define

$$C_1 := 2QBK(6Nv_1/v_2 + 2)\frac{{v_2}^{-D_1}}{v_1(\rho^{-1} - 1)}$$

and

$$C_2 := KB(6Nv_1/v_2 + 2)v_1 \ .$$

We have that

$$R_\alpha(T) \leq C_1\rho^{-H(D_1+1)}\log(Tv_3) + C_2\rho^H T \ .$$

If we let

$$H = \frac{\log T - \log(\log(Tv_3))}{(D_1+2)\log(1/\rho)}$$

so that (equivalently) we have

$$\begin{aligned} H &= \frac{\log T - \log(\log(Tv_3))}{(D_1+2)\log(1/\rho)} = \frac{\log\left(\frac{T}{\log(Tv_3)}\right)}{(D_1+2)\log(1/\rho)} = \frac{1}{(D_1+2)}\log_{1/\rho}\left(\frac{T}{\log(Tv_3)}\right) \\ &= -\log_\rho\left(\frac{T}{\log(Tv_3)}\right)^{\frac{1}{D_1+2}} \ , \end{aligned}$$

then we would obtain

$$\rho^{-H} = \left(\frac{T}{\log(Tv_3)}\right)^{\frac{1}{D_1+2}} \text{ and } \rho^H = \left(\frac{\log(Tv_3)}{T}\right)^{\frac{1}{D_1+2}} \ .$$

And thus, substituting in the bounds we get:

$$\begin{aligned} R_\alpha(T) &\leq C_1\rho^{-H(D_1+1)}\log(Tv_3) + C_2\rho^H T \\ &= C_1 \cdot \left(\frac{T}{\log(Tv_3)}\right)^{\frac{D_1+1}{D_1+2}} \cdot \log(Tv_3) + C_2 \cdot \left(\frac{\log(Tv_3)}{T}\right)^{\frac{1}{D_1+2}} \cdot T \\ &= C_1 \cdot T^{1-\frac{1}{D_1+2}} \cdot (\log(Tv_3))^{1-\frac{D_1+1}{D_1+2}} + C_2 \cdot T^{1-\frac{1}{D_1+2}} \cdot (\log(Tv_3))^{\frac{1}{D_1+2}} \\ &= C_1 \cdot T^{1-\frac{1}{D_1+2}} \cdot (\log(Tv_3))^{\frac{1}{D_1+2}} + C_2 \cdot T^{1-\frac{1}{D_1+2}} \cdot (\log(Tv_3))^{\frac{1}{D_1+2}} \ . \end{aligned}$$

■

A.7 Proof of Theorem 2

Let $|\mathcal{X}| = \Pi$. The r-packing number of any subset X of $\mathcal{X}$ cannot exceed Π. Let $f(r) = BK(6Nv_1/v_2 + 2)(v_1/v_2)r$ and $\kappa = \alpha u^*_{\min}$. Let

$$P(f, u, \kappa) := \limsup_{r\to 0} M(\mathcal{X}^\kappa_{f(r)}, \|\cdot\|_2, r)$$

be the largest possible r-packing number depending on on f, u and κ. Note that the r-packing number of a set cannot exceed its cardinality, and thus, we have that $P(f, u, \kappa) \leq \Pi$. Therefore,

$$D^u(f, \kappa) = \limsup_{r\to 0} \frac{\log(M(\mathcal{X}^\kappa_{f(r)}, \|\cdot\|_2, r))}{\log(r^{-1})}$$

$$\leq \limsup_{r\to 0} \frac{\log \Pi}{\log(r^{-1})}$$
$$= 0\ .$$

For this case, Lemma 5 implies that for any $\epsilon > 0$, there exists some positive constant $Q(\epsilon) > 0$ such that we have

