Combinatorial Semi-Bandit in the Non-Stationary Environment Supplementary Material

Wei Chen[1,*] **Liwei Wang**[2] **Haoyu Zhao**[3] **Kai Zheng**[4]

[1]Microsoft Research, Beijing, China. weic@microsoft.com
[2]Key Laboratory of Machine Perception, MOE, School of EECS,
Center for Data Science, Peking University, Beijing, China. wanglw@cis.pku.edu.cn
[3] Princeton University, NJ, USA. haoyu@princeton.edu
[4]Kuaishou Inc., Beijing, China. zhengk92@gmail.com
[*]Alphabetic order

APPENDIX

1 OMITTED PROOFS IN SECTION 3

In this section, we give the performance guarantees of our algorithm CUCB-SW and CUCB-BoB in the general case. We first give some definitions and prove some basic lemmas in the first part. Then, as a warm up, we prove the corresponding result of Theorem 1 in main content without the probabilistically triggered arms (Theorem 1 in appendix). Next, we prove Theorem 1 in main content with probabilistically triggered arms (Theorem 2 in appendix). Finally, we prove Theorem 2 in main content (Theorem 3 in appendix), which applies the Bandit-over-Bandit technique to achieve parameter-free.

1.1 FUNDAMENTAL DEFINITIONS AND TOOLS

First, we define the event-filtered regret. Generally speaking, it is the regret when some event happens.

Definition 1 (Event-Filtered Regret)**.** *For any series of events $\{\mathcal{E}_t\}_{t\geq 1}$ indexed by round number t, we define $Reg_\alpha^{\mathcal{A}}(T, \{\mathcal{E}_t\}_{t\geq 1})$ as the regret filtered by events $\{\mathcal{E}_t\}_{t\geq 1}$, that is, regret is only counted in round t if $\mathcal{E}_t$ happens in round t. Formally,*

$$Reg_\alpha^{\mathcal{A}}(T, \{\mathcal{E}_t\}_{t\geq 1}) = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}\{\mathcal{E}_t\}(\alpha \cdot opt_{\boldsymbol{\mu}_t} - r_{\boldsymbol{\mu}_t}(S_t^{\mathcal{A}})\right].$$

For convenience, $\mathcal{A}$, α, or T can be omitted when the context is clear, and we simply use $Reg_\alpha^{\mathcal{A}}(T, \mathcal{E}_t)$ instead of $Reg_\alpha^{\mathcal{A}}(T, \{\mathcal{E}_t\}_{t\geq 1})$.

Then, we define two important events that will use in the event-filtered regret. The two events are Sampling is Nice (Definition 2 and Triggering is Nice (Definition 5. We will also show that these two events happen with high probability. The following propositions, definitions, and lemmas are all related with these two definitions.

Proposition 1 (Hoeffding Inequality)**.** *Suppose $X_i \in [0,1]$ for all $i \in [n]$ and X_i are independent, then we have*

$$\Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]\right| \geq \varepsilon\right\} \leq 2\exp\left(-2n\varepsilon^2\right).$$

Definition 2 (Sampling is Nice)**.** *We say that the sampling is nice at the beginning of round t if for any arm $i \in [m]$, we have $|\hat{\mu}_{i,t} - \nu_{i,t}| < \rho_{i,t}$, where $\rho_{i,t} = \sqrt{\frac{3\ln T}{2T_{i,t}}}$ (∞ if $T_{i,t} = 0$) and $\hat{\mu}_{i,t}$ are defined in the algorithm, and*

$$\nu_{i,t} = \frac{1}{T_{i,t}} \sum_{s=t-w+1}^{t-1} \mathbb{I}\left\{i \text{ is triggered at time } s\right\} \mu_{i,t}.$$

If i is not triggered during time $(t-w, t-1]$, we define $\nu_{i,t} = \mu_{i,t}$. We use $\mathcal{N}_t^s$ to denote this event.

We have the following lemma saying that $\mathcal{N}_t^s$ is a high probability event.

Lemma 1. *For each round $t \geq 1$, $\Pr\{\neg\mathcal{N}_t^s\} \leq 2mT^{-2}$.*

Proof. The proof is a direct application of Hoeffding inequality and a union bound. First when $T_{i,t} = 0$, we have $\rho_{i,t} = \infty$ and the event $\mathcal{N}_t^s$ happens. We first have

$$\begin{aligned}
\Pr\{\neg\mathcal{N}_t^s\} =& \Pr\{\exists i \in [m], |\hat{\mu}_{i,t} - \nu_{i,t}| \geq \rho_{i,t}\} \\
\leq& \sum_{i=1}^{m} \Pr\{|\hat{\mu}_{i,t} - \nu_{i,t}| \geq \rho_{i,t}\} \\
=& \sum_{i=1}^{m} \Pr\left\{|\hat{\mu}_{i,t} - \nu_{i,t}| \geq \sqrt{\frac{3\ln T}{2T_{i,t}}}\right\} \\
=& \sum_{i=1}^{m}\sum_{k=1}^{\Gamma_t} \Pr\left\{T_{i,t} = k, |\hat{\mu}_{i,t} - \nu_{i,t}| \geq \sqrt{\frac{3\ln T}{2T_{i,t}}}\right\}.
\end{aligned}$$

Then, by the conditional probability and the Hoeffding inequality, we have

$$\begin{aligned}
&\Pr\left\{T_{i,t} = k, |\hat{\mu}_{i,t} - \nu_{i,t}| \geq \sqrt{\frac{3\ln T}{2T_{i,t}}}\right\} \\
=& \Pr\{T_{i,t} = k\} \Pr\left\{|\hat{\mu}_{i,t} - \nu_{i,t}| \geq \sqrt{\frac{3\ln T}{2T_{i,t}}} \middle| T_{i,t} = k\right\} \\
\leq& \Pr\{T_{i,t} = k\} 2\exp\left(-2k\frac{3\ln T}{2k}\right) \\
\leq& 2\exp\left(-2k\frac{3\ln T}{2k}\right) \\
=& \frac{2}{T^3}.
\end{aligned}$$

Then we know that

$$\begin{aligned}
\Pr\{\neg\mathcal{N}_t^s\} \leq& \sum_{i=1}^{m}\sum_{k=1}^{\Gamma_t} \Pr\left\{T_{i,t} = k, |\hat{\mu}_{i,t} - \nu_{i,t}| \geq \sqrt{\frac{3\ln T}{2T_{i,t}}}\right\} \\
\leq& \sum_{i=1}^{m}\sum_{k=1}^{\Gamma_t} \frac{2}{T^3} \\
\leq& \sum_{i=1}^{m}\sum_{k=1}^{t} \frac{2}{T^3} \\
=& 2mT^{-2}.
\end{aligned}$$

□

Proposition 2 (Multiplicative Chernoff Bound)**.** *Suppose X_i are Bernoulli variables for all $i \in [n]$ and $\mathbb{E}[X_i|X_1, \ldots, X_{i-1}] \geq \mu$ for every $i \leq n$. Let $Y = X_1 + \cdots + X_n$, then we have*

$$\Pr\{Y \leq (1-\delta)n\mu\} \leq \exp\left(-\frac{\delta^2 n\mu}{2}\right).$$

Definition 3 (Triggering Probability (TP) Group)**.** *Let i be an arm and j be a positive natural number, define the triggering probability group (of actions)*

$$G_{i,j} = \{S^D \in \mathbb{S} \times \mathbb{D} | 2^{-j} < p_i^{D,S} \leq 2^{-j+1}\}.$$

Definition 4 (Main content definition 3 restated)**.** *Given the sliding window size w of the algorithm, in a run of the algorithm, we define the counter $N_{i,j,t}$ as the following number*

$$N_{i,j,t} := \sum_{s=\max\{t-w+1,0\}}^{t} \mathbb{I}\left\{2^{-j} < p_i^{D_s,S_s} \leq 2^{-j+1}\right\}.$$

Definition 5 (Triggering is Nice)**.** *Given integers $\{j_{\max}^i\}_{i\in[m]}$, we call that the triggering is nice at the beginning of round t if for any arm i and any $1 \leq j \leq j_{\max}^i$, as long as $6\ln t \leq \frac{1}{3}N_{i,j,t-1}\cdot 2^{-j}$, we have*

$$T_{i,t-1} \geq \frac{1}{3}N_{i,j,t-1}\cdot 2^{-j}.$$

We use $\mathcal{N}_t^t$ to denote this event.

Lemma 2. *Given a series of integers $\{j_{\max}^i\}_{i\in[m]}$, we have for every round $t \geq 1$,*

$$\Pr\{\neg\mathcal{N}_t^t\} \leq \sum_{i\in[m]} j_{\max}^i t^{-2}.$$

This lemma is exactly the same as Lemma 4 in Wang and Chen [2017]. The proof is a direct application of the Multiplicative Chernoff Bound. We omit the proof here.

Finally, we extend the definition of gap for the ease of the analysis. First recall that we have the following definition of gap.

Definition 6 (Main content definition 2 restated)**.** *For any distribution D with mean vector $\boldsymbol{\mu}$. For each action S, we define the gap $\Delta_S^D := \max\{0, \alpha\cdot opt_{\boldsymbol{\mu}} - r_S(\boldsymbol{\mu})\}$. For each arm i, we define*

$$\Delta_{\min}^{i,t} = \inf_{S\in\mathbb{S}: p_i^{D_t,S}>0, \Delta_S^{D_t}>0} \Delta_S^{D_t},$$
$$\Delta_{\max}^{i,t} = \sup_{S\in\mathbb{S}: p_i^{D_t,S}>0, \Delta_S^{D_t}>0} \Delta_S^{D_t}.$$

We define $\Delta_{\min}^i = +\infty$ and $\Delta_{\max}^i = 0$ if they are not properly defined by the above definitions. Furthermore, we define $\Delta_{\min}^i := \min_{t\leq T}\Delta_{\min}^{i,t}$, $\Delta_{\max}^i := \max_{t\leq T}\Delta_{\max}^{i,t}$ as the minimum and maximum gap for each arm.

The previous definition of gap focus on a single distribution and a single arms. Furthermore, we define $\Delta_{\min}^t := \inf_{i\in[m]}\Delta_{\min}^{i,t}$, $\Delta_{\max}^t := \sup_{i\in[m]}\Delta_{\max}^{i,t}$ as the minimum and maximum gap in each round, and $\Delta_{\min} := \inf_{t\leq T}\Delta_{\min}^t, \Delta_{\max} := \sup_{t\leq T}\Delta_{\max}^t$ as the minimum and maximum gap.

1.2 NON-STATIONARY CMAB WITHOUT PROBABILISTICALLY TRIGGERED ARMS

As a warm up, we first consider the case without the probabilistically triggered arms, i.e. $p_i^{D,S} \in \{0,1\}$. Then $\tilde{S}^D = S$ and we denote $K = \max_S |S|$. Then, the TPM bounded smoothness becomes the following,

Assumption 1 (1-Norm Bounded Smoothness)**.** *For any two distributions D, D' with expectation vectors $\boldsymbol{\mu}$ and $\boldsymbol{\mu}'$ and any action S, we have*

$$|r_S(\boldsymbol{\mu}) - r_S(\boldsymbol{\mu}')| \leq B\sum_{i\in S}|\boldsymbol{\mu}_i - \boldsymbol{\mu}_i'|.$$

We define the following number:

$$\kappa_T(M,s) = \begin{cases} 2B\sqrt{6\ln T}, & \text{if } s = 0, \\ 2B\sqrt{\dfrac{6\ln T}{s}}, & \text{if } 1 \leq s \leq \ell_T(M), \\ 0, & \text{if } s \geq \ell_T(M)+1, \end{cases}$$

where

$$\ell_T(M) = \left\lfloor \frac{24B^2K^2 \ln T}{M^2} \right\rfloor.$$

Generally speaking, we bridge the regret and the upper bound by this number, and we use the technique similar to that in Wang and Chen [2017].

Lemma 3. *Suppose that the sliding window size is w. For any arm $i \in [m]$, any T, and any numbers $\{M_i\}_{i \le m}$,*

$$\sum_{t=1}^{T} \mathbb{I}(i \in S_t) \cdot \kappa_T(M_i, T_{i,t}) \le \left(\frac{T}{w} + 1\right)\left(2B\sqrt{6 \ln T} + \frac{48B^2K \ln T}{M_i}\right).$$

Proof. We devide the time $\{1, 2, \ldots, T\}$ into the following Γ segments $[1 = t_0 + 1, w = t_1], [w + 1 = t_1 + 1, 2w = t_2], \ldots, [t_{\Gamma-1} + 1, t_\Gamma = T]$, where $t_{j-1} = t_j - w$. Each segment has length w, except for the last segment. It is easy to show that $\Gamma \le \lceil \frac{T}{w} \rceil$.

