

# Pure Exploration of Combinatorial Bandits

Shouyuan Chen

April 29, 2014

## 1 Pure Exploration of Combinatorial Bandits

**ExpCMAB: problem formulation.** Suppose that the arms are numbered  $1, 2, \dots, n$ . Each arm  $e \in [n]$  is associated with a reward distribution  $\varphi_e$ . We assume that all reward distributions are  $B$ -sub-Gaussian, i.e.  $(\cdot)$ . Notice that all distributions that are supported on  $[0, B]$  are  $B$ -sub-Gaussian distributions  $\square$ . Let  $w(e)$  denote the expected reward of arm  $e$ , i.e.  $w(e) = \mathbb{E}_{X \sim \varphi_e}[X]$ . In addition, for any set of arms  $M \subseteq [n]$ , we define  $w(M) = \sum_{e \in M} w(e)$  as the sum of expected rewards of arms that belong to  $M$ .

The learning problem of pure exploration combinatorial bandit can be formalized as a game between a learner and a stochastic environment. At the beginning of the game, the learner is given a collection of feasible sets  $\mathcal{M} \subseteq 2^{[n]}$  which corresponds to some combinatorial problem. And the reward distributions  $\{\varphi_e\}_{e \in [n]}$  are unknown to the learner. Then, the game is played for multiple rounds; on each round  $t$ , the learner pulls an arm  $p_t \in [n]$  and observes a reward sampled from the associated reward distribution  $\varphi_{p_t}$ . The game continues until certain stopping condition is satisfied, which will be specified later. After the game finishes, the learner is asked to output a set of arms  $\text{Out} \in \mathcal{M}$  which approximately maximizes the sum of expected weight, i.e.  $w(M_*) - w(\text{Out}) \leq \epsilon$ , where we denote  $M_* = \arg \max_{M \in \mathcal{M}} w(M)$  to be the optimal set of arms. For the sake of simplicity, we shall assume that the optimal set  $M_*$  is unique throughout the paper. Notice that, if  $\epsilon = 0$ , then the learner is required to identify the optimal set, i.e.  $\text{Out} = M_*$ .

**Fixed confidence and fixed budget.** We consider two different stopping conditions of the game, which are known as *fixed confidence* setting and *fixed budget* setting. In the fixed confidence setting, the learner can stop the game at any point and her goal is to achieve a fixed confidence about the optimality of the returned set using a small number of samples (pulls). Specifically, given a confidence parameter  $\delta$ , the learner need to guarantee that  $\Pr[w(M_*) - w(\text{Out}) \leq \epsilon] \geq 1 - \delta$ . The performance is evaluated by the number of pulls used by the learner. Notice that the learner can stop the game at any point in this setting. In the fixed budget setting, the game stops after a fixed number rounds. The learner tries to minimize the probability of error  $\Pr[w(M_*) - w(\text{Out}) > \epsilon]$  within these rounds. In this case, the learner's performance is measured by the probability of error.

**Applications.** Our formulation of the ExpCMAB problem covers many online learning tasks. We consider the following applications as running examples.

- **Multi.**
- **Match.**
- **Path.**

## 2 Algorithm and Main Results

Our main contribution is an algorithm for solving the ExpCMAB problem. Our algorithm Then, we analyze the sample complexity and the probability of error of our algorithm.

**Maximization oracle.** For most non-trivial combinatorial problems, the size of the collection of feasible sets  $\mathcal{M}$  is exponential in  $n$ . Therefore, the learning algorithm needs a succinct representation of  $\mathcal{M}$ . In particular, we allow the learning algorithm to use a *maximization oracle* which can find the optimal set  $M \in \mathcal{M}$  when the expected reward of each arm is known. Specifically, we assume that there exists an oracle which takes a vector  $\mathbf{v} = (v(1), \dots, v(n))^T$  as input and returns a set  $\text{Oracle}(\mathbf{v}) = \arg \max_{M \in \mathcal{M}} v(M)$ . It is clear that a large class of combinatorial problems admit efficient maximization oracles.

**Algorithm.** Our algorithm works for both fixed confidence and fixed budget settings. In either settings, the behaviors of our algorithm only differ in the construction of confidence radius and the stopping condition. In the following, we describe the procedure of our algorithm. Our algorithm maintains empirical mean  $\bar{w}_t(e)$  and confidence radius  $\text{rad}_t(e)$  for each arm  $e \in [n]$  and each round  $t$ . The construction of confidence radius ensures that  $|w(e) - \bar{w}_t(e)| \leq \text{rad}_t(e)$  holds with high probability for each arm  $e \in [n]$  and each round  $t > 0$ . At each round  $t$ , our algorithm accesses the maximization oracle twice. The first access to the oracle computes the set  $M_t = \arg \max_{M \in \mathcal{M}} \bar{w}_t(M)$ . Notice that  $M_t$  is the “best” set according to the empirical means  $\bar{w}_t$ . Then, in order to explore possible refinements of  $M_t$ , the algorithm uses the confidence radius to compute an adjusted expectation vector  $\tilde{w}_t$  in the following way: for each arm  $e \in M_t$ ,  $\tilde{w}_t(e)$  equals to the lower confidence bound  $\tilde{w}_t(e) = \bar{w}_t(e) - \text{rad}_t(e)$ ; and for each arm  $e \notin M_t$ ,  $\tilde{w}_t(e)$  equals to the upper confidence bound  $\tilde{w}_t(e) = \bar{w}_t(e) + \text{rad}_t(e)$ . Intuitively, the adjusted expectation vector  $\tilde{w}_t$  penalizes arms belonging to  $M_t$  and encourages exploring arms out of  $M_t$ . The algorithm then calls the oracle using the adjusted expectation vector  $\tilde{w}_t$  as input, which returns another set  $\tilde{M}_t = \arg \max_{M \in \mathcal{M}} \tilde{w}_t(M)$ . The algorithm stops if  $\tilde{M}_t = M_t$  or the budget is exhausted, i.e.  $t = T$ , in the fixed budget setting. In either cases, the algorithm outputs  $\text{Out} = M_t$  as result. Otherwise, the algorithm plays the arm belonging to the symmetric difference  $(\tilde{M}_t \setminus M_t) \cup (M_t \setminus \tilde{M}_t)$  with the largest confidence radius in the end of round  $t$ . The pseudo-code of the algorithm is shown in Algorithm 1.

---

**Algorithm 1** CGapExp: Combinatorial Gap Exploration

---

**Require:** Confidence parameter:  $\delta \in (0, 1)$ ; Maximization oracle:  $\text{Oracle}(\cdot) : \mathbb{R}^n \rightarrow \mathcal{M}$ .

**Initialize:** Play each arm  $e \in [n]$  once. Initialize empirical means  $\bar{w}_n$  and set  $T_n(e) \leftarrow 1$  for all  $e$ .

```

1: for  $t = n, n + 1, \dots$  do
2:    $M_t \leftarrow \text{Oracle}(\bar{w}_t)$ 
3:   for  $e \in [n]$  do
4:     if  $e \in M_t$  then
5:        $\tilde{w}_t(e) \leftarrow \bar{w}_t(e) - \text{rad}_t(e)$ 
6:     else
7:        $\tilde{w}_t(e) \leftarrow \bar{w}_t(e) + \text{rad}_t(e)$ 
8:     end if
9:   end for
10:   $\tilde{M}_t \leftarrow \text{Oracle}(\tilde{w}_t)$ 
11:  if  $\tilde{M}_t = M_t$  then
12:     $\text{Out} \leftarrow M_t$ 
13:    return Out
14:  end if
15:   $p_t \leftarrow \arg \max_{e \in (\tilde{M}_t \setminus M_t) \cup (M_t \setminus \tilde{M}_t)} \text{rad}_t(e)$ 
16:  Play arm  $p_t$  and observe the reward
17:  Update empirical means  $\bar{w}_{t+1}$  using the observed reward
18:  Update number of samples:  $T_{t+1}(p_t) \leftarrow T_t(p_t) + 1$  and  $T_{t+1}(e) \leftarrow T_t(e)$  for all  $e \neq p_t$ 
19: end for
```

---

## 2.1 Analysis

In this part, we analyze the performance of Algorithm 1 for both fixed confidence and fixed budget settings.