$$M(\mathcal{X}^{\kappa}_{f(r)}, \|\cdot\|_2, r) \leq Q r^{-\epsilon}\ .$$

Since $M(\mathcal{X}^{\kappa}_{f(r)}, \|\cdot\|_2, r) \leq P(f, u, \kappa)$, we can let $Q = P(f, u, \kappa)$. Now, using the same approach as in Theorem 1, we fix an H, the value of which we will determine later, and first consider the contribution to the regret of levels $h < H$. We have

$$\begin{aligned}
R^1_{\alpha}(T) &:= \sum_{\substack{t\leq T:\\ h(t)<H}} BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^{h(t)} \\
&\leq \sum_{h=0}^{H-1} M(\mathcal{X}^{\alpha u^*_{\min}}_{BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^h}, \|\cdot\|_2, v_2\rho^h) \cdot BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^h \cdot q_h \\
&\leq \sum_{h=0}^{H-1} P(f,u,\kappa)\cdot (v_2\rho^h)^{-\epsilon} BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^h \cdot \frac{2\log(Tv_3)}{(v_1\rho^h)^2} \\
&\leq 2P(f,u,\kappa)BK(6Nv_1/v_2+2)\frac{{v_2}^{-\epsilon}}{v_1(\rho^{-1}-1)}\rho^{-H(\epsilon+1)}\log(Tv_3) \\
&= C_3(\epsilon)\rho^{-H(\epsilon+1)}\log(Tv_3)\ .
\end{aligned}$$

For $R^2_{\alpha}(T)$, we have

$$\begin{aligned}
R^2_{\alpha}(T) &:= \sum_{\substack{t\leq T:\\ h(t)\geq H}} BK(6Nv_1/v_2+2)(v_1/v_2)v_2\rho^{h(t)} \\
&\leq TKB(6Nv_1/v_2+2)v_1\rho^H \\
&= C_4\rho^H T\ .
\end{aligned}$$

Again, we define H as,

$$H = \frac{\log T - \log(\log(Tv_3))}{(\epsilon+2)\log(1/\rho)}$$

and obtain the following bound by summing up $R^1_{\alpha}(T)$ and $R^2_{\alpha}(T)$:

$$R_{\alpha}(T) \leq C_3(\epsilon)\cdot T^{1-\frac{1}{\epsilon+2}}\cdot(\log(Tv_3))^{\frac{1}{\epsilon+2}} + C_4\cdot T^{1-\frac{1}{\epsilon+2}}\cdot(\log(Tv_3))^{\frac{1}{\epsilon+2}}\ .$$

■

A.8 Proof of Theorem 3

Fix an ordered partition $\Gamma \in \mathcal{P}(\pi([T]))$. Note that

$$R_{\alpha}(T) = \sum_{\lambda=1}^{|\Gamma|}\left(\alpha \sum_{t\in\mathcal{T}^{\Gamma}_{\lambda}} \mathrm{opt}(\mu^t) - \sum_{t\in\mathcal{T}^{\Gamma}_{\lambda}} u(\mu(\boldsymbol{x}^t_{S^t}))\right) = \sum_{\lambda=1}^{|\Gamma|} R_{\alpha,\lambda}(\mathcal{T}^{\Gamma}_{\lambda})$$

where $R_{\alpha,\lambda}(\mathcal{T}^{\Gamma}_{\lambda})$ denotes the total regret incurred in rounds in $\mathcal{T}^{\Gamma}_{\lambda}$.

For any $\lambda \leq |\Gamma|$, we can define a permutation $\zeta^{\Gamma}_{\lambda} : \mathcal{T}^{\Gamma}_{\lambda} \to [T^{\Gamma}_{\lambda}]$, such that $\zeta_{\lambda}(t^{\Gamma_i}_{\lambda}) = i$, for any $i \in [T^{\Gamma}_{\lambda}]$, where $t^{\Gamma_i}_{\lambda}$ is an element of the ordered set $\mathcal{T}^{\Gamma}_{\lambda} = \{t^{\Gamma_1}_{\lambda}, \ldots, t^{\Gamma_{\lambda}}_{\lambda}\}$. Thus, abusing the terminology, we can solve the bandit subproblems with T^{Γ}_{λ} rounds, for each natural $\lambda \leq |\Gamma|$.