Then we bound $\sum_{t=1}^{T} \mathbb{I}(i \in S_t) \cdot \kappa_T(M_i, T_{i,t})$. We first define another variable $T'_{i,t}$ for every i, t. Suppose that $t_{j-1} < t \le t_j$, which means that t lies in the jth time segment, let $T'_{i,t}$ denote the number of times arm i has been triggered in time $[t_{j-1} + 1, t - 1]$.

Then we know that $T_{i,t} \ge T'_{i,t}$, since the counter $T'_{i,t}$ counts the triggered times in a time interval which is a subset of the time interval for $T_{i,t}$. Because $\kappa_T(M, s)$ is decreasing when s is increasing, we know that

$$\sum_{t=1}^{T} \mathbb{I}(i \in S_t) \cdot \kappa_T(M_i, T_{i,t}) \le \sum_{t=1}^{T} \mathbb{I}(i \in S_t) \cdot \kappa_T(M_i, T'_{i,t})$$

Then we bound the right hand side, and we have

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{I}(i \in S_t) \cdot \kappa_T(M_i, T'_{i,t}) &= \sum_{j=1}^{\Gamma} \sum_{t=t_{j-1}+1}^{t_j} \mathbb{I}(i \in S_t) \cdot \kappa_T(M_i, T'_{i,t}) \\
&\le \sum_{j=1}^{\Gamma} \sum_{s=0}^{w-1} \kappa_T(M_i, s) \\
&\le \sum_{j=1}^{\Gamma} \left(2B\sqrt{6 \ln T} + \sum_{s=1}^{\ell_T(M_i)} \kappa_T(M_i, s)\right) \\
&= \sum_{j=1}^{\Gamma} \left(2B\sqrt{6 \ln T} + \sum_{s=1}^{\ell_T(M_i)} 2B\sqrt{\frac{6 \ln T}{s}}\right) \\
&\le \sum_{j=1}^{\Gamma} \left(2B\sqrt{6 \ln T} + \int_0^{\ell_T(M_i)} 2B\sqrt{\frac{6 \ln T}{s}} ds\right) \\
&\le \sum_{j=1}^{\Gamma} \left(2B\sqrt{6 \ln T} + 4B\sqrt{6 \ln T \ell_T(M_i)}\right) \\
&\le \sum_{j=1}^{\Gamma} \left(2B\sqrt{6 \ln T} + 4B\sqrt{6 \ln T \frac{24B^2K^2 \ln T}{M_i^2}}\right) \\
&\le \left(\frac{T}{w} + 1\right)\left(2B\sqrt{6 \ln T} + \frac{48B^2K \ln T}{M_i}\right).
\end{aligned}$$

□

Then, we have the following simple lemma to bound the difference between the true mean of each round and the actual mean for the round that we trigger. The lemma is simple to proof, and a detailed proof can be found in Zhao and Chen [2019].

Lemma 4. *Suppose that the size of the sliding window is w. For every t and every possible triggering, we have*

$$||\nu_t - \mu_t||_\infty \leq \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty.$$

Denote Δ_S^t as $\Delta_S^{\mathcal{D}_t}$ for simplicity. At round t with action S_t, we use Δ_{S_t} for short.

Lemma 5. *Suppose that the size of the sliding window is w and fix the parameters M_i for each $i \in [m]$ and defining $M_{S_t} = \max_{i \in S_t} M_i$. Then we have*

$$\textit{Reg}(\{\Delta_{S_t}^t \geq M_{S_t}\} \wedge \mathcal{N}_t^s \wedge \neg \mathcal{F}_t) \leq \sum_{i \in [m]} \left(\frac{T}{w} + 1\right) \left(2B\sqrt{6 \ln T} + \frac{48B^2 K \ln T}{M_i}\right) + 2(1+\alpha)KB \sum_{s=2}^{t} ||\mu_s - \mu_{s-1}||_\infty \cdot w.$$

where $\mathcal{F}_t$ is denoted as the event that $\{r_{S_t}(\bar{\mu}_t) < \alpha \cdot opt_{\bar{\mu}_t}\}$

Proof. From the assumption of our oracle, we know that $\Pr\{\mathcal{F}_t\} \leq 1 - \beta$. We also define $M_S = \max_{i \in \bar{S}} M_i$ for each possible action S, and use define $M_S = 0$ if $\bar{S} = \phi$. We first show that when $\{\Delta_{S_t}^t \geq M_{S_t}\}, \mathcal{N}_t^s, \neg \mathcal{F}_t$ all happens, we have

$$\Delta_{S_t}^t \leq \sum_{i \in \bar{S}_t} \kappa_T(M_i, T_{i,t-1}) + 2(1+\alpha)KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty.$$

First when $\Delta_{S_t}^t = 0$, the inequality holds, and we just have to prove the case when $\Delta_{S_t}^t > 0$. Let R_1 denote the optimal strategy when the mean vector is μ_t' in which the i-th entry is $\mu_{i,t}' = \min\{\nu_{i,t} + \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty, 1\}$. Then we know that $\mu_{i,t}' \geq \mu_{i,t}$. From $\mathcal{N}_t^s$ and $\neg \mathcal{F}_t$, we have

$$\begin{aligned}
r_{S_t}(\bar{\mu}_t) \geq & \alpha \cdot \text{opt}_{\bar{\mu}_t} \geq \alpha \cdot r_{R_1}(\bar{\mu}_t) \geq \alpha \cdot r_{R_1}(\nu_t) \\
\geq & \alpha \cdot r_{R_1}(\mu_t') - \alpha KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty \\
\geq & \alpha \cdot \text{opt}_{\mu_t} - \alpha KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty \\
= & r_{S_t}(\mu_t) + \Delta_{S_t}^t - \alpha KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty \\
\geq & r_{S_t}(\nu_t) + \Delta_{S_t}^t - (1+\alpha)KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty,
\end{aligned}$$

so we get

$$\begin{aligned}
\Delta_{S_t} \leq & r_{S_t}(\bar{\mu}_t) - r_{S_t}(\nu_t) + (1+\alpha)KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty \\
\leq & B \sum_{i \in S_t} (\bar{\mu}_{i,t} - \nu_{i,t}) + (1+\alpha)KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty.
\end{aligned}$$

Then when $\{\Delta^t_{S_t} \geq M_{S_t}\}, \mathcal{N}^s_t, \neg\mathcal{F}_t$ all happens, we have

$$\begin{aligned}
\Delta^t_{S_t} \leq & B\sum_{i\in S_t}(\bar{\mu}_{i,t} - \nu_{i,t}) + (1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s - \mu_{s-1}||_\infty \\
\leq & -M_{S_t} + 2B\sum_{i\in S_t}(\bar{\mu}_{i,t} - \nu_{i,t}) + 2(1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s - \mu_{s-1}||_\infty \\
\leq & 2B\sum_{i\in S_t}\left(\bar{\mu}_{i,t} - \nu_{i,t} - \frac{M_{S_t}}{2B|\bar{S}_t|}\right) + 2(1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s - \mu_{s-1}||_\infty \\
\leq & 2B\sum_{i\in S_t}\left(\bar{\mu}_{i,t} - \nu_{i,t} - \frac{M_{S_t}}{2BK}\right) + 2(1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s - \mu_{s-1}||_\infty \\
\leq & 2B\sum_{i\in S_t}\left(\bar{\mu}_{i,t} - \nu_{i,t} - \frac{M_i}{2BK}\right) + 2(1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s - \mu_{s-1}||_\infty.
\end{aligned}$$

By the same proof in Wang and Chen [2017], it can be shown that

$$2B\sum_{i\in S_t}\left(\bar{\mu}_{i,t} - \nu_{i,t} - \frac{M_i}{2BK}\right) \leq \sum_{i\in S_t}\kappa_T(M_i, T_{i,t-1}),$$

and thus we have

$$\Delta^t_{S_t} \leq \sum_{i\in S_t}\kappa_T(M_i, T_{i,t-1}) + 2(1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s - \mu_{s-1}||_\infty.$$

From the previous 2 lemmas, we know that

$$\text{Reg}(\{\Delta^t_{S_t} \geq M_{S_t}\}\wedge\mathcal{N}^s_t\wedge\neg\mathcal{F}_t) \leq \sum_{i\in[m]}\left(\frac{T}{w}+1\right)\left(2B\sqrt{6\ln T} + \frac{48B^2K\ln T}{M_i}\right) + 2(1+\alpha)KB\sum_{s=2}^{t}||\mu_s - \mu_{s-1}||_\infty \cdot w.$$

□

Theorem 1. *Choosing the length of the sliding window to be* $w = \min\left\{\sqrt{\frac{T}{V}}, T\right\}$*, we have the following distribution dependent bound,*

$$\textit{Reg}_{\alpha,\beta} = \tilde{O}\left(\sum_{i\in[m]}\frac{K\sqrt{VT}}{\Delta^i_{\min}} + \sum_{i\in[m]}\frac{K}{\Delta^i_{\min}} + mK\right).$$

If we choose the length of the sliding window to be $w = \min\left\{m^{1/3}T^{2/3}K^{-1/3}V^{-2/3}, T\right\}$*, we have the following distribution independent bound,*

$$\textit{Reg}_{\alpha,\beta} = \tilde{O}\left((mV)^{1/3}(KT)^{2/3} + \sqrt{mKT} + mK\right).$$

The proof is the same as the proof of Theorem 2, and we omit the proof here. The only difference is that, without the probabilistically triggered arms, the constants in Lemma 5 is better than the corresponding lemma with the probabilistically triggered arms.

1.3 NON-STATIONARY CMAB WITH PROBABILISTICALLY TRIGGERED ARMS

In this part, we consider the case with probabilistically triggered arms. Recall that the we have the main TPM bounded smoothness assumption,

Assumption 2 (Main content assumption 2 restated)**.** *For any two distributions* D, D' *with expectation vectors* $\boldsymbol{\mu}$ *and* $\boldsymbol{\mu}'$ *and any action* S*, we have*

$$|r_S(\boldsymbol{\mu}) - r_S(\boldsymbol{\mu}')| \leq B\sum_{i\in[m]}p^{D,S}_i|\boldsymbol{\mu}_i - \boldsymbol{\mu}'_i|.$$

Recall that $\tilde{S}^D = \{i \in [m] : p_i^{D,S} > 0\}$ is the set that can be triggered by action S with distribution D, and we denote $K = \max_{S_D} |\tilde{S}|$. We define the following number:

$$\kappa_{j,T}(M,s) = \begin{cases} 2B\sqrt{72 \cdot 2^{-j} \cdot \ln T}, & \text{if } s = 0, \\ 2B\sqrt{\dfrac{72 \cdot 2^{-j} \cdot \ln T}{s}}, & \text{if } 1 \le s \le \ell_{j,T}(M), \\ 0, & \text{if } s \ge \ell_{j,T}(M) + 1, \end{cases}$$

where

$$\ell_{j,T}(M) = \left\lfloor \frac{288 \cdot 2^{-j} \cdot B^2 K^2 \ln T}{M^2} \right\rfloor.$$

This number is similar to the number defined in the previous part, but this time, we need to consider the probabilistically triggered arms. Besides the M, s that are taken as inputs, we also have j and T as parameters.

Lemma 6. *If $\{\Delta_{S_t} \ge M_{S_t}\}, \neg\mathcal{F}_t, \mathcal{N}_t^s$ and $\mathcal{N}_t^t$ hold, we have*

$$\Delta_{S_t} \le \sum_{i \in \tilde{S}_t^{D_t}} \kappa_{j_i,T}(M_i, N_{i,j_i,t-1}) + 2(1+\alpha)KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty,$$

where j_i is the index of the TP group with $S_t^{D_t} \in G_{i,j_i}$.

Proof. First, similar to the proof with no probabilistic triggering arms, we use the back amortization trick.