**Gap.** We begin with defining a natural complexity measure of the ExpCMAB problem. For each arm  $e \in [n]$ , we define gap  $\Delta_e$  as

$$\Delta_e = \begin{cases} w(M_*) - \max_{M \in \mathcal{M}: e \in M} w(M) & \text{if } e \notin M_*, \\ w(M_*) - \max_{M \in \mathcal{M}: e \notin M} w(M) & \text{if } e \in M_*. \end{cases} \quad (1)$$

By this definition of gap  $\Delta_e$ , for each arm  $e \notin M_*$ ,  $\Delta_e$  represents the gap between the optimal set  $M_*$  and the best set that includes arm  $e$ ; and, for each arm  $e \in M_*$ ,  $\Delta_e$  is the sub-optimality of the best set that does not include arm  $e$ . We notice that this definition resembles the definition of gaps for **Multi** proposed by ().

**Exchange class.** The analysis of our algorithm depends on certain exchange properties of combinatorial structures. To study these properties, we introduce notions of *exchange set* and *exchange class* as our tools for the analysis. We present their definitions in the following.

We begin with the definition of exchange set. We define an exchange set  $b$  as an ordered pair of disjoint sets  $b = (b_+, b_-)$ . Then, we define operator  $\oplus$  such that, for any set  $M$  and any exchange set  $b = (b_+, b_-)$ , we have  $M \oplus b \triangleq M \setminus b_- \cup b_+$ . Similarly, we also define operator  $\ominus$  such that  $M \ominus b \triangleq M \setminus b_+ \cup b_-$ . We call a set of exchange sets  $\mathcal{B}$  an *exchange class* for  $\mathcal{M}$  if  $\mathcal{B}$  satisfies the following property. Let  $M$  and  $M'$  be two elements of  $\mathcal{M}$ . Then, for any  $e \in (M \setminus M')$ , there exists an exchange set  $(b_+, b_-) \in \mathcal{B}$  which satisfies  $e \in b_-$ ,  $b_+ \subseteq M' \setminus M$ ,  $b_- \subseteq M \setminus M'$ ,  $(M \oplus b) \in \mathcal{M}$  and  $(M' \ominus b) \in \mathcal{M}$ . Finally, we define the *width* of exchange class  $\mathcal{B}$  as follows

$$\text{width}(\mathcal{B}) = \max_{(b_+, b_-) \in \mathcal{B}} |b_+| + |b_-|.$$

Intuitively, for any feasible sets  $M$  and  $M'$ , there exists an exchange set  $(b_+, b_-) \in \mathcal{B}$  which can be seen as an “operation” that transforms  $M$  one step towards  $M'$ : this operation generates a new feasible set  $M \oplus b$  by removing elements (including  $e$ ) from  $M$  and adding elements which belongs to  $M'$ . One can chain these operations together such that, for any  $M \neq M'$ , there exists a sequence of exchange sets  $b_1, \dots, b_k$  of  $\mathcal{B}$  such that  $M' = M \oplus b_1 \dots \oplus b_k$ .

Next, we construct the exchange classes for our running examples. Our constructions are summarized in Lemma 1.

**Lemma 1.** *There exist exchange classes  $\mathcal{B}_{\text{Multi}}$ ,  $\mathcal{B}_{\text{Matroid}}$ ,  $\mathcal{B}_{\text{Match}}$  and  $\mathcal{B}_{\text{Path}}$  for  $\mathcal{M}_{\text{Multi}}$ ,  $\mathcal{M}_{\text{Matroid}}$ ,  $\mathcal{M}_{\text{Match}}$  and  $\mathcal{M}_{\text{Path}}$ , respectively. These exchange classes can be constructed as follows*

1.  $\mathcal{B}_{\text{Multi}} = \{(\{i\}, \{j\}) \mid \forall i \in [n], j \in [n]\}$ .
2.  $\mathcal{B}_{\text{Matroid}} = \{(\{i\}, \{j\}) \mid \forall i \in [n], j \in [n]\}$ .
3.  $\mathcal{B}_{\text{Match}} = \{(C_+, C_-) \mid C_+ \cup C_- \text{ is a cycle of } G\}$ .
4.  $\mathcal{B}_{\text{Path}} = \{(P_1, P_2) \mid P_1, P_2 \text{ are two disjoint paths of } G \text{ with same endpoints}\}$ .

In addition, we have  $\text{width}(\mathcal{B}_{\text{Multi}}) = 2$ ,  $\text{width}(\mathcal{B}_{\text{Matroid}}) = 2$ ,  $\text{width}(\mathcal{B}_{\text{Match}}) = |V|$  and  $\text{width}(\mathcal{B}_{\text{Path}}) = |V|$ .

The construction for **Multi** problem is straightforward. For **Matroid** problem, we leverage the basis exchange property of matroids (see Lemma 12 in the appendix). And for **Match** and **Path** problems, we appeal to the graph-theoretical properties of matchings and paths. We illustrate these exchange classes in Figure Y. A detailed proof of Lemma 1 is deferred to the supplementary material.

**Main results.** Our main results are problem-dependent bounds of our algorithm in both fixed confidence and fixed budget settings. First, we show that, in the fixed confidence setting, our algorithm returns the optimal set with high probability and uses at most  $\tilde{O}(\sum_e \text{width}(\mathcal{B})^2 / \Delta_e^2)$  samples.

**Theorem 1.** *Given any  $\delta \in (0, 1)$ , any combinatorial problem  $\mathcal{M} \subseteq 2^{[n]}$  and any vector  $\mathbf{w} \in \mathbb{R}^n$ . Assume that the reward distribution  $\varphi_e$  for each arm  $e \in [n]$  is  $B$ -sub-Gaussian with mean  $w(e)$ . Let  $\mathcal{B}$  be an exchange*

class for  $\mathcal{M}$  and let  $\{\Delta_e\}_{e \in [n]}$  be the gaps defined in Eq. (1). Then, with probability at least  $1 - \delta$ , Algorithm 1 returns the optimal set and

$$T \leq O \left( \frac{1}{B^2} \cdot \sum_{e \in [n]} \frac{\text{width}(\mathcal{B})^2}{\Delta_e^2} \log \left( \frac{1}{B^2} \cdot \frac{n}{\delta} \sum_{e \in [n]} \frac{\text{width}(\mathcal{B})^2}{\Delta_e^2} \right) \right),$$

where  $T$  denotes the number of samples used by Algorithm 1.

Theorem 1 is a general result which provides a sample complexity for any combinatorial problem  $\mathcal{M}$ . In addition, notice that  $\text{width}(\mathcal{B}_{\text{Multi}}) = O(1)$ . Therefore, the sample complexity bound of our algorithm for the **Multi** problem is  $O(\sum_e \Delta_e^{-2} \log(n\delta^{-1} \sum_a \Delta_a^{-2}))$ . This matches the best known problem-dependent bounds for the **Multi** problem due to XXX [], within logarithmic factors.

## 2.2 Lower bounds

In this part, we establish a problem-dependent lower bound on the sample complexity of the **ExpCMAB** problem. To state our results, we first define the notion of  $\delta$ -correct algorithm as follows. For any  $\delta \in (0, 1)$ , we call an algorithm  $\mathbb{A}$  a  $\delta$ -correct algorithm if, for any expected reward  $\mathbf{w}$ , the probability of error of  $\mathbb{A}$  is at most  $\delta$ , i.e  $\Pr[w(M_*) - w(\text{Out}) \geq \epsilon] \leq \delta$ , where  $\text{Out}$  is the output of algorithm  $\mathbb{A}$ .