From Lemma 4 we have:

$$R_{\alpha,\lambda}(\mathcal{T}_\lambda^\Gamma) \leq \sum_{t\in\mathcal{T}_\lambda^\Gamma}\sum_{k=1}^{K} B(6Nv_1/v_2+2)v_1\rho^{H_k^t}$$
$$\leq \sum_{t\in\mathcal{T}_\lambda^\Gamma} BK(6Nv_1/v_2+2)v_1\rho^{h(t)} \, .$$

We use an approach that is similar to the proof of Theorem 1. Fix a positive integer H_λ^Γ whose value will be determined later. Let

$$R^1_{\alpha,\lambda}(\mathcal{T}_\lambda^\Gamma) := \sum_{t\in\mathcal{T}_\lambda^\Gamma : h(t)<H_\lambda^\Gamma} BK(6Nv_1/v_2+2)v_1\rho^{h(t)}$$

and

$$R^2_{\alpha,\lambda}(\mathcal{T}_\lambda^\Gamma) := \sum_{t\in\mathcal{T}_\lambda^\Gamma : h(t)\geq H_\lambda^\Gamma} BK(6Nv_1/v_2+2)v_1\rho^{h(t)} \, .$$

We first bound $R^1_{\alpha,\lambda}(T_\lambda^\Gamma)$. Following steps analogous to (19)-(22) in the proof of Theorem 1, it can be shown that there exists some $Q_\lambda^\Gamma > 0$ (that depends only on $\mathcal{X}$, u, μ, $\alpha u^*_{\min}(\Gamma,\lambda)$ and c but not explicitly on T) such that for $D_\lambda^\Gamma > \bar{D}_\lambda^\Gamma$ we have

$$R^1_{\alpha,\lambda}(\mathcal{T}_\lambda^\Gamma) \leq C_5(\lambda,\Gamma)\cdot \rho^{-H_\lambda^\Gamma(D_\lambda^\Gamma+1)}\cdot \log(Tv_3)$$

where

$$C_5(\lambda,\Gamma) := 2Q_\lambda^\Gamma KB(6Nv_1/v_2+2)\frac{{v_2}^{-D_\lambda^\Gamma}}{v_1(1/\rho-1)} \, .$$

For $R^2_{\alpha,\lambda}(T_\lambda^\Gamma)$, we have

$$R^2_{\alpha,\lambda}(\mathcal{T}_\lambda^\Gamma) \leq C_6\cdot\rho^{H_\lambda^\Gamma}\cdot T_\lambda^\Gamma$$

where

$$C_6 := KB(6Nv_1/v_2+2)v_1 \, .$$

Now let us define H_λ^Γ analogously with Theorem 1, in order to obtain sublinear regret:

$$H_\lambda^\Gamma = \frac{\log T_\lambda^\Gamma - \log(\log(Tv_3))}{(D_\lambda^\Gamma+2)\log(1/\rho)} \, .$$

Substituting H_λ^Γ, we obtain the following.

$$R_{\alpha,\lambda}(\mathcal{T}_\lambda^\Gamma) \leq C_5(\lambda,\Gamma)\rho^{-H_\lambda^\Gamma(D_\lambda^\Gamma+1)}\log(Tv_3) + C_6\rho^{H_\lambda^\Gamma}T_\lambda^\Gamma$$
$$\leq C_5(\lambda,\Gamma)\cdot (T_\lambda^\Gamma)^{1-\frac{1}{D_\lambda^\Gamma+2}}\cdot(\log(Tv_3))^{\frac{1}{D_\lambda^\Gamma+2}} + C_6\cdot(T_\lambda^\Gamma)^{1-\frac{1}{D_\lambda^\Gamma+2}}\cdot(\log(Tv_3))^{\frac{1}{D_\lambda^\Gamma+2}}$$