First when $\Delta_{S_t} = 0$, the inequality holds, and we just have to prove the case when $\Delta_{S_t} > 0$. Let R_1 denote the optimal strategy when the mean vector is $\boldsymbol{\mu}_t'$, where $\boldsymbol{\mu}_t'$ is the vector constituted by $\mu_{i,t}' = \min\{\nu_{i,t} + \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty, 1\}$. Then we know that $\mu_{i,t}' \ge \mu_{i,t}$. From $\mathcal{N}_t^s$ and $\neg\mathcal{F}_t$, we have

$$\begin{aligned} r_{S_t}(\bar{\mu}_t) \ge& \alpha \cdot \text{opt}_{\bar{\mu}_t} \ge \alpha \cdot r_{R_1}(\bar{\mu}_t) \ge \alpha \cdot r_{R_1}(\nu_t) \\ \ge& \alpha \cdot r_{R_1}(\mu_t') - \alpha KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty \\ \ge& \alpha \cdot \text{opt}_{\mu_t} - \alpha KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty \\ =& r_{S_t}(\mu_t) + \Delta_{S_t} - \alpha KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty \\ \ge& r_{S_t}(\nu_t) + \Delta_{S_t} - (1+\alpha)KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty, \end{aligned}$$

so we get

$$\begin{aligned} \Delta_{S_t} \le& r_{S_t}(\bar{\mu}_t) - r_{S_t}(\nu_t) + (1+\alpha)KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty \\ \le& B \sum_{i \in \tilde{S}_t} p_i^{D_t,S_t}(\bar{\mu}_{i,t} - \nu_{i,t}) + (1+\alpha)KB \sum_{s=t-w+2}^{t} ||\mu_s - \mu_{s-1}||_\infty. \end{aligned}$$

Then when $\{\Delta_{S_t}^t \geq M_{S_t}\}, \mathcal{N}_t^s, \neg\mathcal{F}_t$ all happens, we have

$$\begin{aligned}
\Delta_{S_t} \leq & B\sum_{i\in\tilde{S}_t} p_i^{D_t,S_t}(\bar{\mu}_{i,t}-\nu_{i,t}) + (1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s-\mu_{s-1}||_\infty \\
\leq & -M_{S_t} + 2B\sum_{i\in\tilde{S}_t} p_i^{D_t,S_t}(\bar{\mu}_{i,t}-\nu_{i,t}) + 2(1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s-\mu_{s-1}||_\infty \\
\leq & 2B\sum_{i\in\tilde{S}_t} p_i^{D_t,S_t}\left(\bar{\mu}_{i,t}-\nu_{i,t}-\frac{M_{S_t}}{2B|\tilde{S}_t|}\right) + 2(1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s-\mu_{s-1}||_\infty \\
\leq & 2B\sum_{i\in\tilde{S}_t} p_i^{D_t,S_t}\left(\bar{\mu}_{i,t}-\nu_{i,t}-\frac{M_{S_t}}{2BK}\right) + 2(1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s-\mu_{s-1}||_\infty \\
\leq & 2B\sum_{i\in\tilde{S}_t} p_i^{D_t,S_t}\left(\bar{\mu}_{i,t}-\nu_{i,t}-\frac{M_i}{2BK}\right) + 2(1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s-\mu_{s-1}||_\infty.
\end{aligned}$$

Because of $\mathcal{N}_t^t$, same as the proof of Lemma 5 of Wang and Chen [2017], we can show that

$$2B\sum_{i\in\tilde{S}_t} p_i^{D_t,S_t}\left(\bar{\mu}_{i,t}-\nu_{i,t}-\frac{M_i}{2BK}\right) \leq \sum_{i\in(\tilde{S}_t)^{D_t}} \kappa_{j_i,T}(M_i, N_{i,j_i,t-1}).$$

In this way, we prove the following inequality

$$\Delta_{S_t} \leq \sum_{i\in(\tilde{S}_t)^{D_t}} \kappa_{j_i,T}(M_i, N_{i,j_i,t-1}) + 2(1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s-\mu_{s-1}||_\infty,$$

when $\{\Delta_{S_t} \geq M_{S_t}\}, \neg\mathcal{F}_t, \mathcal{N}_t^s$ and $\mathcal{N}_t^t$ hold. □

Then we have the following main lemma to bound the regret with probabilistically triggered arms.

Lemma 7. *Suppose that the size of the sliding window is w and fix choose the parameters M_i for each $i \in [m]$ and defining $M_{S_t} = \min_{i\in\hat{S}} M_i$. Then we have*

$$\begin{aligned}
& \textit{Reg}(\{\Delta_{S_t}^{D_t} \geq M_{S_t}\} \wedge \mathcal{N}_t^s \wedge \mathcal{N}_t^t \wedge \neg\mathcal{F}_t) \\
\leq & \sum_{i\in[m]} \left(\frac{T}{w}+1\right)\left(12(2+\sqrt{2})B\sqrt{\ln T} + \frac{576B^2K\ln T}{M_i}\right) + 2(1+\alpha)KB\sum_{s=2}^{t}||\mu_s-\mu_{s-1}||_\infty \cdot w.
\end{aligned}$$

Proof. From Lemma 6, we know that when $\{\Delta_{S_t}^{D_t} \geq M_{S_t}\}, \neg\mathcal{F}_t, \mathcal{N}_t^s$ and $\mathcal{N}_t^t$ hold, we have

$$\Delta_{S_t}^{D_t} \leq \sum_{i\in(\tilde{S}_t)^{D_t}} \kappa_{j_i,T}(M_i, N_{i,j_i,t-1}) + 2(1+\alpha)KB\sum_{s=t-w+2}^{t}||\mu_s-\mu_{s-1}||_\infty.$$

Then, sum over $t = 1, \ldots, T$, we have

$$\begin{aligned}
\text{Reg}(\{\Delta_{S_t}^{D_t} \geq M_{S_t}\} \wedge \mathcal{N}_t^s \wedge \mathcal{N}_t^t \wedge \neg\mathcal{F}_t) \leq & \sum_{t=1}^{T}\sum_{i\in(\tilde{S}_t)^{D_t}} \kappa_{j_i,T}(M_i, N_{i,j_i,t-1}) + 2(1+\alpha)KB\sum_{t=1}^{T}\sum_{s=t-w+2}^{t}||\mu_s-\mu_{s-1}||_\infty \\
\leq & \sum_{t=1}^{T}\sum_{i\in(\tilde{S}_t)^{D_t}} \kappa_{j_i,T}(M_i, N_{i,j_i,t-1}) + 2(1+\alpha)KB\sum_{s=2}^{t}||\mu_s-\mu_{s-1}||_\infty \cdot w.
\end{aligned}$$

Then we bound the first term. Like the proof without probabilistically triggered arms, we construct another counter $N'_{i,j,t-1}$, which lower bound $N_{i,j,t-1}$. We divide the time $\{1, 2, \ldots, T\}$ into the following Γ segments $[1 = t_0 + 1, w = t_1], [w + 1 =$

$t_1+1, 2w=t_2], \ldots, [t_{\Gamma-1}+1, t_\Gamma = T]$, where $t_{j-1} = t_j - w$. Each segment has length w, except for the last segment. It is easy to show that $\Gamma \leq \lceil \frac{T}{w} \rceil$. Suppose that $t_{k-1} < t \leq t_k$, then define

$$N'_{i,j,t} := \sum_{s=t_k+1}^{t} \mathbb{I}\left\{2^{-j} < p_i^{D_s,S_s} \leq 2^{-j+1}\right\}.$$

Because $\kappa_{j,T}(M,s)$ is monotonically decreasing in terms of s, we have

$$\begin{aligned}
&\sum_{t=1}^{T} \sum_{i\in(\tilde{S}_t)^{D_t}} \kappa_{j_i,T}(M_i, N_{i,j_i,t-1}) \\
\leq &\sum_{t=1}^{T} \sum_{i\in(\tilde{S}_t)^{D_t}} \kappa_{j_i,T}(M_i, N'_{i,j_i,t-1}) \\
\leq &\sum_{i\in[m]} \sum_{k=1}^{\Gamma} \sum_{j=1}^{+\infty} \sum_{s=t_{k-1}+1}^{t_k} \kappa_{j,T}(M_i, s-t_{k-1}-1) \\
\leq &\sum_{i\in[m]} \sum_{k=1}^{\Gamma} \sum_{j=1}^{+\infty} \sum_{s=0}^{\ell_{j,T}(M_i)} \kappa_{j,T}(M_i, s-t_{k-1}-1) \\
\leq &\sum_{i\in[m]} \sum_{k=1}^{\Gamma} \sum_{j=1}^{+\infty} \left(2B\sqrt{72\cdot 2^{-j}\cdot \ln T} + \sum_{s=1}^{\ell_{j,T}(M_i)} 2B\sqrt{\frac{72\cdot 2^{-j}\cdot \ln T}{s}}\right) \\
\leq &\sum_{i\in[m]} \sum_{k=1}^{\Gamma} \sum_{j=1}^{+\infty} \left(2B\sqrt{72\cdot 2^{-j}\cdot \ln T} + 2\cdot 2B\sqrt{72\cdot 2^{-j}\cdot \ln T}\cdot\sqrt{\ell_{j,T}(M_i)}\right) \\
\leq &\sum_{i\in[m]} \sum_{k=1}^{\Gamma} \sum_{j=1}^{+\infty} \left(2B\sqrt{72\cdot 2^{-j}\cdot \ln T} + 2\cdot 2B\sqrt{72\cdot 2^{-j}\cdot \ln T}\cdot\sqrt{\frac{288\cdot 2^{-j}\cdot B^2K^2\ln T}{M_i^2}}\right) \\
\leq &\sum_{i\in[m]} \sum_{k=1}^{\Gamma} \left(12(2+\sqrt{2})B\cdot\sqrt{\ln T} + \frac{576\cdot B^2K\cdot \ln T}{M_i}\right) \\
\leq &\sum_{i\in[m]} \left(\frac{T}{w}+1\right)\left(12(2+\sqrt{2})B\cdot\sqrt{\ln T} + \frac{576\cdot B^2K\cdot \ln T}{M_i}\right).
\end{aligned}$$

Then combining with Lemma 6, we have

$$\begin{aligned}
&\text{Reg}(\{\Delta_{S_t}^{D_t} \geq M_{S_t}\} \wedge \mathcal{N}_t^s \wedge \mathcal{N}_t^t \wedge \neg\mathcal{F}_t) \\
\leq &\sum_{i\in[m]} \left(\frac{T}{w}+1\right)\left(12(2+\sqrt{2})B\sqrt{\ln T} + \frac{576B^2K\ln T}{M_i}\right) + 2(1+\alpha)KB\sum_{s=2}^{t}||\mu_s-\mu_{s-1}||_\infty\cdot w.
\end{aligned}$$

□

Theorem 2 (Main content theorem 1 restated)**.** *Choosing the length of the sliding window to be* $w = \min\left\{\sqrt{\frac{T}{V}}, T\right\}$*, we have the following distribution dependent bound,*

$$\textit{Reg}_{\alpha,\beta} = \tilde{O}\left(\sum_{i\in[m]}\frac{K\sqrt{VT}}{\Delta_{\min}^i} + \sum_{i\in[m]}\frac{K}{\Delta_{\min}^i} + mK\right).$$

If we choose the length of the sliding window to be $w = \min\left\{m^{1/3}T^{2/3}K^{-1/3}V^{-2/3}, T\right\}$*, we have the following distribution independent bound,*

$$\textit{Reg}_{\alpha,\beta} = \tilde{O}\left((mV)^{1/3}(KT)^{2/3} + \sqrt{mKT} + mK\right).$$

Proof. First, from the definition of the filtered regret, we know that

$$\text{Reg}(\{\}) \leq \text{Reg}(\{\Delta_{S_t}^{D_t} \geq M_{S_t}\} \wedge \mathcal{N}_t^s \wedge \mathcal{N}_t^t \wedge \neg\mathcal{F}_t) + \text{Reg}(\{\Delta_{S_t}^{D_t} < M_{S_t}\}) + \text{Reg}(\neg\mathcal{N}_t^s) + \text{Reg}(\neg\mathcal{N}_t^t) + \text{Reg}(\mathcal{F}_t).$$

