Appendices

A  Detailed Proofs of Theorem 1

A.1  Proof of Lemma 1

Proof of Lemma 1 Define that $R(s) = \sum_{t=1}^{s} \max_{k \in K} X_{k,t} - X_{A_t,t}$, then we have that $R(T) = \mathbb{E}[R(T)]$.

$$R(T) = \mathbb{E}[R(T)] = \mathbb{E}[R(T)I\{\tau_1 \leq T\}] + \mathbb{E}[R(T)I\{\tau_1 > T\}] \leq T \cdot \mathbb{P}(\tau_1 \leq T) + \mathbb{E}[R(T)I\{\tau_1 > T\}] \tag{10}$$

Define $N_k(t)$ as the number of times arm $k$ has been selected by the Algorithm 2 in the first $t$ steps, i.e., $N_k(t) = \sum_{i=1}^{t} I(A_i = k)$. No false alarm is raised and we do not restart the UCB algorithm if the event $\{\tau_1 > T\}$ happens. Therefore, we have the following equation:

$$\mathbb{E}[R(T)I\{\tau_1 > T\}] = \sum_{\Delta^{(1)}_k > 0} \Delta^{(1)}_k \cdot \mathbb{E}[N_k(T)I\{\tau_1 > T\}] \tag{11}$$

Thus, it remains to show an upper bound for $\mathbb{E}[N_k(T)I\{\tau_1 > T\}]$. By the definition of Algorithm 2 we have that for any $k \in K$$

$$N_k(T)I\{\tau_1 > T\} = \sum_{t=1}^{T} I\{A_t = k, \tau_1 > T, N_k(t) < l\} + \sum_{t=1}^{T} I\{A_t = k, \tau_1 > T, N_k(t) \geq l\} \leq l + \sum_{t=1}^{T} I\{t \mod \lfloor K/\gamma \rfloor = k, N_k(t) \geq l\} + \sum_{t=1}^{T} I\{k = \arg\max_{k \in K} \text{UCB}_k, N_k(t) \geq l\} \leq l + \lfloor T \gamma/K \rfloor + \sum_{t=1}^{T} I\{k = \arg\max_{k \in K} \text{UCB}_k, N_k(t) \geq l\},$$

where the first inequality is due to the fact that if the event $\{A_t = k, \tau_1 > T\}$ happens, then we do not restart the UCB algorithm before time $T$ and the selection of the $k$th arm is based on either the uniform sampling or the largest UCB index in a stochastic bandit setting. Setting $l = \lceil 8 \log T/(\Delta^{(1)}_k)^2 \rceil$ and following the same argument as in the proof of Theorem 1 of [Auer et al., 2002a], we have that

$$\mathbb{E}[N_k(T)I\{\tau_1 > T\}] \leq \frac{T \gamma}{K} + \frac{8 \log T}{(\Delta^{(1)}_k)^2} + 1 + \frac{\pi^2}{3} + K.$$ Summing over $k \in K$ we prove the result.

A.2  Proof of Lemma 2

Proof of Lemma 2 Define $\tau_{k,1}$ as the first detection time of the $k$th arm. Then, $\tau_1 = \min_{k \in K} \{\tau_{k,1}\}$ since Algorithm 2 is designed to reinitialize the UCB algorithm if a change is detected on any of the $K$ arms. Using the union bound, we have
Nearly Optimal Adaptive Procedure with Change Detection for Piecewise-Stationary Bandit

that

\[ P(\tau_1 \leq T) \leq \sum_{k=1}^{K} P(\tau_{k,1} \leq T). \]

Define that for any \( k \in \mathcal{K} \) and \( t \geq w \)

\[ S_{k,t} = \left| \sum_{i=t-w/2+1}^{t} Z_{k,i} - \sum_{i=t-w+1}^{t} Z_{k,i} \right|. \tag{12} \]

Then, for any \( k \in \mathcal{K} \), \( \tau_{k,1} \) is given by

\[ \tau_{k,1} = \inf \{ t \geq w : S_{k,t} > b \} \]

Let \( \mathbb{Z}^+ \) be the set of all positive integers. Define that for any \( 0 \leq j \leq w - 1 \) the stopping times

\[ \tau_{k,1}^{(j)} = \inf \{ t = j + nw, n \in \mathbb{Z}^+ : S_{k,t} > b \} \].

