Appendix-A

In this appendix we provide the materials to complete the proof of Theorem 3.3.

Lemma 5.1. *Let,* $r^* = \lceil (1/\alpha)\log((1/\rho)\log(1/\rho))\rceil$. *Then, for every phase* $r \geq r^*$, *the size of* $\mathcal{K}_r$ *can be lower bounded as* $n_r = \lceil t_r^\alpha \rceil \geq \lceil \frac{\alpha}{(1+\gamma)\rho} \cdot \ln t_r^{\log \mathrm{e}} \rceil$, *wherein,* $0.53 < \gamma \stackrel{\text{def}}{=} \max_x \frac{\log\log x}{\log x} < 0.531$.

Proof. We notice, for every, $r \geq r^*$, $t_r \geq \left\lceil \left(\frac{1}{\rho}\log\frac{1}{\rho}\right)^{\frac{1}{\alpha}} \right\rceil$. Then, for each $r \geq r^*$, we can lower bound the size of the set $\mathcal{K}_r$ as follows. As, n_r is an integer, to ease the calculation let us define $s_u = 2^u$, where $u \in \mathbb{R}^+$, and s_u does not need to be an integer. Now, letting $u^* \stackrel{\text{def}}{=} \log\left(\frac{1}{\rho}\log\frac{1}{\rho}\right)^{\frac{1}{\alpha}}$, we get

$$
\begin{aligned}
&\frac{1}{\rho}\log s_{u^*} \\
&= \frac{1}{\rho}\log\left(\frac{1}{\rho}\log\frac{1}{\rho}\right)^{\frac{1}{\alpha}} \\
&= \frac{1}{\alpha\rho}\left(\log\frac{1}{\rho} + \log\log\frac{1}{\rho}\right) \\
&= \frac{1}{\alpha\rho}\log\frac{1}{\rho}\left(1 + \frac{\log\log\frac{1}{\rho}}{\log\frac{1}{\rho}}\right) \\
&\leq \frac{1+\gamma}{\alpha\rho}\log\frac{1}{\rho}\ \left[\text{as } \gamma \stackrel{\text{def}}{=} \max_x \frac{\log\log x}{\log x}\right] \\
&= \frac{1+\gamma}{\alpha}\left(\frac{1}{\rho}\log\frac{1}{\rho}\right) \\
&= \frac{1+\gamma}{\alpha} s_{u^*}^\alpha . \\
&\implies s_{u^*}^\alpha \geq \frac{\alpha}{(1+\gamma)\rho}\log s_{u^*}. \qquad (5)
\end{aligned}
$$

As, s_u^α grows with u faster than $\log s_u$, therefore,

$$\forall u \geq u^*,\ s_u^\alpha \geq \frac{\alpha}{(1+\gamma)\rho}\log s_u. \qquad (6)$$

Therefore, recalling that r is an integer, for all values of $r \geq \lceil u^* \rceil$, the statement of the lemma follows. □

Assuming $r^* = \lceil r^* \rceil$, below we present the detailed steps for obtaining (3) in the proof of Theorem 3.3.

$$
\begin{aligned}
&\sum_{r=1}^{\log T} t_r \Pr\{E_r(\rho)\} \\
&= \sum_{r=1}^{r^*-1} t_r \Pr\{E_r(\rho)\} + \sum_{r=r^*}^{\log T} t_r \Pr\{E_r(\rho)\} \\
&\leq \sum_{r=1}^{r^*-1} t_r + \sum_{r=r^*}^{\log T} {t_r}^{1-\frac{\alpha\log \mathrm{e}}{1+\gamma}} \leq t_{r^*} + \sum_{r=r^*}^{\log T} {t_r}^{1-\frac{\alpha\log \mathrm{e}}{1+\gamma}} \\
&\leq t_{r^*} + \sum_{j=1}^{\log T - r^*} (T/2^j)^{1-\frac{\alpha\log \mathrm{e}}{1+\gamma}} \\
&\leq 2^{\left\lceil \log\left(\frac{1}{\rho}\log\frac{1}{\rho}\right)^{\frac{1}{\alpha}} \right\rceil} + T^{1-\frac{\alpha\log \mathrm{e}}{1+\gamma}} \sum_{j=0}^{\log T - r^*} \left(\frac{1}{2}\right)^{j(1-\frac{\alpha\log \mathrm{e}}{1+\gamma})} \\
&< 2^{\log\left(\frac{1}{\rho}\log\frac{1}{\rho}\right)^{\frac{1}{\alpha}}+1} + T^{1-\frac{\alpha\log \mathrm{e}}{1+\gamma}} \sum_{j=0}^{\infty} \left(\frac{1}{2}\right)^{j(1-\frac{\alpha\log \mathrm{e}}{1+\gamma})} \\
&= O\left(2^{\log\left(\frac{1}{\rho}\log\frac{1}{\rho}\right)^{\frac{1}{\alpha}}} + T^{1-\frac{\alpha\log \mathrm{e}}{1+\gamma}}\right) \\
&= O\left(\left(\frac{1}{\rho}\log\frac{1}{\rho}\right)^{\frac{1}{\alpha}} + T^{1-\frac{\alpha\log \mathrm{e}}{1+\gamma}}\right).
\end{aligned}
$$