The last 3 terms are rather easy to bound, we have

$$\begin{aligned}
\text{Reg}(\neg\mathcal{N}_t^s) &= \sum_{t=1}^{T} \Delta_{S_t}^{D_t} \mathbb{I}\{\neg\mathcal{N}_t^s\} \leq \sum_{t=1}^{T} \Pr\{\neg\mathcal{N}_t^s\} \Delta_{\max} \leq \frac{\pi^2}{3} m \cdot \Delta_{\max} \\
\text{Reg}(\neg\mathcal{N}_t^t) &= \sum_{t=1}^{T} \Delta_{S_t}^{D_t} \mathbb{I}\{\neg\mathcal{N}_t^t\} \leq \sum_{t=1}^{T} \Pr\{\neg\mathcal{N}_t^t\} \Delta_{\max} \leq \frac{\pi^2}{6} \sum_{i\in[m]} j_{\max}^i \cdot \Delta_{\max} \\
\text{Reg}(\mathcal{F}_t) &= \sum_{t=1}^{T} \Delta_{S_t}^{D_t} \mathbb{I}\{\mathcal{F}_t\} \leq \sum_{t=1}^{T} \Pr\{\mathcal{F}_t\} \Delta_{\max}^t \leq (1-\beta) \cdot \sum_{t=1}^{T} \Delta_{\max}^t
\end{aligned}$$

We also know that

$$\begin{aligned}
&\text{Reg}_{\alpha,\beta}^{\mathcal{A}} - \text{Reg}(\{\Delta_{S_t}^{D_t} < M_{S_t}\}) \\
=&\alpha \cdot \beta \cdot \sum_{t=1}^{T} \text{opt}_{\boldsymbol{\mu}_t} - \mathbb{E}\left[\sum_{t=1}^{T} r_{S_t^{\mathcal{A}}}(\boldsymbol{\mu}_t)\right] - \text{Reg}(\{\Delta_{S_t}^{D_t} < M_{S_t}\}) \\
=&\text{Reg}(\{\}) - (1-\beta)\alpha \cdot \sum_{t=1}^{T} \text{opt}_{\boldsymbol{\mu}_t} - \text{Reg}(\{\Delta_{S_t}^{D_t} < M_{S_t}\}) \\
\leq&\text{Reg}(\{\Delta_{S_t}^{D_t} \geq M_{S_t}\} \wedge \mathcal{N}_t^s \wedge \mathcal{N}_t^t \wedge \neg\mathcal{F}_t) + \text{Reg}(\neg\mathcal{N}_t^s) + \text{Reg}(\neg\mathcal{N}_t^t) + \text{Reg}(\mathcal{F}_t) - (1-\beta)\alpha \cdot \sum_{t=1}^{T} \text{opt}_{\boldsymbol{\mu}_t} \\
\leq&\text{Reg}(\{\Delta_{S_t}^{D_t} \geq M_{S_t}\} \wedge \mathcal{N}_t^s \wedge \mathcal{N}_t^t \wedge \neg\mathcal{F}_t) + \frac{\pi^2}{3} m \cdot \Delta_{\max} + \frac{\pi^2}{6} \sum_{i\in[m]} j_{\max}^i \cdot \Delta_{\max} \\
&+ (1-\beta) \cdot \sum_{t=1}^{T} \Delta_{\max}^t - (1-\beta)\alpha \cdot \sum_{t=1}^{T} \text{opt}_{\boldsymbol{\mu}_t} \\
\leq&\text{Reg}(\{\Delta_{S_t}^{D_t} \geq M_{S_t}\} \wedge \mathcal{N}_t^s \wedge \mathcal{N}_t^t \wedge \neg\mathcal{F}_t) + \frac{\pi^2}{3} m \cdot \Delta_{\max} + \frac{\pi^2}{6} \sum_{i\in[m]} j_{\max}^i \cdot \Delta_{\max}.
\end{aligned}$$

Then we have

$$\text{Reg}_{\alpha,\beta}^{\mathcal{A}} \leq \text{Reg}(\{\Delta_{S_t}^{D_t} \geq M_{S_t}\} \wedge \mathcal{N}_t^s \wedge \mathcal{N}_t^t \wedge \neg\mathcal{F}_t) + \text{Reg}(\{\Delta_{S_t}^{D_t} < M_{S_t}\}) + \frac{\pi^2}{3} m \cdot \Delta_{\max} + \frac{\pi^2}{6} \sum_{i\in[m]} j_{\max}^i \cdot \Delta_{\max}.$$

Recall that from Lemma 7,

$$\begin{aligned}
&\text{Reg}(\{\Delta_{S_t}^{D_t} \geq M_{S_t}\} \wedge \mathcal{N}_t^s \wedge \mathcal{N}_t^t \wedge \neg\mathcal{F}_t) \\
\leq& \sum_{i\in[m]} \left(\frac{T}{w} + 1\right) \left(12(2+\sqrt{2})B\sqrt{\ln T} + \frac{576B^2 K \ln T}{M_i}\right) + 2(1+\alpha)KB \sum_{s=2}^{t} ||\mu_s - \mu_{s-1}||_\infty \cdot w.
\end{aligned}$$

For the distribution dependent bound, we choose $M_i = \Delta_{\min}^i$. Then, we have $\Delta_{S_t}^{D_t} \geq M_{S_t}$ and $\text{Reg}(\{\Delta_{S_t}^{D_t} < M_{S_t}\}) = 0$. If we set $w = \min\left\{\sqrt{\frac{T}{V}}, T\right\}$, we can get

$$\text{Reg}_{\alpha,\beta}^{\mathcal{A}} = \tilde{O}\left(\sum_{i\in[m]} \frac{K\sqrt{VT}}{\Delta_{\min}^i} + \sum_{i\in[m]} \frac{K}{\Delta_{\min}^i} + mK\right).$$

As for the distribution independent bound, if we set $w = \min\left\{m^{1/3}T^{2/3}K^{-1/3}V^{-2/3}, T\right\}, M_i = \sqrt{mK/w} = \Theta(\max\{(mKV)^{1/3}T^{-1/3}), \sqrt{mK/T}\}$, we can get

$$\text{Reg}_{\alpha,\beta}^{\mathcal{A}} = \tilde{O}\left((mV)^{1/3}(KT)^{2/3} + \sqrt{mKT} + mK\right) = \tilde{O}\left((mN)^{1/3}(KT)^{2/3} + \sqrt{mKT} + mK\right).$$

□

1.4 THEORETICAL GUARANTEES OF CUCB-BoB

In this section, we show the performance guarantee of our algorithm CUCB-BoB. Before moving into the formal proof, we will first introduce more on the EXP3 algorithm and its variant: EXP3.P algorithm.

Background on the EXP3 algorithm and its variant First we introduce the EXP3 algorithm and its variant EXP3.P algorithm. EXP3 algorithm is a famous algorithm for the adversarial bandit problem. In the original paper that introduce the Bandit-over-Bandit technique Cheung et al. [2019], the authors apply the EXP3 algorithm. However in our case, the regret is complicated and to make the proof easier, we apply the EXP3.P algorithm. The difference is that, the EXP3 algorithm has bounded "pseudo-regret", but the EXP3.P algorithm has bounded "regret" with high probability, and thus has bounded "expected regret". It is know that the "pseudo-regret" is a weaker measurement than the "expected regret", so for the ease of analysis, we apply EXP3.P algorithm.

Algorithm 1 EXP3.P

Input: Number of arms K', Total time horizon T', Parameters $\eta \in \mathbb{R}^+$, $\gamma, \beta \in [0, 1]$.
Let p_1 denote the uniform distribution over $[K']$.
for $t = 1, 2, \ldots, T'$ **do**
  Draw an arm I_t according to the probability distribution p_t.
  Compute the estimated gain for each arm

$$\tilde{g}_{i,t} = \frac{g_{i,t}\mathbb{I}\{I_t = i\} + \beta}{p_{i,t}}$$

  Update the estimated gain $\tilde{G}_{i,t} = \sum_{s=1}^{t} \tilde{g}_{i,s}$.
  Compute the new probability distribution over the arms $p_{t+1} = (p_{1,t+1}, \ldots, p_{K',t+1})$, where

$$p_{i,t+1} = (1-\gamma)\frac{\exp(\eta\tilde{G}_{i,t})}{\sum_{k=1}^{K'}\exp(\eta\tilde{G}_{k,t})} + \frac{\gamma}{K'}.$$

end for

Algorithm 1 is the pseudo-code for the EXP3.P algorithm. In the algorithm, $p_{i,t}$ is the gain (reward) in round t of arm i, and it satisfies $0 \le p_{i,t} \le 1$. It is easy to generalize the algorithm into the case where $0 \le p_{i,t} \le R'$, and we only have to normalize to $[0, 1]$ each time.

By choosing the parameters

$$\beta = \sqrt{\frac{\ln K'}{K'T'}}, \eta = 0.95\sqrt{\frac{\ln K'}{T'K'}}, \gamma = 1.05\sqrt{\frac{K'\ln K'}{T'}},$$

we have the following performance guarantee for the EXP3.P algorithm.

Proposition 3 (Main content proposition 1 restated)**.** *Suppose that the reward of each arm in each round is bounded by $0 \le r_{i,t} \le R'$, the number of arms is K', and the total time horizon is T'. The expected regret of EXP3.P algorithm is bounded by $O(R'\sqrt{K'T'\log K'})$.*

Proof of Theorem 2 in main content Now we prove Theorem 2 in main content (Theorem 3 in appendix). The main part of the proof is to decompose the regret into 2 parts, and optimize the length of each block to balance 2 parts. Recall that we have the following theorem.

Theorem 3 (Main content theorem 2 restated)**.** *Suppose that there exist R_1, R_2 such that $R_1 \le r_S(\mathbf{0}) \le r_S(\mathbf{1}) \le R_2$ for any $S \in \mathbb{S}$ and $R = R_2 - R_1$. Choosing $L = \sqrt{mKT}/R$, we have the following distribution-independent regret bound for $Reg_{\alpha,\beta}$,*

$$\tilde{O}\left((mV)^{\frac{1}{3}}(KT)^{\frac{2}{3}} + \sqrt{R}(mK)^{\frac{1}{4}}T^{\frac{3}{4}} + R\sqrt{mKT}\right).$$

Choosing $L = K^{2/3}T^{1/3}$, we have the following distribution-dependent regret bound

$$\tilde{O}\left(K\sqrt{\sum_{i\in[m]}\frac{TV}{\Delta^i_{\min}}} + \sum_{i\in[m]}\frac{K^{\frac{1}{3}}T^{\frac{2}{3}}}{\Delta^i_{\min}} + RK^{\frac{1}{3}}T^{\frac{2}{3}}\right).$$

Proof. We suppose that each block has length L, and there are $\lceil \frac{T}{L} \rceil$ blocks in total. Then, the reward in each block is bounded by $R' = RL$, since the reward in each round is bounded by R. We also know that the total number of possible length of sliding window is $K' = \lceil \log_2 L \rceil$, and the time horizon for the EXP3.P algorithm is $T' = \lceil \frac{T}{L} \rceil$.

From the definition of the (α, β)-approximation regret, we have

$$\begin{aligned}\text{Reg}^{\mathcal{A}}_{\boldsymbol{\mu},\alpha,\beta} =& \alpha \cdot \beta \cdot \sum_{t=1}^{T} \text{opt}_{\boldsymbol{\mu}_t} - \mathbb{E}\left[\sum_{t=1}^{T} r_{S_t^{\mathcal{A}}}(\boldsymbol{\mu}_t)\right] \\ =& \underbrace{\alpha \cdot \beta \cdot \sum_{t=1}^{T} \text{opt}_{\boldsymbol{\mu}_t} - \mathbb{E}\left[\sum_{t=1}^{T} r_{S_t^{\mathcal{B}}}(\boldsymbol{\mu}_t)\right]}_{\text{Term } \mathbb{A}} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T} r_{S_t^{\mathcal{B}}}(\boldsymbol{\mu}_t)\right] - \mathbb{E}\left[\sum_{t=1}^{T} r_{S_t^{\mathcal{A}}}(\boldsymbol{\mu}_t)\right]}_{\text{Term } \mathbb{B}},\end{aligned}$$

where $\mathcal{B}$ is another algorithm with the same block size but with fixed window size $w = 2^k$ for some number k. From Proposition 3, it is easy to know that for any fixed window size w and the induced algorithm $\mathcal{B}$, the second term (Term $\mathbb{B}$) is bounded by

$$\text{Term } \mathbb{B} \leq \tilde{O}(R'\sqrt{K'T'}) = \tilde{O}\left(RL\sqrt{\frac{T}{L}}\right) = \tilde{O}\left(R\sqrt{TL}\right).$$