We have that \( \tau_{k,1} = \min \{ \tau_{1}^{(0)}, \ldots, \tau_{1}^{(w-1)} \} \). Note that under the stationary environment, for any \( 0 \leq j \leq w - 1 \), \( \tau_{k,1}^{(j)} \) is a random variable with the geometric distribution

\[ P(\tau_{k,1}^{(j)} = nw + j) = p(1 - p)^{n-1}, \]

where \( p = P(S_{k,w} > b) \). Therefore, considering union bound we have that for any \( k \in \mathcal{K} \)

\[ P(\tau_{k,1} \leq T) \leq w \left( 1 - (1 - p)^{\lceil T/w \rceil} \right). \]

The remaining task is to find an upper bound for \( p \). Note that for any \( k \in \mathcal{K} \), \( S_{k,w} \) is a random variable with zero mean. We have by the McDiarmid’s inequality and the union bound that

\[ p \leq 2 \cdot \exp \left( -\frac{2b^2}{w} \right). \]

Combining the above analysis we conclude the result. \( \square \)

A.3 Proof of Lemma 3

**Proof of Lemma 3** Assume that \( \delta_{\hat{k}}^{(1)} \geq 2b/w + c \) for some \( \hat{k} \in \mathcal{K} \). Since the uniformly sampling scheme (step 2-4 of Algorithm 2) guarantees that in any time interval with length larger than \( L/2 \) each arm is sampled at least \( w/2 \) times, conditioning on \( \{ \tau_1 > \nu_1 \} \), we have that

\[
P(\nu_1 < \tau_1 \leq \nu_1 + L/2 \mid \tau_1 > \nu_1) \\
\geq P(S_{\hat{k},w} > b) \\
\geq 1 - 2 \exp \left( -\frac{(w|\delta_{\hat{k}}^{(1)}/2 - b|^2)}{w} \right) \\
\geq 1 - 2 \exp \left( -\frac{wc^2}{4} \right), \tag{13}
\]

where \( S_{k,t} \) is defined in (12) and we use McDiarmid’s inequality in the second inequality. \( \square \)
A.4 Proof of Lemma 4

Proof of Lemma 4 First, define that $N = \lceil b/\delta_k^{(1)} \rceil \cdot \lceil K/\gamma \rceil$, we obtain a simple upper bound for the EDD as follows.

\[
\begin{align*}
\mathbb{E}[\tau_1 - \nu_1 | \nu_1 < \tau_1 & \leq \nu_1 + L/2] \\
= & \sum_{i=1}^{L/2} \mathbb{P}(\tau_1 \geq \nu_1 + i | \nu_1 < \tau_1 \leq \nu_1 + L/2) \\
& \leq N + \sum_{i=N}^{L/2} (\mathbb{P}(\tau_1 \geq \nu_1 + i | \nu_1 < \tau_1 \leq \nu_1 + L/2)).
\end{align*}
\]

Since the uniformly sampling scheme guarantees that we have at least $i/\lceil K/\gamma \rceil$ samples from each arm within $i$ time steps, we use McDiarmid’s inequality and Lemma 3 to have that

\[
\sum_{i=N}^{L/2} (\mathbb{P}(\tau_1 \geq \nu_1 + i | \nu_1 < \tau_1 \leq \nu_1 + L/2)) \leq \frac{N}{1 - 2 \exp(-w^2/4)} \cdot \sum_{i=N}^{L/2} 2 \exp\left( -\frac{(i/\lceil (K/\gamma) \rceil \delta_k^{(1)} - b)^2}{w} \right) \leq \frac{\lceil K/\gamma \rceil}{1 - 2 \exp(-w^2/4)} \cdot \sum_{j=[b/\delta_k^{(1)}]}^{w/2} 2 \exp\left( -\frac{(j \delta_k^{(1)} - b)^2}{w} \right).
\]

Define $q = \lceil (w/2) \cdot \delta_k^{(1)} \rceil - b$ and we have $q > 1$ from the assumption that $\delta_k^{(1)} > 2b/w + c$. Combining the above analysis, we have that

\[
\begin{align*}
& (1 - 2 \exp(-w^2/4)) \cdot \mathbb{E}[\tau_1 - \nu_1 | \nu_1 < \tau_1 \leq \nu_1 + L/2] \\
\leq & N \cdot \lceil K/\gamma \rceil \cdot \sum_{j=[b/\delta_k^{(1)}]}^{w/2} 2 \exp\left( -\frac{(j \delta_k^{(1)} - b)^2}{w} \right) \\
\leq & N \cdot 2 \lceil K/\gamma \rceil \cdot \left( 1 + \int_{1}^{q} \exp\left( -\frac{t^2}{w} \right) dt \right) \\
\leq & N \cdot 2 \lceil K/\gamma \rceil \cdot \left( \frac{\sqrt{\frac{1}{w}}}{1} - \frac{1}{\sqrt{w}} + \int_{1}^{q/\sqrt{w}} \exp(-u^2) du \right) \\
\leq & N \cdot 2 \lceil K/\gamma \rceil \cdot \left( \sqrt{\frac{\sqrt{w}}{w}} + \int_{1}^{q/\sqrt{w}} u \exp(-u^2) du \right) \\
\leq & \left( \lceil b/\delta_k^{(1)} \rceil + 3\sqrt{w} \right) \cdot \lceil K/\gamma \rceil,
\end{align*}
\]

where we transform $l$ into $u = l/\sqrt{w}$ in the third inequality and we use the fact that $\exp(-u^2) \leq u \exp(-u^2)$, $u \geq 1$ in the fourth inequality. On the other hand, by the definition of the conditioning event we also have that

\[
(1 - 2 \exp(-w^2/4)) \cdot \mathbb{E}[\tau_1 - \nu_1 | \nu_1 < \tau_1 \leq \nu_1 + L/2] \leq L/2.
\]