Our next theorem shows that, for any combinatorial problem  $\mathcal{M}$ , any expected rewards  $\mathbf{w}$  and any  $\delta$ -correct algorithm  $\mathbb{A}$ , algorithm  $\mathbb{A}$  must use at least  $\tilde{\Omega}(\sum_e \frac{1}{\Delta_e^2})$  samples in expectation.

**Theorem 2.** Fix any  $\mathcal{M} \subseteq 2^{[n]}$  and any vector  $\mathbf{w} \in \mathbb{R}^n$ . Suppose that, for each arm  $e \in [n]$ , the reward distribution  $\varphi_e$  is given by  $\varphi_e = \mathcal{N}(w(e), 1)$ , where  $\mathcal{N}(\mu, \sigma^2)$  denotes a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ . Then, for any  $\delta \in (0, e^{-16}/4)$  and any  $\delta$ -correct algorithm  $\mathbb{A}$ , we have

$$\mathbb{E}[T] \geq \sum_e \frac{1}{16\Delta_e^2} \log(1/4\delta),$$

where  $T$  denote the number of total samples used by algorithm  $\mathbb{A}$  and  $\Delta_e$  is defined in Eq. (1).

Now, we compare the sample complexity of Algorithm 1 to the lower bound provided in Theorem 2 on our running examples **Multi**, **Matroid**, **Match** and **Path**. For clarity, we consider the case that  $\epsilon = 0$  which corresponds to the learning problem of finding the optimal set. We see that Algorithm 1 uses at most  $\tilde{O}(\sum_e \text{width}(\mathcal{B})^2 / \Delta_e^2)$  samples. Recall that, for **Multi** and **Matroid** problems, Lemma X shows that  $\text{width}(\mathcal{B}) = 2$ . Hence, for these two problems, Algorithm 1 achieves optimal sample complexity within logarithmic factors.

On the other hand, for **Match**( $V, E$ ) and **Path**( $V, E$ ), Lemma X indicates that  $\text{width}(\mathcal{B}) = |V| \leq n$ . This means that the gap between our algorithm and this lower bound is a factor of  $|V|^2$ . Notice this gap only depends on the underlying combinatorial structure of  $\mathcal{M}$  and is independent of expected rewards  $\mathbf{w}$ . This means that the sample complexity of Algorithm 1 has an optimal dependency on the gaps  $\{\Delta_e\}_{e \in [n]}$ .

However, we still remain to investigate the necessity of the dependency on  $\text{width}(\mathcal{B})$  of Algorithm 1. To this end, we provide evidence showing that the sample complexity of any  $\delta$ -correct algorithm should be related to size of exchange sets. In fact, we show that, for any optimal exchange set  $b \in \mathcal{B}_{\text{opt}}$  and any  $\delta$ -correct algorithm, the algorithm must spend  $\tilde{\Omega}(|b|^2/w(b)^2)$  samples on the arms belonging to  $b$ . This result is formalized in the following theorem.

**Theorem 3.** Fix any  $\mathcal{M} \subseteq 2^{[n]}$  and any vector  $\mathbf{w} \in \mathbb{R}^n$ . Suppose that, for each arm  $e \in [n]$ , the reward distribution  $\varphi_e$  is given by  $\varphi_e = \mathcal{N}(w(e), 1)$ , where  $\mathcal{N}(\mu, \sigma^2)$  denotes a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ . Fix any  $\delta \in (0, e^{-16}/4)$  and any  $\delta$ -correct algorithm  $\mathbb{A}$ .

Then, for any  $b \in \mathcal{B}_{\text{opt}}$ , we have

$$\mathbb{E}[T_b] \geq \frac{|b|^2}{32w(b)^2} \log(1/4\delta),$$

where  $T_b$  denotes the number of samples of arms belonging to  $b$  used by algorithm  $\mathbb{A}$ .

Notice that

### 3 Proof of Main Results

**Notations.** For any set  $a \subseteq [n]$ , let  $\chi_a \in \{0, 1\}^n$  denote the incidence vector of set  $a \subseteq [n]$ , i.e.  $\chi_a(e) = 1$  if and only if  $e \in a$ . For an exchange set  $b = (b_+, b_-)$ , we define  $\chi_b \triangleq \chi_{b_+} - \chi_{b_-}$  as the incidence vector of  $b$ . We notice that  $\chi_b \in \{-1, 0, 1\}^n$ .

For each round  $t$ , we define vector  $\mathbf{rad}_t = (\text{rad}_t(1), \dots, \text{rad}_t(n))^T$  and recall that  $\bar{\mathbf{w}}_t \in \mathbb{R}^n$  is the empirical mean rewards of arms up to round  $t$ .

Let  $\mathbf{u} \in \mathbb{R}^n$  and  $\mathbf{v} \in \mathbb{R}^n$  be two vectors. Let  $\langle \mathbf{u}, \mathbf{v} \rangle$  denote the inner product of  $\mathbf{u}$  and  $\mathbf{v}$ . We define  $\mathbf{u} \circ \mathbf{v} \triangleq (u(1) \cdot v(1), \dots, u(n) \cdot v(n))^T$  as the element-wise product of  $\mathbf{u}$  and  $\mathbf{v}$ . For any  $s \in \mathbb{R}$ , we also define  $\mathbf{u}^s \triangleq (u(1)^s, \dots, u(n)^s)^T$  as the element-wise exponentiation of  $\mathbf{u}$ . Let  $|\mathbf{u}| = (|u(1)|, \dots, |u(n)|)^T$  denote the element-wise absolute value of  $\mathbf{u}$ .

#### 3.1 Preparatory Lemmas

**Lemma 2.** Let  $M_1 \subseteq [n]$  be a set. Let  $b = (b_+, b_-)$  be an exchange set such that  $b_- \subseteq M_1$  and  $b_+ \cap M_1 = \emptyset$ . Define  $M_2 = M_1 \oplus b$ . Then, we have

$$\chi_{M_1} + \chi_b = \chi_{M_2}.$$

*Proof.* Recall that  $M_2 = M_1 \setminus b_- \oplus b_+$  and  $b_+ \cap b_- = \emptyset$ . Therefore we see that  $M_2 \setminus M_1 = b_+$  and  $M_1 \setminus M_2 = b_-$ . Then, we decompose  $\chi_{M_1}$  as  $\chi_{M_1} = \chi_{M_1 \setminus M_2} + \chi_{M_1 \cap M_2}$ . Hence, we have

$$\begin{aligned} \chi_{M_1} + \chi_b &= \chi_{M_1 \setminus M_2} + \chi_{M_1 \cap M_2} + \chi_{b_+} - \chi_{b_-} \\ &= \chi_{M_1 \cap M_2} + \chi_{M_2 \setminus M_1} \\ &= \chi_{M_2}. \end{aligned}$$

□

**Lemma 3.** Let  $\mathcal{M} \subseteq 2^{[n]}$  and  $\mathcal{B}$  be an exchange class for  $\mathcal{M}$ . Then, for any two different elements  $M, M'$  of  $\mathcal{M}$  and any  $e \in (M \setminus M') \cup (M' \setminus M)$ , there exists an exchange set  $b = (b_+, b_-) \in \mathcal{B}$  such that  $e \in (b_+ \cup b_-)$ ,  $b_- \subseteq (M \setminus M')$ ,  $b_+ \subseteq (M' \setminus M)$ ,  $(M \oplus b) \in \mathcal{M}$  and  $(M' \ominus b) \in \mathcal{M}$ . Moreover, if  $M' = M_*$ , then we have  $\langle \mathbf{w}, \chi_b \rangle \geq \Delta_e > 0$ , where  $\Delta_e$  is the gap defined in Eq. (1).

*Proof.* We decompose our proof into two cases.