Then, the remaining part is to select a window size w and bound Term $\mathbb{A}$. We decompose Term $\mathbb{A}$ into sum of regret of each block,

$$\text{Term } \mathbb{A} = \alpha \cdot \beta \cdot \sum_{t=1}^{T} \text{opt}_{\boldsymbol{\mu}_t} - \mathbb{E}\left[\sum_{t=1}^{T} r_{S_t^{\mathcal{B}}}(\boldsymbol{\mu}_t)\right] = \sum_{\ell=1}^{\lceil \frac{T}{L} \rceil} \left(\alpha \cdot \beta \cdot \sum_{s=L(\ell-1)+1}^{\min\{\ell L, T\}} \text{opt}_{\boldsymbol{\mu}_t} - \mathbb{E}\left[\sum_{s=L(\ell-1)+1}^{\min\{\ell L, T\}} r_{S_t^{\mathcal{B}}}(\boldsymbol{\mu}_t)\right]\right).$$

Suppose that in each block $\ell \leq \lceil \frac{T}{L} \rceil$, the variation in block ℓ is denoted by V_ℓ. Formally, we define

$$V_\ell = \sum_{s=L(\ell-1)+2}^{\min\{\ell L, T\}} ||\boldsymbol{\mu}_s - \boldsymbol{\mu}_{s-1}||_\infty.$$

Now we bound the regret in each block. The bound is similar to the proof in Theorem 2. Choosing $w = 2^k$ where $2^k \leq \min\{m^{1/3}T^{2/3}K^{-1/3}V^{-2/3}, L\} < 2^{k+1}$ and $M_i = \sqrt{mK/w}$. If we have $m^{1/3}T^{2/3}K^{-1/3}V^{-2/3} \leq L$, then the regret in block $\ell < \frac{T}{L}$ is bounded by

$$\tilde{O}\left((mV)^{1/3}K^{2/3}T^{-1/3} \cdot L + m^{1/3}(KT)^{2/3}V^{-2/3} \cdot V_\ell + mK\right).$$

The regret in last block is bounded by L, and Term $\mathbb{A}$ can be bounded by

$$\tilde{O}\left((mV)^{1/3}(KT)^{2/3} + L + mK\frac{T}{L}\right).$$

Then we consider the case when $(mK)^{1/3}T^{2/3}V^{-2/3} > L$. This time, the regret in each block is bounded by

$$\tilde{O}\left(\sqrt{mKL} + mK\right).$$

Then sum the regret in each block, we bound Term $\mathbb{A}$ by the following

$$\tilde{O}\left(\sqrt{mKL}\frac{T}{L} + L + mK\frac{T}{L}\right) = \tilde{O}\left(\sqrt{mK/L} \cdot T + L + mK\frac{T}{L}\right),$$

where the last term is the regret for the last block. Sum them up, we know that Term $\mathbb{A}$ is bounded by

$$\text{Term } \mathbb{A} \leq \tilde{O}\left((mV)^{1/3}(KT)^{2/3} + \sqrt{mK/L} \cdot T + L + mK\frac{T}{L}\right).$$

Then combining Term $\mathbb{B}$, we have

$$\text{Reg}_{\alpha,\beta}^{\mathcal{A}} = \tilde{O}\left((mV)^{1/3}(KT)^{2/3} + \sqrt{mK/L}\cdot T + L + R\sqrt{TL} + mK\frac{T}{L}\right).$$

Choosing $L = \sqrt{mKT}/R$, the regret is bounded by

$$\text{Reg}_{\alpha,\beta}^{\mathcal{A}} = \tilde{O}\left((mV)^{1/3}(KT)^{2/3} + \sqrt{R}(mK)^{1/4}T^{3/4} + R\sqrt{mKT}\right).$$

Next, we consider the distribution dependent bound. Now, we choose $w = 2^k$ where $2^k \leq \min\left\{\sqrt{\frac{T}{V}\cdot\sum_{i\in[m]}\frac{1}{\Delta_{\min}^i}}, L\right\} < 2^{k+1}$. First we consider the case when $\sqrt{\frac{T}{V}\cdot\sum_{i\in[m]}\frac{1}{\Delta_{\min}^i}} \leq L$. In this case, the regret in block ℓ (except for the last one) is bounded by

$$\tilde{O}\left(\frac{L}{w}\cdot\sum_{i\in[m]}\frac{K}{\Delta_{\min}^i} + w\cdot V_\ell + mK\right).$$

Summing up the regret in each block, we can know that Term $\mathbb{A}$ in this case is bounded by

$$\tilde{O}\left(K\sqrt{TV\cdot\sum_{i\in[m]}\frac{1}{\Delta_{\min}^i}} + mKL\right).$$

Then consider the case when $\sqrt{\frac{T}{V}\cdot\sum_{i\in[m]}\frac{1}{\Delta_{\min}^i}} > L$. In this case, the regret for block ℓ is bounded by

$$\tilde{O}\left(\sum_{i\in[m]}\frac{K}{\Delta_{\min}^i} + mK\right).$$

Summing up the regret in each block, we know that Term $\mathbb{A}$ is bounded by

$$\tilde{O}\left(\frac{T}{L}\cdot\sum_{i\in[m]}\frac{K}{\Delta_{\min}^i} + mK\frac{T}{L}\right).$$

Combining the regret bound in each case, we know that

$$\text{Term } \mathbb{A} = \tilde{O}\left(K\sqrt{TV\cdot\sum_{i\in[m]}\frac{1}{\Delta_{\min}^i}} + \frac{T}{L}\cdot\sum_{i\in[m]}\frac{K}{\Delta_{\min}^i} + mK\frac{T}{L}\right).$$

Take Term $\mathbb{B}$ into account, we have

$$\text{Reg}_{\alpha,\beta}^{\mathcal{A}} = \tilde{O}\left(K\sqrt{TV\cdot\sum_{i\in[m]}\frac{1}{\Delta_{\min}^i}} + \frac{T}{L}\cdot\sum_{i\in[m]}\frac{K}{\Delta_{\min}^i} + mK\frac{T}{L} + R\sqrt{TL}\right).$$

Choosing $L = K^{2/3}T^{1/3}$, we can get

$$\text{Reg}_{\alpha,\beta}^{\mathcal{A}} = \tilde{O}\left(K\sqrt{TV\cdot\sum_{i\in[m]}\frac{1}{\Delta_{\min}^i}} + \sum_{i\in[m]}\frac{K^{\frac{1}{3}}T^{\frac{2}{3}}}{\Delta_{\min}^i} + RK^{\frac{1}{3}}T^{\frac{2}{3}}\right).$$

□

Algorithm 2 ADA-LCMAB

Input: confidence δ, time horizon T, action space $\mathbb{S}$

Definition: $\nu_j = \sqrt{\frac{C_0}{m2^j L}}$, where $C_0 = \ln\left(\frac{8T^3|\mathbb{S}|^2}{\delta}\right)$, $L = \lceil 4mC_0 \rceil$, $\mathcal{B}_{(i,j)} := [\iota_i, \iota_i + 2^j L - 1]$.

Initialize: $t = 1, i = 1$

$\iota_i \leftarrow t$

for $j = 0, 1, 2, \ldots$ **do**

  If $j = 0$, set $Q_{(i,j)}$ as an arbitrary distribution over $\mathbb{S}$; otherwise, let $(\boldsymbol{q}_{(i,j)}^{\nu_j}, Q_{(i,j)}^{\nu_j})$ be the associated solution and distribution of equation (5) with inputs $\mathcal{I} = \mathcal{B}_{(i,j-1)}$ and $\nu = \nu_j$

  $\mathcal{E} \leftarrow \emptyset$

  while $t \leqslant \iota_i + 2^j L - 1$ **do**

    Draw $\mathrm{REP} \sim \mathrm{Bernoulli}\left(\frac{1}{L} \times 2^{-j/2} \times \sum_{k=0}^{j-1} 2^{-k/2}\right)$

    if $\mathrm{REP} = 1$ **then**

      Sample n from $\{0, \ldots, j-1\}$ s.t. $\Pr[n = b] \propto 2^{-b/2}$

      $\mathcal{E} \leftarrow \mathcal{E} \cup \{(n, [t, t + 2^n L - 1])\}$

    end if

    Let $N_t := \{n | \exists \mathcal{I}$ such that $t \in \mathcal{I}$ and $(n, \mathcal{I}) \in \mathcal{E}\}$

    If N_t is empty, play $S_t \sim Q_{(i,j)}^{\nu_j}$; otherwise, sample $n \sim \mathrm{Uniform}(N_t)$, and play $S_t \sim Q_{(i,n)}^{\nu_n}$

    Receive $\{X_i^t | i \in S_t\}$ and calculate $\hat{\mu}_t$ according to equation (9)

    for $(n, [s, s']) \in \mathcal{E}$ **do**

      if $s' = t$ and ENDOFREPLAYTEST$(i, j, n, [s, t]) = Fail$ **then**

        $t \leftarrow t + 1, i \leftarrow i + 1$ and return to Line 4

      end if

    end for

    if $t = \iota_i + 2^j L - 1$ and ENFOFBLOCKTEST$(i, j) = Fail$ **then**

      $t \leftarrow t + 1, i \leftarrow i + 1$ and return to Line 4

    end if

  end while

end for

Procedure: ENDOFREPLAYTEST$(i, j, n, \mathcal{A})$:

Return *Fail* if there exists $S \in \mathbb{S}$ such that any of the following inequalities holds:

$$\widehat{\mathrm{Reg}}_{\mathcal{A}}(S) - 4\widehat{\mathrm{Reg}}_{\mathcal{B}(i,j-1)}(S) \geqslant 34mK\nu_n \log T \tag{1}$$

$$\widehat{\mathrm{Reg}}_{\mathcal{B}(i,j-1)}(S) - 4\widehat{\mathrm{Reg}}_{\mathcal{A}}(S) \geqslant 34mK\nu_n \log T \tag{2}$$

Procedure: ENDOFBLOCKTEST(i, j):

Return *Fail* if there exists $k \in \{0, 1, \ldots, j-1\}$ and $S \in \mathbb{S}$ such that any of the following inequalities holds:

$$\widehat{\mathrm{Reg}}_{\mathcal{B}(i,j)}(S) - 4\widehat{\mathrm{Reg}}_{\mathcal{B}(i,k)}(S) \geqslant 20mK\nu_k \log T \tag{3}$$

$$\widehat{\mathrm{Reg}}_{\mathcal{B}(i,k)}(S) - 4\widehat{\mathrm{Reg}}_{\mathcal{B}(i,j)}(S) \geqslant 20mK\nu_k \log T \tag{4}$$

2 MORE DETAILS IN SECTION 4

2.1 DETAILED ALGORITHM

In this part, we give our full algorithm pseudo-code. Please see Algorihtm 2 for more details.

2.2 OMITTED PROOFS IN SECTION 4

Lemma 8 (Main content lemma 1 restated)**.** *For any time interval I, its empirical reward estimation $\hat{\mu}_I$, and exploration parameter $\nu > 0$, let $\boldsymbol{q}_I^\nu$ be the solution to following optimization problem (5) with constant $C = 100$:*

$$\boldsymbol{q}_I^\nu = \underset{\boldsymbol{q} \in \mathrm{Conv}(\mathbb{S})_\nu}{\operatorname{argmax}} \langle \boldsymbol{q}, \hat{\boldsymbol{\mu}}_I \rangle + C\nu \sum_{i=1}^{m} \log q_i \tag{5}$$

Let Q_I^ν be the distribution over $\mathbb{N}$ such that $\mathbb{E}_{S \sim Q_I^\nu}[\mathbf{1}_S] = \boldsymbol{q}_I^\nu$, then there is

$$\sum_{S \in \mathbb{S}} Q_I^\nu(S) \widehat{\mathrm{Reg}}_I(S) \leqslant Cm\nu \tag{6}$$

$$\forall S \in \mathbb{S},\ \mathrm{Var}(Q_I^\nu, S) \leqslant m + \frac{\widehat{\mathrm{Reg}}_I(S)}{C\nu} \tag{7}$$

Proof. Define loss function $F_{\mathcal{I}}(Q) := \sum_{S \in \mathbb{S}} Q(S) \widehat{\mathrm{Reg}}_{\mathcal{I}}(S) + C\nu \sum_{i=1}^{m} \ln(1/q_i)$ with decision domain $\Delta(\mathbb{S})_\nu := \{Q \in \mathbb{R}_+^{|\mathbb{S}|} | \sum_{S \in \mathbb{S}} Q(S) = 1, \forall i \in [m], q_i \geqslant \nu\}$ (recall $\boldsymbol{q}$ is the expectation vector of Q). Because the decision domain $\Delta(\mathbb{S})_\nu$ is compact and loss function $F_{\mathcal{I}}(Q)$ is strictly convex in $\Delta(\mathbb{S})_\nu$, there exists a unique minimizer. What's more, it is not difficult to see $Q_{\mathcal{I}}^\nu$ induced by the solution to equation (5) is exactly the minimizer of loss function $F_{\mathcal{I}}(Q)$. Now we prove the lemma.