Combining the above analysis we conclude the result.
A.5 Proof of Theorem 1

Proof of Theorem 1: Recall that $L = w \lceil K/\gamma \rceil$. Algorithm 2 guarantees that in any time interval with length larger than $L$ each arm is sampled at least $w$ times. Define events $F_i = \{ \tau_i > \nu_i \}, 1 \leq i \leq M - 1$. Define events $D_i = \{ \tau_i \leq \nu_i + L/2 \}, 1 \leq i \leq M - 2$ and event $D_{M-1} = \{ \tau_{M-1} \leq T \}$. Therefore, the event $F_i D_i$ is the good event where the $i$th change can be detected correctly and efficiently. Define that $R(s) = \sum_{i=1}^s \max_{k \in K} X_{k,t} - X_{A_i,t}$, then we have that $\mathcal{R}(T) = \mathbb{E}[R(T)]$. Equipped with the sequence of good events, we have that

\[
\mathcal{R}(T) = \mathbb{E}[R(T)] \leq \mathbb{E}[R(T|F_1)] + T \cdot (1 - \mathbb{P}(F_1)) \leq \mathbb{E}[R(\nu_1|F_1)] + \mathbb{E}[R(T|D_1)] + 1 \leq \hat{C}_1 + \gamma \nu_1 + \mathbb{E}[R(T) - R(\nu_1)] + 1.
\]

Above, the second inequality is due to Lemma 2 that $\mathbb{P}(F_1) \geq 1 - 1/T$ provided that $b = [w \log(2KT^2)/2]^{1/2}$ and the third inequality is due to the bound in the end of Lemma 1, which is the bound for the UCB algorithm in a stochastic bandit setting.

The next step is to bound $\mathbb{E}[R(T) - R(\nu_1)]$. Using the law of total expectation, we have that

\[
\mathbb{E}[R(T) - R(\nu_1)] \leq \mathbb{E}[R(T) - R(\nu_1) | F_1 D_1] + T \cdot (1 - \mathbb{P}(F_1 D_1)) \leq \mathbb{E}[R(T) - R(\nu_1) | F_1 D_1] + 2,
\]

where the last inequality is due to Lemma 3 that we have $\mathbb{P}(D_1 | F_1) \geq 1 - 1/T$ provided that $c = 2\sqrt{\log(2T)}/w$ and the fact that $\mathbb{P}(F_1 D_1) = \mathbb{P}(D_1 | F_1) \cdot \mathbb{P}(F_1)$ for any probability measure $\mathbb{P}$.

Therefore, the remaining task is to bound $\mathbb{E}[R(T) - R(\nu_1) | F_1 D_1]$. Denote $\hat{\mathbb{E}}$ as the expectation according to the piecewise-stationary bandit starts from the second segment. Further splitting the regret, we have that

\[
\mathbb{E}[R(T) - R(\nu_1) | F_1 D_1] \leq \mathbb{E}[R(T) - R(\tau_1) | F_1 D_1] + \mathbb{E}[R(\tau_1) - R(\nu_1) | F_1 D_1] \leq \hat{\mathbb{E}}[R(T - \nu_1)] + \mathbb{E}[\tau_1 - \nu_1 | F_1 D_1] \leq \hat{\mathbb{E}}[R(T - \nu_1)] + \min(L/2, ([b/\delta(1)] + 3\sqrt{w}) \cdot [K/\gamma])/(1 - 1/T)
\]

where the second inequality is due to the renewal property given that the whole algorithm restarts in the time interval between $\nu_1$ and $\nu_1 + L/2$ and the last inequality is due to Lemma 4 by setting $c = 2\sqrt{\log(2T)}/w$.

Combining the above analysis, we bound the regret in a recursive manner as follows (assuming $T \geq 2$):

\[
\mathbb{E}[R(T)] \leq \hat{\mathbb{E}}[R(T - \nu_1)] + \hat{C}_1 + \gamma \nu_1 + 2 \min(L/2, ([b/\delta(1)] + 3\sqrt{w}) \cdot [K/\gamma]) + 3.
\]

The recursive manner means that we can apply the same method to bound $\hat{\mathbb{E}}[R(T - \nu_1)]$, by conditioning on the event $D_2 F_2$. Repeating this procedure $M - 1$ times, we obtain that

\[
\mathbb{E}[R(T)] \leq \sum_{i=1}^M \hat{C}_i + \gamma T + \sum_{i=1}^{M-1} \frac{2K \cdot \min([b/\delta(1)] + 3\sqrt{w})}{\gamma} + 3M.
\]