**Case (1):**  $e \in M \setminus M'$ .

By the definition of exchange class, we know that there exists  $b = (b_+, b_-) \in \mathcal{B}$  which satisfies that  $e \in b_-$ ,  $b_- \subseteq (M \setminus M')$ ,  $b_+ \subseteq (M' \setminus M)$ ,  $(M \oplus b) \in \mathcal{M}$  and  $(M' \ominus b) \in \mathcal{M}$ .

Next, if  $M' = M_*$ , we see that  $e \notin M_*$ . Let us consider the set  $M_1 = \arg \max_{M': M' \in \mathcal{M} \wedge e \in M'} w(M')$ . Also define  $M_0 = M_* \ominus b$ . We have already proved that  $M_0 \in \mathcal{M}$ . Combining with the fact that  $e \in M_0$ , we see that  $w(M_0) \leq w(M_1)$ . Therefore, we obtain that  $w(M_*) - w(M_0) \geq w(M_*) - w(M_1) = \Delta_e$ . Notice that the left-hand side of the former inequality can be rewritten using Lemma 2 as follows

$$w(M_*) - w(M_0) = \langle \mathbf{w}, \chi_{M_*} \rangle - \langle \mathbf{w}, \chi_{M_0} \rangle = \langle \mathbf{w}, \chi_{M_*} - \chi_{M_0} \rangle = \langle \mathbf{w}, \chi_b \rangle.$$

Therefore, we obtain  $\langle \mathbf{w}, \chi_b \rangle \geq \Delta_e$ .

**Case (2):**  $e \in M' \setminus M$ .

Using the definition of exchange class, we see that there exists  $c = (c_+, c_-) \in \mathcal{B}$  such that  $e \in c_-$ ,  $c_- \subseteq (M' \setminus M)$ ,  $c_+ \subseteq (M \setminus M')$ ,  $(M' \oplus c) \in \mathcal{M}$  and  $(M \oplus c) \in \mathcal{M}$ .

We construct  $b = (b_+, b_-)$  by setting  $b_+ = c_-$  and  $b_- = c_+$ . Notice that, by the construction of  $b$ , we have  $M \oplus b = M \oplus c$  and  $M' \oplus b = M' \oplus c$ . Therefore, it is clear that  $b$  satisfies the requirement of the lemma.

Now, suppose that  $M' = M_*$ . In this case, we have  $e \in M_*$ . Consider the set  $M_3 = \arg \max_{M': M' \in \mathcal{M} \wedge e \notin M'} w(M')$ . We see that  $w(M_*) - w(M_3) = \Delta_e$ . Define  $M_2 = M_* \oplus b$  and notice that  $M_2 \in \mathcal{M}$ . Combining with the fact that  $e \notin M_2$ , we obtain that  $w(M_2) \leq w(M_3)$ . Hence, we have  $w(M_*) - w(M_2) \geq w(M_*) - w(M_3) = \Delta_e$ . Similar to Case (1), applying Lemma 2 again, we have

$$\langle \mathbf{w}, \chi_b \rangle = w(M_*) - w(M_2) \geq \Delta_e.$$

□

**Lemma 4.** Let  $M$  and  $M'$  be two sets. Then, we have

$$\max_{e \in (M \setminus M') \cup (M' \setminus M)} \text{rad}_t(e) = \|\mathbf{rad}_t \circ |\chi_{M'} - \chi_M|\|_\infty.$$

*Proof.* Notice that  $\chi_{M'} - \chi_M = \chi_{M' \setminus M} - \chi_{M \setminus M'}$ . In addition, since  $(M' \setminus M) \cap (M \setminus M') = \emptyset$ , we have  $\chi_{M' \setminus M} \circ \chi_{M \setminus M'} = \mathbf{0}_n$ . Also notice that  $\chi_{M' \setminus M} - \chi_{M \setminus M'} \in \{-1, 0, 1\}^n$ . Therefore, we have

$$\begin{aligned} |\chi_{M' \setminus M} - \chi_{M \setminus M'}| &= (\chi_{M' \setminus M} - \chi_{M \setminus M'})^2 \\ &= \chi_{M' \setminus M}^2 + \chi_{M \setminus M'}^2 + 2\chi_{M' \setminus M} \circ \chi_{M \setminus M'} \\ &= \chi_{M' \setminus M} + \chi_{M \setminus M'} \\ &= \chi_{(M' \setminus M) \cup (M \setminus M')}, \end{aligned}$$

where the third equation follows from the fact that  $\chi_{M \setminus M'} \in \{0, 1\}^n$  and  $\chi_{M' \setminus M} \in \{0, 1\}^n$ . The lemma follows immediately from the fact that  $\text{rad}_t(e) \geq 0$  and  $\chi_{(M \setminus M') \cup (M' \setminus M)} \in \{0, 1\}^n$ . □

**Lemma 5.** Let  $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$  be three vectors. Then, we have  $\langle \mathbf{a}, \mathbf{b} \circ \mathbf{c} \rangle = \langle \mathbf{a} \circ \mathbf{b}, \mathbf{c} \rangle$ .

*Proof.* We have

$$\langle \mathbf{a}, \mathbf{b} \circ \mathbf{c} \rangle = \sum_{i=1}^n a(i)(b(i)c(i)) = \sum_{i=1}^n (a(i)b(i))c(i) = \langle \mathbf{a} \circ \mathbf{b}, \mathbf{c} \rangle.$$

□

**Lemma 6.** Let  $M_t$  and  $\tilde{\mathbf{w}}_t$  be defined in Algorithm 1. Let  $M' \in \mathcal{M}$  be a feasible set. We have

$$\tilde{w}_t(M') - \tilde{w}_t(M_t) = \langle \tilde{\mathbf{w}}_t, \chi_{M'} - \chi_{M_t} \rangle = \langle \bar{\mathbf{w}}_t, \chi_{M'} - \chi_{M_t} \rangle + \langle \mathbf{rad}_t, |\chi_{M'} - \chi_{M_t}| \rangle.$$

*Proof.* We begin with proving the first part. It is easy to verify that  $\tilde{\mathbf{w}}_t = \bar{\mathbf{w}}_t + \mathbf{rad}_t \circ (\mathbf{1}_n - 2\chi_{M_t})$ . Then, we have

$$\begin{aligned} \langle \tilde{\mathbf{w}}_t, \chi_{M'} - \chi_{M_t} \rangle &= \langle \bar{\mathbf{w}}_t + \mathbf{rad}_t \circ (\mathbf{1}_n - 2\chi_{M_t}), \chi_{M'} - \chi_{M_t} \rangle \\ &= \langle \bar{\mathbf{w}}_t, \chi_{M'} - \chi_{M_t} \rangle + \langle \mathbf{rad}_t, (\mathbf{1}_n - 2\chi_{M_t}) \circ (\chi_{M'} - \chi_{M_t}) \rangle \\ &= \langle \bar{\mathbf{w}}_t, \chi_{M'} - \chi_{M_t} \rangle + \langle \mathbf{rad}_t, \chi_{M'} - \chi_{M_t} - 2\chi_{M_t} \circ \chi_{M'} + 2\chi_{M_t}^2 \rangle \end{aligned} \tag{2}$$

$$= \langle \bar{w}_t, \chi_{M'} - \chi_{M_t} \rangle + \langle \text{rad}_t, \chi_{M'}^2 - \chi_{M_t}^2 - 2\chi_{M_t} \circ \chi_{M'} + 2\chi_{M_t}^2 \rangle \quad (3)$$

$$= \langle \bar{w}_t, \chi_{M'} - \chi_{M_t} \rangle + \langle \text{rad}_t, (\chi_{M'} - \chi_{M_t})^2 \rangle \\ = \langle \bar{w}_t, \chi_{M'} - \chi_{M_t} \rangle + \langle \text{rad}_t, |\chi_{M'} - \chi_{M_t}| \rangle, \quad (4)$$

where Eq. (2) follows from Lemma 5; Eq. (3) holds since  $\chi_{M'} \in \{0, 1\}^n$  and  $\chi_{M_t} \in \{0, 1\}^n$  and therefore  $\chi_{M'} = \chi_{M'}^2$  and  $\chi_{M_t} = \chi_{M_t}^2$ ; and Eq. (4) follows since  $\chi_{M'} - \chi_{M_t} \in \{-1, 0, 1\}^n$ .