Define $\Delta(\mathbb{S})'_\nu := \{Q \in \mathbb{R}_+^{|\mathbb{S}|} | \sum_{S \in \mathbb{S}} Q(S) \leqslant 1, \forall i \in [m], q_i \geqslant \nu\}$. We claim there is $\min_{Q \in \Delta(\mathbb{S})} F_{\mathcal{I}}(Q) = \min_{Q \in \Delta(\mathbb{S})'} F_{\mathcal{I}}(Q)$, otherwise we can increase the weight of $\hat{S}_{\mathcal{I}}$ in $\Delta(\mathbb{S})'_\nu$ until it reaches the boundary, which always decreases the loss value.

Since $\nabla F_{\mathcal{I}}(Q)|_{Q(S)} = \widehat{\mathrm{Reg}}_{\mathcal{I}}(S) - Cv \sum_{i \in S} 1/q_i$, according to KKT conditions, we have

$$\widehat{\mathrm{Reg}}_{\mathcal{I}}(S) - C\nu \sum_{i \in S} \frac{1}{\boldsymbol{q}_{\mathcal{I},i}^\nu} - \lambda_S - \sum_{i \in S} \lambda_i + \lambda = 0 \tag{8}$$

for some Lagrangian multipliers $\lambda_S \geqslant 0, \lambda_i \geqslant 0, \lambda \geqslant 0$. Multiplying both sides by $Q_{\mathcal{I}}^\nu(S)$ and summing over $S \in \mathbb{S}$ give

$$\begin{aligned}
\sum_{S \in \mathbb{S}} Q_{\mathcal{I}}^\nu(S) \widehat{\mathrm{Reg}}_{\mathcal{I}}(S) &= C\nu \sum_{S \in \mathbb{S}} Q_{\mathcal{I}}^\nu(S) \sum_{i \in S} \frac{1}{\boldsymbol{q}_{\mathcal{I},i}^\nu} + \sum_{S \in \mathbb{S}} Q_{\mathcal{I}}^\nu(S) \lambda_S + \sum_{S \in \mathbb{S}} \sum_{i \in S} Q_{\mathcal{I}}^\nu(S) \lambda_i - \lambda \\
&= C\nu \sum_{S \in \mathbb{S}} Q_{\mathcal{I}}^\nu(S) \sum_{i \in S} \frac{1}{\boldsymbol{q}_{\mathcal{I},i}^\nu} - \lambda \\
&= Cm\nu - \lambda \\
&\leqslant Cm\nu
\end{aligned}$$

where the second equality is because of complementary slackness. Now we have proved the inequality (6) stated in the theorem. What's more, as $\widehat{\mathrm{Reg}}_{\mathcal{I}}(S) \geqslant 0$ for $\forall S \in \mathbb{S}$, there is $\lambda \leqslant Cm\nu$.

Rearranging from equation (8), we know

$$\begin{aligned}
\sum_{i \in S} \frac{1}{\boldsymbol{q}_{\mathcal{I},i}^\nu} &= \frac{1}{C\nu} \left(\widehat{\mathrm{Reg}}_{\mathcal{I}}(S) - \lambda_S - \sum_{i \in S} \lambda_i + \lambda \right) \\
&\leqslant m + \frac{\widehat{\mathrm{Reg}}_{\mathcal{I}}(S)}{C\nu}
\end{aligned}$$

which finishes the proof of inequality (7). □

For any interval $\mathcal{I}$ that lies in a block j of epoch i (i.e. $[\iota_i + 2^{j-1}L, \iota_i + 2^j L - 1]$), define $\varepsilon_{\mathcal{I}} := \max_{S\in\mathbb{S}} \mathrm{Reg}_{\mathcal{I}}(S) - 8\widehat{\mathrm{Reg}}_{\mathcal{B}_{(i,j-1)}}(S), \alpha_{\mathcal{I}} = \sqrt{\frac{2mC_0}{|\mathcal{I}|}}\log_2 T$, where $\mathrm{Reg}_{\mathcal{I}}(S) := \sum_{t\in\mathcal{I}} \mathrm{opt}_{\mu_t} - r_{S_t}(\mu_t)$. In Lemma 9 and Lemma 10, since we consider the regret in epoch i, we use $\mathcal{B}_j$ to represent $\mathcal{B}_{(i,j)}$ for simplicity.

Lemma 9. *With probability $1-\delta$,* ADA-LCMAB *guarantees for any block j and any interval $\mathcal{I}$ lies in block j,*

$$\sum_{t\in\mathcal{I}} \mathrm{opt}_{\mu_t} - r_{S_t}(\mu_t) = \tilde{O}\left(|\mathcal{I}|mK\nu_n + |\mathcal{I}|(K\alpha_{\mathcal{I}} + K\Delta_{\mathcal{I}} + \epsilon_{\mathcal{I}}\mathbb{I}_{\varepsilon_{\mathcal{I}} > D_3 K\alpha_{\mathcal{I}}})\right)$$

where $D_3 = 170$.

Proof. First, according to Azuma's inequality and a union bound over all T^2 intervals, with probability $1-\delta$, for any interval $\mathcal{I}$, there is

$$\sum_{t\in\mathcal{I}} \mathrm{opt}_{\mu_t} - r_{S_t}(\mu_t) \leqslant \sum_{t\in\mathcal{I}} \mathbb{E}_t[\mathrm{opt}_{\mu_t} - r_{S_t}(\mu_t)] + O\left(K\sqrt{|\mathcal{I}|\log(T^2/\delta)}\right) \tag{9}$$

Now we bound the conditional expectation in above inequality.

Note

$$\mathbb{E}_t[\mathrm{opt}_{\mu_t} - r_{S_t}(\mu_t)] = \begin{cases} \sum_{S\in\mathbb{S}} Q_j^{\nu_j}(S)(\mathrm{opt}_{\mu_t} - r_S(\mu_t)) & \text{if } N_t = \emptyset \\ \sum_{S\in\mathbb{S}} \sum_{n\in N_t} \frac{Q_n^{\nu_n}(S)}{|N_t|}(\mathrm{opt}_{\mu_t} - r_S(\mu_t)) & \text{if } N_t \neq \emptyset \end{cases} \tag{10}$$

$$= \begin{cases} \sum_{S\in\mathbb{S}} Q_j^{\nu_j}(S)\mathrm{Reg}_t(S) & \text{if } N_t = \emptyset \\ \sum_{S\in\mathbb{S}} \sum_{n\in N_t} \frac{Q_n^{\nu_n}(S)}{|N_t|}\mathrm{Reg}_t(S) & \text{if } N_t \neq \emptyset \end{cases} \tag{11}$$

Now, for any $t \in \mathcal{I}$ and $n \in [j]$, there is

$$\begin{aligned}
\sum_{S\in\mathbb{S}} Q_n^{\nu_n}(S)\mathrm{Reg}_t(S) &\leqslant \sum_{S\in\mathbb{S}} Q_n^{\nu_n}(S)\mathrm{Reg}_{\mathcal{I}}(S) + O(K\Delta_{\mathcal{I}}) \qquad \text{(nearly the same as Lemma 8 in Chen et al. [2019])} \\
&= 8\sum_{S\in\mathbb{S}} Q_n^{\nu_n}(S)\widehat{\mathrm{Reg}}_{\mathcal{B}_{j-1}}(S) + O(K\Delta_{\mathcal{I}}) + \varepsilon_{\mathcal{I}} \\
&\leqslant 8\sum_{S\in\mathbb{S}} Q_n^{\nu_n}(S)\left(4\widehat{\mathrm{Reg}}_{\mathcal{B}_{n-1}}(S) + 20mK\nu_{n-1}\log T\right) + O(K\Delta_{\mathcal{I}}) + \varepsilon_{\mathcal{I}} \\
&\quad \text{(condition (3) doesn't hold)} \\
&\leqslant \tilde{O}(mK\nu_n + K\Delta_{\mathcal{I}}) + \epsilon_{\mathcal{I}} \\
&\leqslant \tilde{O}(mK\nu_n + K\Delta_{\mathcal{I}} + K\alpha_{\mathcal{I}}) + \varepsilon_{\mathcal{I}}\mathbb{I}_{\varepsilon_{\mathcal{I}} > D_3 K\alpha_{\mathcal{I}}}
\end{aligned}$$

Combining all above inequalities and using the fact $\sqrt{|\mathcal{I}|\log(T^2/\delta)} \leqslant O(|\mathcal{I}|\alpha_{\mathcal{I}})$ finish the proof. □

Next, we bound the dynamic regret in block j within epoch i, that is $\mathcal{J} := [\iota_i, \iota_{i+1} - 1] \cap [\iota_i + 2^{j-1}L, \iota_i + 2^j L - 1]$.

Lemma 10. *With probability $1-\delta$, Algorithm 2 has the following regret for any block $\mathcal{J}$:*

$$\sum_{t\in\mathcal{J}} (\mathrm{opt}_{\mu_t} - r_{S_t}(\mu_t)) = \tilde{O}\left(\min\left\{\sqrt{mC_0\mathcal{S}_{\mathcal{J}}|\mathcal{J}|}, \sqrt{mC_0|\mathcal{J}|} + C_0^{\frac{1}{3}} m^{\frac{4}{3}} \Delta_{\mathcal{J}}^{\frac{1}{3}} |\mathcal{J}|^{\frac{2}{3}}\right\}\right)$$

To prove this lemma, we first partition the block into several intervals with some desired properties. As the greedy algorithm in Chen et al. [2019] used to partition the block $\mathcal{J}$ is only based on the total variation of underlying distribution, we can directly use the same greedy algorithm in non-stationary CMAB and have the same result:

Lemma 11 (Lemma 5 in Chen et al. [2019])**.** *There exists a partition $\mathcal{I}_1 \cup \mathcal{I}_2 \cup \cdots \cup \mathcal{I}_\Gamma$ of block $\mathcal{J}$ such tht $\Delta_{\mathcal{I}_k} \leqslant \alpha_{\mathcal{I}_k}, \forall k \in [\Gamma]$, and $\Gamma = O(\min\{\mathcal{S}_{\mathcal{J}}, (mC_0)^{-\frac{1}{3}}\Delta_{\mathcal{J}}^{\frac{2}{3}}|\mathcal{J}|^{\frac{1}{3}} + 1\})$*

Next, we give some basic concentration results for Linear CMAB. Define $U_t(S) := \mathbb{E}_t[(r_S(\hat{\mu}_t) - r_S(\mu_t))^2]$.

Lemma 12. *For any $S \in \mathbb{S}$ and any time t in epoch i and block j, there is*

$$U_t(S) \leqslant \begin{cases} K\mathrm{Var}(Q^{\nu_n}_{(i,n)}, S)\log T\ (\forall n \in [N_t]) & \text{if } N_t \neq \emptyset \\ K\mathrm{Var}(Q^{\nu_j}_{(i,j)}, S) & \text{if } N_t = \emptyset \end{cases}$$

Proof. If $N_t \neq \emptyset$, then $U_t(S) \leqslant \mathbb{E}_t[r^2_S(\hat{\mu}_t)] = \mathbb{E}_t[(\hat{\mu}_t^\top \mathbf{1}_S)^2] \leqslant K\sum_{k\in S}\mathbb{E}_t[\hat{\mu}^2_{t,k}] \leqslant K\sum_{k\in S}\frac{1}{q_{t,k}}$, where $\boldsymbol{q}_t$ is the expectation of distribution Q_t played at round t. According to our Algorithm 2, we know $Q_t = \frac{1}{|N_t|}\sum_{n\in N_t} Q^{\nu_n}_{(i,n)}$ when $N_t \neq \emptyset$. Thus, $\boldsymbol{q}_t = \frac{1}{|N_t|}\sum_{n\in N_t}\boldsymbol{q}^{\nu_n}_{(i,n)}$ where $\boldsymbol{q}^{\nu_n}_{(i,n)}$ is the expectation of distribution $Q^{\nu_n}_{(i,n)}$, and $q_{t,k} \geqslant q^{\nu_n}_{(i,n),k}/|N_t|$. What's more, as $|N_t| \leqslant logT$, we then finish the proof when $N_t \neq \emptyset$. If N_t is empty, the proof is exactly the same. □

Lemma 13. *With probability at least $1-\delta/4$, for any $S \in \mathbb{S}$, we have*

$$|r_S(\hat{\mu}_{\mathcal{B}_{(i,j)}}) - r_S(\mu_{\mathcal{B}_{(i,j)}})| \leqslant \frac{\lambda}{|\mathcal{B}_{(i,j)}|}\sum_{t\in\mathcal{B}_{(i,j)}} U_t(S) + \frac{C_0}{\lambda|\mathcal{B}_{(i,j)}|} \quad (\forall\lambda \in (0, \frac{\nu_j}{K}])$$

and for any interval $\mathcal{A}$ covered by some replay phase of index n,

$$|r_S(\hat{\mu}_{\mathcal{A}}) - r_S(\mu_{\mathcal{A}})| \leqslant \frac{\lambda}{|\mathcal{A}|}\sum_{t\in\mathcal{A}} U_t(S) + \frac{C_0}{\lambda|\mathcal{A}|} \quad (\forall\lambda \in (0, \frac{\nu_n}{K}])$$

Proof. Using Freedman's inequality with respect to each term in the summation just like Lemma 14 in Chen et al. [2019]. □

Define $\textsc{Event}_1$ as the event that bounds in Lemma 13 holds, then $\textsc{Event}_1$ holds with probability at least $1-\delta/4$.