Finally, by summing over all $\lambda \leq |\Gamma|$, we obtain

$$R_\alpha(T) = \sum_{\lambda=1}^{|\Gamma|} R_{\alpha,\lambda}(T_\lambda^\Gamma)$$
$$\leq \sum_{\lambda=1}^{|\Gamma|}\left(C_5(\lambda,\Gamma)\cdot (T_\lambda^\Gamma)^{1-\frac{1}{D_\lambda^\Gamma+2}}\cdot(\log(Tv_3))^{\frac{1}{D_\lambda^\Gamma+2}} + C_6\cdot(T_\lambda^\Gamma)^{1-\frac{1}{D_\lambda^\Gamma+2}}\cdot(\log(Tv_3))^{\frac{1}{D_\lambda^\Gamma+2}}\right)$$
$$\leq \sum_{\lambda=1}^{|\Gamma|} O\left((T_\lambda^\Gamma)^{1-\frac{1}{D_\lambda^\Gamma+2}}\cdot(\log(Tv_3))^{\frac{1}{D_\lambda^\Gamma+2}}\right) \, .$$

■

A.9 Proof of Corollary 1

Fix $\xi \in (0,1)$. Let $\mathcal{T}^\xi = \{\pi(1), \dots, \pi(T - \lfloor T^\xi \rfloor)\}$, where π is the permutation defined in Section 4.4. The idea is to bound separately the regret coming from the rounds in $\mathcal{T}^\xi$ and those in $[T] - \mathcal{T}^\xi$. First we consider $\mathcal{T}^\xi$. From Lemma 4 we have:

$$R_\alpha(\mathcal{T}^\xi) := \alpha \sum_{t \in \mathcal{T}^\xi} \text{opt}(\mu^t) - \sum_{t \in \mathcal{T}^\xi} u(\mu(\boldsymbol{x}^t_{S^t})) \leq \sum_{t \in \mathcal{T}^\xi} BK(6Nv_1/v_2 + 2)v_2 \rho^{h(t)} .$$

Using the result of Theorem 3 the above argument can be bounded as

$$R_\alpha(\mathcal{T}^\xi) \leq C_5(\xi) \cdot T^{1 - \frac{1}{D(\xi)+2}} \cdot (\log(Tv_3))^{\frac{1}{D(\xi)+2}} + C_6 \cdot T^{1 - \frac{1}{D(\xi)+2}} \cdot (\log(Tv_3))^{\frac{1}{D(\xi)+2}}$$

where

$$C_5(\xi) := 2Q(\xi)KB(6Nv_1/v_2 + 2) \frac{{v_2}^{-D(\xi)}}{v_1(1/\rho - 1)} .$$

Here, $Q(\xi) > 0$ is a constant that depends on $\mathcal{X}$, u, μ, $\alpha u^*_{\min}(\xi)$ and c but not explicitly on T. On the other hand, the regret incurred in round in $[T] - \mathcal{T}^\xi$ is upper bounded by $\alpha u^*_{\max} T^\xi$. The final result follows from summing these two bounds and taking the infimum. ■

A.10 Proof of Example 1

We have

$$\begin{aligned}
\mathcal{X}^\kappa_{c\epsilon} &= \{x \in \mathcal{X} : \kappa - u(\mu(x)) \leq c\epsilon\} \\
&= \{x \in \mathcal{X} : (1 - \|x^*\|_2^a) - (1 - \|x\|_2^a) \leq c\epsilon\} \\
&= \{x \in \mathcal{X} : \|x\|_2^a \leq c\epsilon\} \\
&= \{x \in \mathcal{X} : \|0 - x\|_2^a \leq ((c\epsilon)^{1/a})^a\} \\
&= \{x \in \mathcal{X} : \|0 - x\|_2 \leq (c\epsilon)^{1/a}\} .
\end{aligned}$$