Next, recall that  $\tilde{M}_t = \arg \max_{M \in \mathcal{M}} \tilde{w}_t(M)$ . Therefore, we have  $\tilde{w}_t(\tilde{M}_t) \geq \tilde{w}_t(M')$ . Subtracting  $\tilde{w}_t(M_t)$  from both sides of the former inequality, we have

$$\tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) \geq \tilde{w}_t(M') - \tilde{w}_t(M_t). \quad (5)$$

The lemma follows by noticing that the left-hand side of Eq. (5) equals to  $\langle \bar{w}_t, \chi_{\tilde{M}_t} - \chi_{M_t} \rangle + \langle \text{rad}_t, |\chi_{\tilde{M}_t} - \chi_{M_t}| \rangle$  and the right-hand side equals to  $\langle \bar{w}_t, \chi_{M'} - \chi_{M_t} \rangle + \langle \text{rad}_t, |\chi_{M'} - \chi_{M_t}| \rangle$ .  $\square$

### 3.2 Confidence Intervals

For all  $t > 0$ , we define random event  $\xi_t$  as follows

$$\xi_t = \left\{ \forall i \in [n], \quad |w(i) - \bar{w}_t(i)| \leq \text{rad}_t(i) \right\}. \quad (6)$$

We notice that random event  $\xi_t$  characterizes the event that the confidence bounds of all arms are valid at round  $t$ .

**Lemma 7.** *Suppose that the reward distribution  $\varphi_e$  is a  $B$ -sub-Gaussian distribution for all  $e \in [n]$ . And if, for all  $t > 0$  and all  $e \in [n]$ , the confidence interval  $\text{rad}_t(e)$  is given by*

$$\text{rad}_t(e) = B \sqrt{\frac{2 \log \left( \frac{4nt^2}{\delta} \right)}{T_e(t)}},$$

where  $T_e(t)$  is the number of samples of arm  $e$  up to round  $t$ . Then, we have

$$\Pr \left[ \bigcap_{t=1}^{\infty} \xi_t \right] \geq 1 - \delta.$$

*Proof.* For any  $t > 0$  and  $e \in [n]$ , notice  $\varphi_e$  is a  $B$ -sub-Gaussian distribution with mean  $w(e)$  and  $\bar{w}_t(e)$  is the empirical mean of  $\varphi_e$  for  $T_e(t)$  samples. Using Hoeffding's inequality (see XXX), we obtain

$$\Pr \left[ |\bar{w}_t(e) - w(e)| \geq B \sqrt{\frac{2 \log \left( \frac{4nt^2}{\delta} \right)}{T_e(t)}} \right] \leq \frac{\delta}{2nt^2}.$$

By union bound over all  $e \in [n]$ , we see that  $\Pr[\xi_t] \geq 1 - \frac{\delta}{2t^2}$ . Using a union bound again over all  $t > 0$ , we have

$$\begin{aligned} \Pr \left[ \bigcap_{t=1}^{\infty} \xi_t \right] &\geq 1 - \sum_{t=1}^{\infty} \Pr[\neg \xi_t] \\ &\geq 1 - \sum_{t=1}^{\infty} \frac{\delta}{2t^2} \\ &= 1 - \frac{\pi^2}{12} \delta \geq 1 - \delta. \end{aligned}$$

□

**Lemma 8.** *Given any  $t > 0$ , assume that event  $\xi_t$  as defined in Eq. (6) occurs. Then, for any vector  $\mathbf{a} \in \mathbb{R}^n$ , we have*

$$|\langle \mathbf{w}, \mathbf{a} \rangle - \langle \bar{\mathbf{w}}_t, \mathbf{a} \rangle| \leq \langle \mathbf{rad}_t, |\mathbf{a}| \rangle.$$

*Proof.* Suppose that  $\xi$  occurs. Then, we have

$$\begin{aligned} |\langle \mathbf{w}, \mathbf{a} \rangle - \langle \bar{\mathbf{w}}_t, \mathbf{a} \rangle| &= |\langle \mathbf{w} - \bar{\mathbf{w}}_t, \mathbf{a} \rangle| \\ &= \left| \sum_{i=1}^n (w(i) - \bar{w}_t(i)) a(i) \right| \\ &\leq \sum_{i=1}^n |w(i) - \bar{w}_t(i)| |a(i)| \\ &\leq \sum_{i=1}^n \text{rad}_t(i) \cdot |a(i)| \\ &= \langle \mathbf{rad}_t, |\mathbf{a}| \rangle, \end{aligned} \tag{7}$$

where Eq. (7) follows the definition of event  $\xi_t$  in Eq. (6) and the assumption that it occurs. □

### 3.3 Main Lemmas

**Lemma 9.** *Given any  $t > 0$ , assume that event  $\xi_t$  (defined in Eq. (6)) occurs. Then, if Algorithm 1 terminates at round  $t$ , we have  $M_t = M_*$ .*

*Proof.* Suppose that  $M_t \neq M_*$ . By definition, we have  $w(M_*) > w(M_t)$ . Rewriting the former inequality, we obtain that  $\langle \mathbf{w}, \chi_{M_*} \rangle > \langle \mathbf{w}, \chi_{M_t} \rangle$ .

Applying Lemma 3 by setting  $M = M_t$  and  $M' = M_*$ , we see that there exists  $b = (b_+, b_-) \in \mathcal{B}$  such that  $(M_t \oplus b) \in \mathcal{M}$ .

Now define  $M'_t = M_t \oplus b$ . Recall that  $\tilde{M}_t = \arg \max_{M \in \mathcal{M}} \tilde{w}_t(M)$  and therefore  $\tilde{w}_t(\tilde{M}_t) \geq \tilde{w}_t(M'_t)$ . Hence, we have

$$\begin{aligned} \tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) &\geq \tilde{w}_t(M'_t) - \tilde{w}_t(M_t) \\ &= \langle \bar{\mathbf{w}}_t, \chi_{M'_t} - \chi_{M_t} \rangle + \langle \mathbf{rad}_t, |\chi_{M'_t} - \chi_{M_t}| \rangle \end{aligned} \tag{8}$$

$$\geq \langle \mathbf{w}, \chi_{M'_t} - \chi_{M_t} \rangle \tag{9}$$

$$= w(M'_t) - w(M_t) > 0, \tag{10}$$

where Eq. (8) follows from Lemma 6; and Eq. (9) follows the assumption that event  $\xi$  occurs and Lemma X; Therefore Eq. (10) shows that  $\tilde{w}_t(\tilde{M}_t) > \tilde{w}_t(M_t)$ . However, this contradicts to the stopping condition of Algorithm 1:  $\tilde{w}_t(\tilde{M}_t) \leq \tilde{w}_t(M_t)$  and the assumption that the algorithm terminates on round  $t$ . □

**Lemma 10.** *Given any  $t > 0$  and suppose that event  $\xi_t$  (defined in Eq. (6)) occurs. For any  $e \in [n]$ , if  $\text{rad}_t(e) < \frac{\Delta_e}{3 \text{width}(\mathcal{B})}$ , then, arm  $e$  will not be pulled on round  $t$ , i.e.  $p_t \neq e$ .*

*Proof.* Suppose, in the contrary, that  $p_t = e$ . By Lemma 3, there exists an exchange set  $c = (c_+, c_-) \in \mathcal{B}$  such that  $e \in (c_+ \cup c_-)$ ,  $c_- \subseteq (M_t \setminus \tilde{M}_t)$ ,  $c_+ \subseteq (\tilde{M}_t \setminus M_t)$ ,  $(M_t \oplus c) \in \mathcal{M}$  and  $(\tilde{M}_t \ominus c) \in \mathcal{M}$ .