Lemma 14. *Assume $\textsc{Event}_1$ holds, and there is no restart triggered in $\mathcal{B}_j$, then the following hold for any $S \in \mathbb{S}$:*

$$\mathrm{Reg}_{\mathcal{B}_j}(S) \leqslant 2\widehat{\mathrm{Reg}}_{\mathcal{B}_j}(S) + 10mK\nu_j$$
$$\widehat{\mathrm{Reg}}_{\mathcal{B}_j}(S) \leqslant 2\mathrm{Reg}_{\mathcal{B}_j}(S) + 10mK\nu_j$$

Proof. We prove this lemma by induction. When $j = 0$, it's not hard to see $\mathrm{Reg}_{\mathcal{B}_0}(S) \leqslant K \leqslant 10mK\nu_0$,

$$\begin{aligned}
\widehat{\mathrm{Reg}}_{\mathcal{B}_0}(S) - \mathrm{Reg}_{\mathcal{B}_0}(S) &= r_{\hat{S}_{\mathcal{B}_0}}(\hat{\mu}_{\mathcal{B}_0}) - r_S(\hat{\mu}_{\mathcal{B}_0}) - r_{S_{\mathcal{B}_0}}(\mu_{\mathcal{B}_0}) + r_S(\mu_{\mathcal{B}_0}) \\
&\leqslant r_{\hat{S}_{\mathcal{B}_0}}(\hat{\mu}_{\mathcal{B}_0}) - r_S(\hat{\mu}_{\mathcal{B}_0}) - r_{\hat{S}_{\mathcal{B}_0}}(\mu_{\mathcal{B}_0}) + r_S(\mu_{\mathcal{B}_0}) \quad \text{(by the optimality of } S_{\mathcal{B}_0}) \\
&\leqslant 2\left(\frac{\nu_0}{KL}\sum_{t\in\mathcal{B}_0} U_t(S) + \frac{KC_0}{\nu_0 L}\right) \quad \text{(by the definition of } \textsc{Event}_1 \text{ with } \lambda = \nu_0/K\ ) \\
&\leqslant 2(K + K/2) \\
&\leqslant 4K
\end{aligned}$$

which implies $\widehat{\mathrm{Reg}}_{\mathcal{B}_0}(S) \leqslant 5K \leqslant 10mK\nu_0$.

Now, assume the inequalities hold for $\{0, \ldots, j-1\}$, then for any $t \in \mathcal{B}_j$ and any $n \in [1, j]$, there is

$$\begin{aligned}
\mathrm{Var}(Q^{\nu_n}_n, S) &\leqslant m + \frac{\widehat{\mathrm{Reg}}_{\mathcal{B}_{n-1}}(S)}{C\nu_n} \\
&\leqslant m + \frac{2\mathrm{Reg}_{\mathcal{B}_{n-1}}(S) + 10mK\nu_{n-1}}{C\nu_n} \\
&\leqslant \frac{\mathrm{Reg}_{\mathcal{B}_{n-1}}(S)}{3\nu_n} + mK \\
&\leqslant \frac{\mathrm{Reg}_{\mathcal{B}_{n-1}}(S)}{3\nu_j} + mK
\end{aligned}$$

Combining Lemma 12 above and Lemma 19 in Chen et al. [2019] gives the result in this theorem. □

Lemma 15. *Assume* EVENT$_1$ *holds. Let $\mathcal{A}$ be a complete replay phase of index n, if for any $S \in \mathcal{S}$, equation (2) in EndofReplayTest doesn't hold, then the following hold for all $S \in \mathbb{S}$:*

$$\begin{aligned}\mathrm{Reg}_{\mathcal{A}}(S) &\leqslant 2\widehat{\mathrm{Reg}}_{\mathcal{A}}(S) + C_3 mK\nu_n \\ \widehat{\mathrm{Reg}}_{\mathcal{A}}(S) &\leqslant 2\mathrm{Reg}_{\mathcal{A}}(S) + C_3 mK\nu_n\end{aligned}$$

where $C_3 = 15$

Proof. According to Lemma 9 and Lemma 12, we have

$$\begin{aligned}\mathrm{Var}(Q_n^{\nu_n}, S) &\leqslant m + \frac{\widehat{\mathrm{Reg}}_{\mathcal{B}_{n-1}}(S)}{C\nu_n} \\ &\leqslant m + \frac{4\widehat{\mathrm{Reg}}_{\mathcal{B}_{j-1}}(S) + 20mK\nu_n \log T}{C\nu_n} \\ &\leqslant \frac{30\log T}{C}mK + \frac{16\widehat{\mathrm{Reg}}_{\mathcal{A}}(S) + 136mK\nu_n \log T}{C\nu_n} \quad \text{(because of EndOfReplayTest)} \\ &\leqslant \frac{\widehat{\mathrm{Reg}}_{\mathcal{A}}(S)}{3\nu_n} + \frac{166\log T}{C}mK\end{aligned}$$

Combining Lemma 12 and Lemma 19 in Chen et al. [2019] proves the result. □

Lemma 16. *Assume* EVENT$_1$ *holds. Let $\mathcal{A} = [s, e]$ be a complete replay phase of index n, then the following hold for all $S \in \mathbb{S}$:*

$$\begin{aligned}\mathrm{Reg}_{\mathcal{A}}(S) &\leqslant 2\widehat{\mathrm{Reg}}_{\mathcal{A}}(S) + 4mK\nu_n + \bar{V}_{[\iota_i, e]} \\ \widehat{\mathrm{Reg}}_{\mathcal{A}}(S) &\leqslant 2\mathrm{Reg}_{\mathcal{A}}(S) + 4mK\nu_n + \bar{V}_{[\iota_i, e]}\end{aligned}$$

Proof. For any $t \in \mathcal{A}$, there is

$$\begin{aligned}\mathrm{Var}(Q_n^{\nu_n}, S) &\leqslant m + \frac{\widehat{\mathrm{Reg}}_{\mathcal{B}_{n-1}}(S)}{C\nu_n} \\ &\leqslant m + \frac{2\mathrm{Reg}_{\mathcal{B}_{n-1}}(S) + 10mK\nu_n}{C\nu_n} \quad \text{(because of Lemma 14)} \\ &\leqslant \frac{1}{2}mK + \frac{2\mathrm{Reg}_{\mathcal{A}}(S) + 2m\bar{V}_{[\iota_i, e]}}{C\nu_n} \quad \text{(because of Lemma 8 in Chen et al. [2019])} \\ &\leqslant \frac{\mathrm{Reg}_{\mathcal{A}}(S)}{3\nu_n} + \frac{1}{2}mK + \frac{2m\bar{V}_{[\iota_i, e]}}{C\nu_n}\end{aligned}$$

Combining Lemma 12 above and Lemma 19 in Chen et al. [2019] proves the result. □

Lemma 17. *Assume* EVENT$_1$ *holds. Let $\mathcal{I} = [s, e]$ be an interval in the fictitious block $\mathcal{J}'$ with index j, and such that $\bar{V}_{\mathcal{I}} \leqslant \alpha_{\mathcal{I}}, \epsilon_{\mathcal{I}} > D_3 K\alpha_{\mathcal{I}}$, then*

(1) there exist an index $n_{\mathcal{I}} \in \{0, 1, \ldots, j-1\}$ such that $D_3 mK\nu_{n+1}\log T \leqslant \epsilon_{\mathcal{I}} \leqslant D_3 mK\nu_n \log T$;

(2) $|\mathcal{I}| \geqslant 2^{n_{\mathcal{I}}} L$;

(3) if the algorithm starts a replay phase $\mathcal{A}$ with index $n_{\mathcal{I}}$ within the range of $[s, e - 2^{n_{\mathcal{I}}} L]$, then the algorithm restarts when the replay phase finishes.

Proof. For (1), on one hand $\epsilon_{\mathcal{I}} \leqslant K \leqslant D_3 mK\nu_0$; on the other hand, $\epsilon_{\mathcal{I}} > D_3 K\alpha_{\mathcal{I}} \geqslant D_3 mK\nu_j \log T$ because of the definition of $\alpha_{\mathcal{I}}, \nu_j$ and $|\mathcal{I}| \leqslant |\mathcal{J}'| \leqslant 2^{j-1}L$. Therefore, there must exist an index $n_{\mathcal{I}}$ such that the condition holds.

For (2), since $D_3 K\alpha_{\mathcal{I}} \leqslant D_3 mK\nu_{n_{\mathcal{I}}} \log T$, we have $|\mathcal{I}| > 2^{n_{\mathcal{I}}} L$.

For (3), we show that the EndOfReplayTest fails when the replay phase finishes. Suppose for $\forall S \in \mathbb{S}$, Eq.(2) doesn't hold, then according to Lemma 15, we know $\mathrm{Reg}_{\mathcal{A}}(S) \leqslant 2\widehat{\mathrm{Reg}}_{\mathcal{A}}(S) + C_3 mK\nu_{n_{\mathcal{I}}}$. Besides, we know there exists S' such that

$$\begin{aligned}
\mathrm{Reg}_{\mathcal{A}}(S') &\geqslant \mathrm{Reg}_{\mathcal{I}}(S') - 2K\bar{V}_{\mathcal{I}} \quad \text{(because of Lemma 8 in Chen et al. [2019])} \\
&\geqslant 8\widehat{\mathrm{Reg}}_{\mathcal{B}_{j-1}}(S') + \epsilon_{\mathcal{I}} - 2K\bar{V}_{\mathcal{I}} \quad \text{(because of the definition of } \epsilon_{\mathcal{I}}) \\
&\geqslant 8\widehat{\mathrm{Reg}}_{\mathcal{B}_{j-1}}(S') + (D_3/2 - 2)mK\nu_{n_{\mathcal{I}}} \log T
\end{aligned}$$

Combining above two inequalities, we have

$$\begin{aligned}
\widehat{\mathrm{Reg}}_{\mathcal{A}}(S') &> 4\widehat{\mathrm{Reg}}_{\mathcal{B}_{j-1}}(S') + \frac{0.5D_3 - 2 - C_3}{2} mK\nu_{n_{\mathcal{I}}} \log T \\
&= 4\widehat{\mathrm{Reg}}_{\mathcal{B}_{j-1}}(S') + 34mK\nu_{n_{\mathcal{I}}} \log T
\end{aligned}$$

which is the Eq.(1) in EndOfReplayTest, thus the algorithm will restart. □

Now, we begin to prove Lemma 10.

Proof. Consider the fictitious partition constructed in Lemma 11, for the first $\Gamma - 1$ intervals, using Lemma 9 with respect to each interval as there is no restart. For the last interval Γ, we also use Lemma 9 but with the fictitious planned interval in the same way as in paper Chen et al. [2019].