This set is an $\|\cdot\|_2$-ball with center at 0 and radius $(c\epsilon)^{1/a}$. What we want to know is, for a fixed ϵ, how many disjoint $\|\cdot\|_2$-balls of radius ϵ can fit inside of it (so that we can estimate its ϵ-packing number). Thus, we have

$$M(\mathcal{X}^\kappa_{c\epsilon}, \|\cdot\|_2, r) \leq \left(\frac{(c\epsilon)^{1/a}}{\epsilon^1} \right)^D = c^{D/a} (\epsilon^{-1})^{(1-1/a)D} .$$

This together with Lemma 5 implies that $\bar{D} \leq (1 - 1/a)D$ for $a > 1$. ■

A.11 Table of Notation

Table 2: Notation.

Symbol	Meaning
$(\mathcal{X}, \|\cdot\|_2)$	The metric space of contexts (D-dimensional)
$\mathcal{M}^t$	The set of available base arms in round t
$\mathcal{S}^t$	$\{S \subset \mathcal{M}^t : \|S\| = K\}$, the set of available super arms in round t
x_m^t	Context of base arm $m \in \mathcal{M}^t$ in round t
$\mathcal{X}^t$	$\{x_m^t\}_{m \in \mathcal{M}^t}$
$r(x_m^t)$	(Random) outcome of base arm $m \in \mathcal{M}^t$ in round t
$\mu(x)$	Expected outcome of a base arm with context x
μ^t	$[\mu(x_m^t)]_{m \in \mathcal{M}^t}$
$\boldsymbol{x}_S^t$	$[x_{s_1}^t, \ldots, x_{s_K}^t]$, K-tuple of contexts of base arms that are in super arm $S \in \mathcal{S}^t$
$u(\cdot)$	Reward function
opt(μ^t)	$\max_{S \in \mathcal{S}^t} u(\mu(\boldsymbol{x}_S^t))$
S^t	Super arm selected by the learner in round t
S^{*t}	$\text{argmax}_{S \in \mathcal{S}^t} u(\mu(\boldsymbol{x}_S^t))$, the optimal super arm in round t
$x_{h,i}$	The ith node in the hth level of the tree of partitions
$X_{h,i}$	The cell associated to node $x_{h,i}$
$\mathcal{X}_h$	$\{x_{h,i},\ 1 \leq i \leq N^h\}$
$M(\mathcal{X}, \|\cdot\|_2, r)$	The r-packing number of $(\mathcal{X}, \|\cdot\|_2)$
$\mathcal{X}_{f(r)}^{\kappa}$	The $(f(r), u, \kappa)$-optimal set
$D^u(f, \kappa)$	The (f, u, κ)-optimality dimension
$g^t(x_{h,i})$	The index of node $x_{h,i}$ at time t
$g^t(x_m^t)$	The index of base arm $m \in \mathcal{M}^t$ in round t
$\hat{\mu}^t(x_{h,i})$	The sample mean of the outcomes of the selected base arms in cell $X_{h,i}$ by round t
$C^t(x_{h,i})$	The number of times a base arm with context in $X_{h,i}$ was selected by round t
$c^t(x_{h,i})$	The confidence radius of cell $X_{h,i}$ in round t
(H_k^t, I_k^t)	(h, i) index of the active leaf node associated with the kth selected base arm (s_k^t) in round t
$(\tilde{h}_m^t, \tilde{i}_m^t)$	(h, i) index of the active leaf node associated with base arm $m \in \mathcal{M}^t$ in round t
$\mathcal{L}^t$	The set of active leaf nodes in round t
$\mathcal{N}^t$	The set of available active leaf nodes in round t
$\mathcal{P}^t$	The set of active leaf nodes selected by the algorithm in round t