Now, we decompose our proof into two cases.

**Case (1):**  $(e \in M_* \wedge e \in c_+) \vee (e \notin M_* \wedge e \in c_-)$ .

Define  $M'_t = \tilde{M}_t \ominus c$  and recall that  $M'_t \in \mathcal{M}$  due to the definition of exchange class.



First, we claim that  $M'_t \neq M_*$ . Suppose that  $e \in M_*$  and  $e \in c_+$ . Then, we see that  $e \notin M'_t$  and hence  $M'_t \neq M_*$ . On the other hand, if  $e \notin M_*$  and  $e \in c_-$ , then  $e \in M'_t$  which also means that  $M'_t \neq M_*$ . Therefore we have  $M'_t \neq M_*$  in either cases.

Next, we apply Lemma 3 by setting  $M = M'_t$  and  $M' = M_*$ . We see that there exists an exchange set  $b \in \mathcal{B}$  such that,  $e \in (b_+ \cup b_-)$ ,  $(M'_t \oplus b) \in \mathcal{M}$  and  $\langle w, \chi_b \rangle \geq \Delta_e > 0$ .

Now, we define vectors  $\mathbf{d} = \chi_{\tilde{M}_t} - \chi_{M_t}$ ,  $\mathbf{d}_1 = \chi_{M'_t} - \chi_{M_t}$  and  $\mathbf{d}_2 = \chi_{M'_t \oplus b} - \chi_{M_t}$ . By the definition of  $M'_t$  and Lemma 3, we see that  $\mathbf{d}_1 = \mathbf{d} - \chi_c$  and  $\mathbf{d}_2 = \mathbf{d}_1 + \chi_b = \mathbf{d} - \chi_c + \chi_b$ .

Then, we claim that  $\|\mathbf{rad}_t \circ (\mathbf{d} - \chi_c)\|_\infty < \frac{\Delta_e}{3 \text{width}(\mathcal{B})}$ . Since  $c_- \subseteq M_t$  and  $c_+ \cap M_t = \emptyset$ , using standard set theoretical manipulations, we can show that  $M_t \setminus \tilde{M}_t = (M_t \setminus M'_t) \cup c_-$ . Similarly, one can show that  $\tilde{M}_t \setminus M_t = (M'_t \setminus M_t) \cup c_+$ . This means that  $((M_t \setminus M'_t) \cup (M'_t \setminus M_t)) \subseteq ((M_t \setminus \tilde{M}_t) \cup (\tilde{M}_t \setminus M_t))$ . Then, applying Lemma Y, we obtain

$$\begin{aligned} \|\mathbf{rad}_t \circ (\mathbf{d} - \chi_c)\|_\infty &= \left\| \mathbf{rad}_t \circ (\chi_{M'_t} - \chi_{M_t}) \right\|_\infty \\ &= \max_{i \in (M_t \setminus M'_t) \cup (M'_t \setminus M_t)} \text{rad}_t(i) \\ &\leq \max_{i \in (M_t \setminus \tilde{M}_t) \cup (\tilde{M}_t \setminus M_t)} \text{rad}_t(i) \\ &= \text{rad}_t(e) < \frac{\Delta_e}{3 \text{width}(\mathcal{B})}. \end{aligned} \quad (11)$$

We claim that  $\|\mathbf{rad}_t \circ \chi_c\|_\infty < \frac{\Delta_e}{3 \text{width}(\mathcal{B})}$ . Recall that, by the definition of  $c$ , we have  $c_+ \subseteq (\tilde{M}_t \setminus M_t)$  and  $c_- \subseteq (M_t \setminus \tilde{M}_t)$ . Hence  $c_+ \cup c_- \subseteq (\tilde{M}_t \setminus M_t) \cup (M_t \setminus \tilde{M}_t)$ . Since  $\chi_c \in [-1, 1]^n$ , we see that

$$\begin{aligned} \|\mathbf{rad}_t \circ \chi_c\|_\infty &= \max_{i \in c_+ \cup c_-} \text{rad}_t(i) \\ &\leq \max_{i \in (\tilde{M}_t \setminus M_t) \cup (M_t \setminus \tilde{M}_t)} \text{rad}_t(i) \\ &= \text{rad}_t(e) < \frac{\Delta_e}{3 \text{width}(\mathcal{B})}. \end{aligned} \quad (12)$$

Next, we claim that  $\mathbf{d} \circ \chi_c = |\chi_c|$ . Recall that  $\chi_c = \chi_{c_+} - \chi_{c_-}$  and  $\mathbf{d} = \chi_{\tilde{M}_t} - \chi_{M_t} = \chi_{\tilde{M}_t \setminus M_t} - \chi_{M_t \setminus \tilde{M}_t}$ . We also notice that  $c_+ \subseteq (\tilde{M}_t \setminus M_t)$  and  $c_- \subseteq (M_t \setminus \tilde{M}_t)$ . This implies that  $c_+ \cap (M_t \setminus \tilde{M}_t) = \emptyset$  and  $c_- \cap (\tilde{M}_t \setminus M_t) = \emptyset$ . Therefore, we have

$$\begin{aligned} \mathbf{d} \circ \chi_c &= (\chi_{\tilde{M}_t \setminus M_t} - \chi_{M_t \setminus \tilde{M}_t}) \circ (\chi_{c_+} - \chi_{c_-}) \\ &= \chi_{\tilde{M}_t \setminus M_t} \circ \chi_{c_+} + \chi_{M_t \setminus \tilde{M}_t} \circ \chi_{c_-} - \chi_{\tilde{M}_t \setminus M_t} \circ \chi_{c_-} - \chi_{M_t \setminus \tilde{M}_t} \circ \chi_{c_+} \\ &= \chi_{\tilde{M}_t \setminus M_t} \circ \chi_{c_+} + \chi_{M_t \setminus \tilde{M}_t} \circ \chi_{c_-} \\ &= \chi_{c_+} + \chi_{c_-} = |\chi_c|. \end{aligned}$$

where the last equality holds since  $c_+ \cap c_- = \emptyset$ .

Now, we bound quantity  $\langle \mathbf{rad}_t, |\mathbf{d}_2| \rangle - \langle \mathbf{rad}_t, |\mathbf{d}| \rangle$  as follows

$$\langle \mathbf{rad}_t, |\mathbf{d}_2| \rangle - \langle \mathbf{rad}_t, |\mathbf{d}| \rangle = \langle \mathbf{rad}_t, |\mathbf{d}_2| - |\mathbf{d}| \rangle = \langle \mathbf{rad}_t, \mathbf{d}_2^2 - \mathbf{d}^2 \rangle \quad (13)$$

$$\begin{aligned} &= \langle \mathbf{rad}_t, (\mathbf{d} - \chi_c + \chi_b)^2 - \mathbf{d}^2 \rangle \\ &= \langle \mathbf{rad}_t, \chi_b^2 + \chi_c^2 - 2\chi_b \circ \chi_c - 2\mathbf{d} \circ \chi_c + 2\mathbf{d} \circ \chi_b \rangle \\ &= \langle \mathbf{rad}_t, \chi_b^2 - \chi_c^2 + 2\chi_b \circ (\mathbf{d} - \chi_c) \rangle \end{aligned} \quad (14)$$

$$\begin{aligned} &= \langle \mathbf{rad}_t, |\chi_b| \rangle - \langle \mathbf{rad}_t, |\chi_c| \rangle - 2 \langle \mathbf{rad}_t, \chi_b \circ (\mathbf{d} - \chi_c) \rangle \\ &= \langle \mathbf{rad}_t, |\chi_b| \rangle - \langle \mathbf{rad}_t, |\chi_c| \rangle - 2 \langle \mathbf{rad}_t \circ (\mathbf{d} - \chi_c), \chi_b \rangle \end{aligned} \quad (15)$$