Thus, for block j (i.e. $[\iota_i, \iota_{i+1} - 1] \cup [\iota_i + 2^{j-1}L - 1, \iota_i + 2^j L - 1]$), there is

$$\begin{aligned}
&\sum_{t \in \mathcal{J}} \mathrm{opt}_{\mu_t} - r_{S_t}(\mu_t) \\
\leqslant &\underbrace{\sum_{k=1}^{\Gamma} \sum_{t \in \mathcal{I}_k} \sum_{n \in N_t \cup \{j\}} mK\nu_n}_{\text{Term1}} + \underbrace{\sum_{k=1}^{\Gamma-1} K|\mathcal{I}_k|\alpha_{\mathcal{I}_k} + K|\mathcal{I}_\Gamma|\alpha_{\mathcal{I}'_\Gamma}}_{\text{Term2}} + \underbrace{\sum_{k=1}^{\Gamma-1} |\mathcal{I}_k|\varepsilon_{\mathcal{I}_k} I_{\varepsilon_{\mathcal{I}_k} > D_3 K \alpha_{\mathcal{I}_k}} + |\mathcal{I}_\Gamma|\varepsilon_{\mathcal{I}'_\Gamma} I_{\varepsilon_{\mathcal{I}'_\Gamma} > D_3 K \alpha_{\mathcal{I}'_\Gamma}}}_{\text{Term3}}
\end{aligned}$$

Using exactly the same technique as Chen et al. [2019] and Lemma 17 above, one can prove

$$\begin{aligned}
\text{Term1} &\leqslant O(\log(1/\delta)\sqrt{C_0 mK 2^j L}) \\
\text{Term2} &\leqslant O(\log T \sqrt{C_0 mK\Gamma|\mathcal{J}|}) \\
\text{Term3} &\leqslant O(\log(1/\delta)\log T \sqrt{C_0 mK\Gamma 2^j L})
\end{aligned}$$

Combining all above inequalities and Lemma 11 finishes the proof.

□

Theorem 4 (Theorem 3 restated)**.** *Algorithm 2 guarantees $\mathrm{Reg}_{1,1}^{\mathcal{A}}$ is upper bounded by*

$$\tilde{O}\left(\min\left\{\sqrt{mK^2NT}, \sqrt{mK^2T} + K(m\bar{V})^{\frac{1}{3}}T^{\frac{2}{3}}\right\}\right).$$

Proof. First, we bound the regret in an epoch i (i.e. $\mathcal{H}_i = [\iota_i, \iota_{i+1} - 1]$). For block j in epoch i, we denote it as $\mathcal{J}_{ij} = [\iota_i + 2^{j-1}L, \iota_i + 2^j L - 1] \cap \mathcal{H}_i$. As the last index of j is at most $j^* = \lceil \log(|\mathcal{H}_i/L|) \rceil$, we have

$$\begin{aligned}
\mathbb{E}\left[\sum_{t \in \mathcal{H}_i} \mathrm{opt}_{\mu_t} - r_{S_t}(\mu_t)\right] &\leqslant \tilde{O}\left(L + \sum_{j=1}^{j^*} \sqrt{C_0 mK^2 \mathcal{S}_{\mathcal{J}_{ij}} 2^j L}\right) \\
&= \tilde{O}\left(\sqrt{C_0 mK^2 \mathcal{S}_{\mathcal{H}_i}|\mathcal{H}_i|}\right)
\end{aligned}$$

Similarily, using Hölder inequality, we have

$$\mathbb{E}\left[\sum_{t\in\mathcal{H}_i}\mathrm{opt}_{\mu_t}-r_{S_t}(\mu_t)\right]\leqslant\tilde{O}\left(\sqrt{C_0mK^2|\mathcal{H}_i|}+KC_0^{\frac{1}{3}}m^{\frac{1}{3}}\bar{V}_{\mathcal{H}_i}^{\frac{1}{3}}|\mathcal{H}_i|^{\frac{2}{3}}\right)$$

According to Lemma 18 below, we know there is at most $E:=\min\{\mathcal{S},(C_0m)^{-\frac{1}{3}}\bar{V}^{\frac{2}{3}}T^{\frac{1}{3}}+1\}$ number of epochs with high probability, thus summing up the regret bound over all epochs, we have

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}\left[\mathrm{opt}_{\mu_t}-r_{S_t}(\mu_t)\right]&\leqslant\tilde{O}\left(\sum_{t=1}^{E}\sqrt{C_0mK^2\mathcal{S}_{\mathcal{H}_i}|\mathcal{H}_i|}\right)\\&\leqslant\tilde{O}\left(\sqrt{C_0mK^2\mathcal{S}T}\right)\end{aligned}$$

and

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}\left[\mathrm{opt}_{\mu_t}-r_{S_t}(\mu_t)\right]&\leqslant\tilde{O}\left(\sum_{t=1}^{E}\left(\sqrt{C_0mK^2|\mathcal{H}_i|}+KC_0^{\frac{1}{3}}m^{\frac{1}{3}}\bar{V}_{\mathcal{H}_i}^{\frac{1}{3}}|\mathcal{H}_i|^{\frac{2}{3}}\right)\right)\\&\leqslant\left(\sqrt{C_0mK^2T}+KC_0^{\frac{1}{3}}m^{\frac{1}{3}}\bar{V}^{\frac{1}{3}}T^{\frac{2}{3}}\right)\end{aligned}$$

□

Lemma 18. *Denote the number of restart by E. With probability $1-\delta$, we have $E\leqslant\min\{\mathcal{S},(C_0m)^{-\frac{1}{3}}\bar{V}^{\frac{2}{3}}T^{\frac{1}{3}}+1\}$.*

Proof. First, we prove that if for all t in epoch i with $\bar{V}_{[\iota_i,t]}\leqslant\sqrt{\frac{mC_0}{t-\iota_i+1}}$, restart will not be triggered at time t.

For EndOfBlockTest, suppose $t=\iota_i+2^jL-1$ for some j, then for any $S\in\mathcal{S},k\in[0,j-1]$, we have

$$\begin{aligned}\widehat{\mathrm{Reg}}_{\mathcal{B}_j}&\leqslant2\mathrm{Reg}_{\mathcal{B}_j}(S)+10mK\nu_j\quad\text{(because of Lemma 14)}\\&\leqslant2\mathrm{Reg}_{\mathcal{B}_k}(S)+10mK\nu_j+4m\bar{V}_{[\iota_i,t]}\quad\text{(because of Lemma 8 in Chen et al. [2019])}\\&\leqslant4\widehat{\mathrm{Reg}}_{\mathcal{B}_k}(S)+34mK\nu_j\quad\text{(because of above condition and definition of }\nu_j)\end{aligned}$$

Similarly, there is $\widehat{\mathrm{Reg}}_{\mathcal{B}_k}\leqslant4\widehat{\mathrm{Reg}}_{\mathcal{B}_j}+34mK\nu_j$. Thus, EndOfBlockTest will not return Fail.

For EndOfReplayTest, suppose $\mathcal{A}\subset[\iota_i,t]$ be a complete replay phase of index n, and $\bar{V}_{[\iota_i,t]}\leqslant\sqrt{\frac{mC_0}{|\mathcal{A}|}}$, we have

$$\begin{aligned}\widehat{\mathrm{Reg}}_{\mathcal{A}}&\leqslant2\mathrm{Reg}_{\mathcal{A}}(S)+4mK\nu_n+m\bar{V}_{[\iota_i,t]}\quad\text{(because of Lemma 16)}\\&\leqslant2\mathrm{Reg}_{\mathcal{B}_{j-1}}(S)+4mK\nu_n+5m\bar{V}_{[\iota_i,t]}\quad\text{(because of Lemma 8 in Chen et al. [2019])}\\&\leqslant4\widehat{\mathrm{Reg}}_{\mathcal{B}_k}(S)+20mK\nu_n\quad\text{(because of above condition and definition of }\nu_j)\end{aligned}$$

Similarly, there is $\widehat{\mathrm{Reg}}_{\mathcal{B}_{j-1}}\leqslant4\widehat{\mathrm{Reg}}_{\mathcal{B}_j}+20mK\nu_n$. Thus, EndOfBlockTest will not return Fail.

With above result, now we prove the theorem. If there is no distribution change which implies $\bar{V}_{[\iota_i,t]}=0$ then the algorithm will not restart. Therefore we have $E\leqslant\mathcal{S}$.

Denote the length of each epoch as $T_1,\ldots,T_E$, according to above result, we know there must be $\bar{V}_{\mathcal{H}_i}>\sqrt{\frac{mC_0}{T_i}}$. By Hölder's inequality, we have

$$\begin{aligned}E-1&\leqslant\sum_{i=1}^{E-1}T_i^{\frac{1}{3}}T_i^{-\frac{1}{3}}\\&\leqslant\left(\sum_{i=1}^{E-1}T_i\right)^{\frac{1}{3}}\left(\sum_{i=1}^{E-1}T_i^{-\frac{1}{2}}\right)^{\frac{2}{3}}\\&\leqslant T^{\frac{1}{3}}\left(\frac{\bar{V}}{\sqrt{mC_0}}\right)^{\frac{2}{3}}\\&\leqslant(mC_0)^{-\frac{1}{3}}\bar{V}^{\frac{2}{3}}T^{\frac{1}{3}}\end{aligned}$$

□

2.3 NON-STATIONARY LINEAR CMAB IN GENERAL CASE

In section 4, we need to solve an FTRL optimization probelm in Algorithm 2 and find a distribution Q over the decision space $\mathbb{S}$ such that its expectation is the solution to FTRL, which can only be implemented efficiently when $\mathrm{Conv}(\mathbb{S})_\nu$ is described by a polynomial number of constraints Zimmert et al. [2019], Combes et al. [2015], Sherali [1987]. In general, the problems with polynomial number of constraints for $\mathrm{Conv}(\mathbb{S})_\nu$ is a subset of all the problem with linear reward function and exact offline oracle, but there are also many of them whose convex hull can be represented by polynomial number of constraints. For example, for the TOP K arm problem, the convex hull of the feasible actions can be represented by polynomial number of constraints. Another non-trivial example is the bipartite matching problem. The convex hull of all the matchings in a bipartite graph can also be represented by polynomial number of constraints. This is due to the fact that, by applying the convex relaxation of the bipartite matching problem, the constraint matrix of the corresponding linear programming is a Totally Unimodular Matrix (TUM), and the resulting polytope of the linear programming is integral, i.e. all the vertices have integer coordinates. In this way, each vertex is a feasible matching, and the polytope is the convex hull.

To make it more general and get rid of the constraint about polynomial description of $\mathrm{Conv}(\mathbb{S})_\nu$, instead of solving FTRL and then calculating corresponding distribution Q, what we need to do is to find a distribution Q such that it satisfies inequalities (6) and (7) given in Lemma 8. In fact, we can achieve this goal using similar methods as in Agarwal et al. [2014], Chen et al. [2019] to find a sparse distribution over $\mathbb{S}$ efficiently through our offline exact oracle or equivalently an ERM oracle [1].

References

Alekh Agarwal, Daniel Hsu, Satyen Kale, John Langford, Lihong Li, and Robert Schapire. Taming the monster: A fast and simple algorithm for contextual bandits. In *International Conference on Machine Learning*, pages 1638–1646, 2014.

Yifang Chen, Chung-Wei Lee, Haipeng Luo, and Chen-Yu Wei. A new algorithm for non-stationary contextual bandits: Efficient, optimal and parameter-free. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 696–726, Phoenix, USA, 25–28 Jun 2019. PMLR.

Wang Chi Cheung, David Simchi-Levi, and Ruihao Zhu. Learning to optimize under non-stationarity. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 1079–1087. PMLR, 16–18 Apr 2019.

Richard Combes, Mohammad Sadegh Talebi Mazraeh Shahi, Alexandre Proutiere, et al. Combinatorial bandits revisited. In *Advances in Neural Information Processing Systems*, pages 2107–2115, 2015.

Hanif D Sherali. A constructive proof of the representation theorem for polyhedral sets based on fundamental definitions. *American Journal of Mathematical and Management Sciences*, 7(3-4):253–270, 1987.

Qinshi Wang and Wei Chen. Improving regret bounds for combinatorial semi-bandits with probabilistically triggered arms and its applications. In *Advances in Neural Information Processing Systems*, pages 1161–1171, 2017.

Haoyu Zhao and Wei Chen. Online second price auction with semi-bandit feedback under the non-stationary setting. *arXiv preprint arXiv:1911.05949*, 2019.

Julian Zimmert, Haipeng Luo, and Chen-Yu Wei. Beating stochastic and adversarial semi-bandits optimally and simultaneously. In *International Conference on Machine Learning*, pages 7683–7692, 2019.

[1] We also need to add a small exploration probability over m super arms where i-th super arm contains base arm i in Step 15 of Algorithm 2 just like Chen et al. [2019].