Below are the detailed steps for obtaining (4) in the proof of Theorem 3.3.

$$
\begin{aligned}
\sum_{r=r^*}^{\log T} C\sqrt{n_r t_r} &= \sum_{j=0}^{\log T - r^*} C\sqrt{\frac{T}{2^j}\left(\frac{T}{2^j}\right)^\alpha} \\
&= CT^{(1+\alpha)/2} \cdot \sum_{j=0}^{\log T - r^*} \left(\frac{1}{2^{(1+\alpha)/2}}\right)^j \\
&< CT^{(1+\alpha)/2} \cdot \sum_{j=0}^{\infty} \left(\frac{1}{2^{(1+\alpha)/2}}\right)^j \\
&\leq C' T^{(1+\alpha)/2}
\end{aligned}
$$

for some constant C'.

Appendix-B

For the experiments in Section 4.1, we have used four problem instances, namely I-P, I-N, I-W, and I-S. For each instance, the mean reward is a Lipschitz-continuous function over the set of arms, which is $[0,1]$. Figure 2 presents a visualisation; the precise mathematical specifications are provided below.

I-P (Parabolic): The mean function is a segment of a parabola with $\mu_{(x)} = 1$ for $x = 0.5$, and $\mu_{(x)} = 0$ for $x \in \{0,1\}$. Precisely, $\mu_{(x)} = 1 - 4(x-0.5)^2$.

I-N (Notch): The mean function has value 0.5 everywhere except in the interval $[0.25, 0.45]$, where it forms a notch and attains the value 1 at 0.35. Precisely,

$$\mu_{(x)} = \begin{cases} 0.5 \text{ if } |x - 0.35| > 0.1 \\ 1 - 5 \cdot |x - 0.35| \text{ otherwise.} \end{cases}$$

I-W (Wave): The mean function is a smooth approximation of a rectangular wave form. The interval $[0, 1]$ is divided into ten equal sub-intervals, each of length 0.1. Let $[a, b]$ be a sub-interval, $\epsilon = 0.01$, and $f(x, c, a, b) = 6y^5 - 15y^4 + 10y^3$ where $y = (x - c)/(b - a)$. Here $f(\cdot)$ is a SMOOTHSTEP function (Ebert et al., 2002). Then $\mu_{(x)}$ on each sub-interval $[a, b]$ is given by

$$\mu_{(x)} = \begin{cases} 0.5 \text{ if } x \in [a, a + 2\epsilon] \cup [a + 8\epsilon, b] \\ 0.5 + 0.5f(x, a, a, b) \text{ if } x \in [a + 2\epsilon, a + 3\epsilon] \\ 0.5 + 0.5f(x, b, a, b) \text{ if } x \in [a + 7\epsilon, a + 8\epsilon] \\ 1 \text{ otherwise.} \end{cases}$$

I-S (Smooth Step): The mean is a sigmoid function whose both ends are flat. Borrowing the definition of $f(\cdot)$ from instance I-W,

$$\mu_{(x)} = \begin{cases} 0.5 \text{ if } x \in [0, 0.4] \\ 1 \text{ if } x \in [0.6, 1] \\ f(x, 0.4, 0.4, 0.6) \text{ otherwise.} \end{cases}$$