$$\geq \langle \mathbf{rad}_t, |\chi_b| \rangle - \langle \mathbf{rad}_t, |\chi_c| \rangle - 2 \|\mathbf{rad}_t \circ (\mathbf{d} - \chi_c)\|_\infty \|\chi_b\|_1 \quad (16)$$

$$> \langle \mathbf{rad}_t, |\chi_b| \rangle - \langle \mathbf{rad}_t, |\chi_c| \rangle - \frac{2\Delta_e}{3 \text{width}(\mathcal{B})} \|\chi_b\|_1 \quad (17)$$

$$\geq \langle \mathbf{rad}_t, |\chi_b| \rangle - \langle \mathbf{rad}_t, |\chi_c| \rangle - \frac{2\Delta_e}{3}, \quad (18)$$

where Eq. (13) holds since  $\mathbf{d} \in \{-1, 0, 1\}^n$  and  $\mathbf{d}_2 \in \{-1, 0, 1\}^n$ ; Eq. (14) follows from the claim that  $\mathbf{d} \circ \chi_c = |\chi_c| = \chi_c^2$ ; Eq. (15) and Eq. (16) follow from Lemma 5 and Hölder's inequality; Eq. (17) follows from Eq. (11); and Eq. (18) holds since  $b \in \mathcal{B}$  and  $\|\chi_b\|_1 = |b_+| + |b_-| \leq \text{width}(\mathcal{B})$ .

Applying Lemma 6 by setting  $M' = M'_t \oplus b$  and using the fact that  $\tilde{w}_t(\tilde{M}_t) \geq \tilde{w}_t(M'_t)$ , we have

$$\begin{aligned} \langle \bar{w}_t, \mathbf{d} \rangle + \langle \mathbf{rad}_t, |\mathbf{d}| \rangle &= \langle \bar{w}_t, \chi_{\tilde{M}_t} - \chi_{M_t} \rangle + \langle \mathbf{rad}_t, |\chi_{\tilde{M}_t} - \chi_{M_t}| \rangle \\ &= \tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) \\ &\geq \tilde{w}_t(M'_t) - \tilde{w}_t(M_t) \\ &= \langle \bar{w}_t, \chi_{M'_t} - \chi_{M_t} \rangle + \langle \mathbf{rad}_t, |\chi_{M'_t} - \chi_{M_t}| \rangle \\ &= \langle \bar{w}_t, \mathbf{d}_2 \rangle + \langle \mathbf{rad}_t, |\mathbf{d}_2| \rangle \\ &= \langle \bar{w}_t, \mathbf{d} \rangle - \langle \bar{w}_t, \chi_c \rangle + \langle \bar{w}_t, \chi_b \rangle + \langle \mathbf{rad}_t, |\mathbf{d}_2| \rangle, \end{aligned}$$

where the last equality follows from the fact that  $\mathbf{d}_2 = \mathbf{d} - \chi_c + \chi_b$ . Rearranging the above inequality, we obtain

$$\begin{aligned} \langle \bar{w}_t, \chi_c \rangle &\geq \langle \bar{w}_t, \chi_b \rangle + \langle \mathbf{rad}_t, |\mathbf{d}_2| \rangle - \langle \mathbf{rad}_t, |\mathbf{d}| \rangle \\ &\geq \langle \bar{w}_t, \chi_b \rangle + \langle \mathbf{rad}_t, |\chi_b| \rangle - \langle \mathbf{rad}_t, |\chi_c| \rangle - \frac{2\Delta_e}{3} \end{aligned} \quad (19)$$

$$> \langle \bar{w}_t, \chi_b \rangle - \langle \mathbf{rad}_t, \chi_c \rangle - \frac{2\Delta_e}{3} \quad (20)$$

$$> \langle \bar{w}_t, \chi_b \rangle - \frac{\Delta_e}{3} - \frac{2\Delta_e}{3} \quad (21)$$

$$= \langle \bar{w}_t, \chi_b \rangle - \Delta_e \geq 0, \quad (22)$$

where Eq. (19) uses Eq. (18); Eq. (20) follows from the assumption that event  $\xi$  occurs and Lemma X; and Eq. (20) holds since Eq. (12).

We have shown that  $\langle \bar{w}_t, \chi_c \rangle > 0$ . Now we can bound  $\bar{w}_t(M'_t)$  as follows

$$\bar{w}_t(M'_t) = \langle \bar{w}_t, \chi_{M'_t} \rangle = \langle \bar{w}_t, \chi_{M_t} + \chi_c \rangle = \langle \bar{w}_t, \chi_{M_t} \rangle + \langle \bar{w}_t, \chi_c \rangle > \langle \bar{w}_t, \chi_{M_t} \rangle = \bar{w}_t(M_t).$$

However, the definition of  $M_t$  ensures that  $M_t = \arg \max_{M \in \mathcal{M}} \bar{w}_t(M)$ , i.e.  $\bar{w}_t(M_t) \geq \bar{w}_t(M'_t)$ . Contradiction.

**Case (2):**  $(e \in M_* \wedge e \in c_-) \vee (e \notin M_* \wedge e \in c_+)$ .

First, we claim that  $\tilde{M}_t \neq M_*$ . Suppose that  $e \in M_*$  and  $e \in c_-$ . Then, we see that  $e \notin \tilde{M}_t$ , which implies that  $\tilde{M}_t \neq M_*$ . If  $e \notin M_*$  and  $e \in c_+$ , then  $e \in \tilde{M}_t$ , which also implies that  $\tilde{M}_t \neq M_*$ . Therefore we have  $\tilde{M}_t \neq M_*$  in either cases.

Hence, by Lemma 3, there exists an exchange set  $b = (b_+, b_-) \in \mathcal{B}$  such that  $e \in (b_+ \cup b_-)$ ,  $b_- \subseteq (\tilde{M}_t \setminus M_*)$ ,  $b_+ \subseteq (M_* \setminus \tilde{M}_t)$  and  $(\tilde{M}_t \oplus b) \in \mathcal{M}$ . Lemma 3 also indicates that  $\langle \bar{w}_t, \chi_b \rangle \geq \Delta_e > 0$ .

Next, we define vectors  $\mathbf{d} = \chi_{\tilde{M}_t} - \chi_{M_t}$  and  $\mathbf{d}_1 = \chi_{\tilde{M}_t \oplus b} - \chi_{M_t}$ . Notice that Lemma 3 gives that  $\mathbf{d}_1 = \mathbf{d} + \mathbf{b}$ . Then, we apply Lemma 4 by setting  $M = M_t$  and  $M' = \tilde{M}_t$ . This shows that

$$\|\mathbf{rad}_t \circ \mathbf{d}\|_\infty \leq \max_{i: (\tilde{M}_t \setminus M_t) \cup (M_t \setminus \tilde{M}_t)} \text{rad}_t(i) = \text{rad}_t(e) < \frac{\Delta_e}{3}. \quad (23)$$

Now, we bound quantity  $\langle \bar{\mathbf{w}}_t, \mathbf{d}_1 \rangle + \langle \mathbf{rad}_t, |\mathbf{d}_1| \rangle - \langle \bar{\mathbf{w}}_t, \mathbf{d} \rangle - \langle \mathbf{rad}_t, |\mathbf{d}| \rangle$  as follows

$$\begin{aligned} \langle \bar{\mathbf{w}}_t, \mathbf{d}_1 \rangle + \langle \mathbf{rad}_t, |\mathbf{d}_1| \rangle - \langle \bar{\mathbf{w}}_t, \mathbf{d} \rangle - \langle \mathbf{rad}_t, |\mathbf{d}| \rangle &= \langle \bar{\mathbf{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\mathbf{d}_1| - |\mathbf{d}| \rangle \\ &= \langle \bar{\mathbf{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, \mathbf{d}_1^2 - \mathbf{d}^2 \rangle \end{aligned} \quad (24)$$

$$= \langle \bar{\mathbf{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, 2\mathbf{d} \circ \boldsymbol{\chi}_b + \boldsymbol{\chi}_b^2 \rangle \quad (25)$$

$$= \langle \bar{\mathbf{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, \boldsymbol{\chi}_b^2 \rangle + 2 \langle \mathbf{rad}_t \circ \mathbf{d}, \boldsymbol{\chi}_b \rangle \quad (26)$$

$$\geq \langle \mathbf{w}, \boldsymbol{\chi}_b \rangle - 2 \langle \mathbf{rad}_t \circ \mathbf{d}, \boldsymbol{\chi}_b \rangle \quad (27)$$

$$\geq \langle \mathbf{w}, \boldsymbol{\chi}_b \rangle - 2 \|\mathbf{rad}_t \circ \mathbf{d}\|_\infty \|\boldsymbol{\chi}_b\|_1 \quad (28)$$

$$> \langle \mathbf{w}, \boldsymbol{\chi}_b \rangle - \frac{2\Delta_e}{3} \quad (29)$$

where Eq. (24) follows from the fact that  $\mathbf{d}_1 \in \{-1, 0, 1\}^n$  and  $\mathbf{d} \in \{-1, 0, 1\}^n$ ; Eq. (25) holds since  $\mathbf{d}_1 = \mathbf{d} + \boldsymbol{\chi}_b$ ; Eq. (26) follows from the assumption that  $\xi$  occurs and Lemma X; Eq. (27) follows from Lemma 5 and Hölder's inequality; and Eq. (28) is due to Eq. (23).

Therefore, we have proved that  $\langle \bar{\mathbf{w}}_t, \mathbf{d} \rangle + \langle \mathbf{rad}_t, |\mathbf{d}| \rangle < \langle \bar{\mathbf{w}}_t, \mathbf{d}_1 \rangle + \langle \mathbf{rad}_t, |\mathbf{d}_1| \rangle$ . However, Lemma 6 shows that

$$\begin{aligned} \langle \bar{\mathbf{w}}_t, \mathbf{d} \rangle + \langle \mathbf{rad}_t, |\mathbf{d}| \rangle &= \langle \bar{\mathbf{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t}| \rangle \\ &= \tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) \\ &\geq \tilde{w}_t(\tilde{M}_t \oplus b) - \tilde{w}_t(M_t) \\ &= \langle \bar{\mathbf{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t \oplus b} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_{\tilde{M}_t \oplus b} - \boldsymbol{\chi}_{M_t}| \rangle \\ &= \langle \bar{\mathbf{w}}_t, \mathbf{d}_1 \rangle + \langle \mathbf{rad}_t, |\mathbf{d}_1| \rangle. \end{aligned}$$

This is a contradiction and therefore  $p_t \neq e$ . □

### 3.4 Proof of Theorem 1

Theorem 1 is now a straightforward corollary of Lemma 9 and Lemma 10.

*Proof.* Lemma 7 indicates that the event  $\xi \triangleq \bigcap_{t=1}^\infty \xi_t$  occurs with probability at least  $1 - \delta$ . In the rest of the proof, we shall assume that this event holds.

By Lemma 9 and the assumption on  $\xi$ , we see that  $\text{Out} = M_*$ . Next, we focus on bounding  $T$ .

Let  $T_e$  denote the total number of pull of arm  $e \in [n]$ . Define

$$t_e^* = \left\lceil \frac{275}{B^2} \cdot \frac{\text{width}(\mathcal{B})^2}{\Delta_e^2} \log \left( \frac{1}{B^2} \cdot \frac{4n}{\delta} \sum_{i \in [n]} \frac{\text{width}(\mathcal{B})^2}{\Delta_i^2} \right) \right\rceil.$$

We claim that, for any  $e \in [n]$ , we have  $T_e \leq \lceil t_e^* \rceil$ . Suppose that  $T_e \geq \lceil t_e^* \rceil + 1$ . Then, consider the round  $t$  that such that  $p_t = e$  and  $T_e(t) = \lceil t_e^* \rceil + 1$ . Notice that  $t \leq T_e(t)$ . Then, by straightforward computations, we obtain

$$\text{rad}_t(e) = B \sqrt{\frac{2 \log \left( \frac{4nt^2}{\delta} \right)}{T_e(t)}} \leq B \sqrt{\frac{2 \log \left( \frac{4n(t_e^*)^2}{\delta} \right)}{t_e^*}} < \frac{\Delta_e}{3 \text{width}(\mathcal{B})}.$$

However, by Lemma 10, we see that  $p_t \neq e$ . This is a contradiction, and therefore the claim that  $T_e \leq \lceil t_e^* \rceil$  is true.

Then, the total number of samples  $T$  can be bounded by

$$T = \sum_{e \in [n]} T_e \leq \sum_{e \in [n]} \lceil t_e^* \rceil \leq O \left( \frac{1}{B^2} \sum_{e \in [n]} \frac{\text{width}(\mathcal{B})^2}{\Delta_e^2} \log \left( \frac{1}{B^2} \cdot \frac{n}{\delta} \sum_{e \in [n]} \frac{\text{width}(\mathcal{B})^2}{\Delta_e^2} \right) \right).$$

□

## 4 Proof of Lower Bounds

**Lemma 11.**

$$\Delta_e = \min_{b: e \in b, b \in \mathcal{B}_{\text{opt}}} w(b).$$

*Proof.*

□

*Proof.* Fix  $\delta > 0$ ,  $\mathbf{w} = \{w(1), \dots, w(n)\}$  and a  $\delta$ -correct policy  $\mathbb{A}$ . For each  $e \in [n]$ , assume that the reward distribution is given by  $\varphi_e = \mathcal{N}(w(e), 1)$ . For any  $e \in [n]$ , let  $T_e$  denote the number of trials of arm  $e$  used by algorithm  $\mathbb{A}$ . In the rest of the proof, we will show that for any  $e \in [n]$ , the number of trials of arm  $e$  is lower-bounded by

$$\mathbb{E}[T_e] \geq \frac{1}{16\Delta_e^2} \log(1/4\delta). \quad (30)$$

Notice that the theorem follows immediately by summing up Eq. (30) for all  $e \in [n]$ .

Fix an arm  $e \in [n]$ . We now focus on proving Eq. (30). Consider two hypothesis  $H_0$  and  $H_1$ . Under hypothesis  $H_0$ , all reward distributions are same with our assumption before

$$H_0 : \varphi_l = \mathcal{N}(w(l), 1) \quad \text{for all } l \in [n].$$

Under hypothesis  $H_1$ , we change the means of reward distributions such that

$$H_1 : \varphi_e = \begin{cases} \mathcal{N}(w(e) - 2\Delta_e, 1) & \text{if } e \in M_* \\ \mathcal{N}(w(e) + 2\Delta_e, 1) & \text{if } e \notin M_* \end{cases} \quad \text{and } \varphi_l = \mathcal{N}(w(l), 1) \quad \text{for all } l \neq e.$$

Define  $M_e$  be the “next-to-optimal” set as follows

$$M_e = \begin{cases} \arg \max_{M \in \mathcal{M}: e \in M} w(M) & \text{if } e \notin M_*, \\ \arg \max_{M \in \mathcal{M}: e \notin M} w(M) & \text{if } e \in M_*. \end{cases}$$

By definition of  $\Delta_e$ , we know that  $w(M_*) - w(M_e) = \Delta_e$ .

Let  $\mathbf{w}_0$  and  $\mathbf{w}_1$  be expected reward vectors under  $H_0$  and  $H_1$  respectively. Notice that  $w_0(M_*) - w_0(M_e) = \Delta_e > 0$ . On the other hand,  $w_1(M_*) - w_1(M_e) = -\Delta < 0$ . This means that under  $H_1$ ,  $M_*$  is not the optimal set. For  $l \in \{0, 1\}$ , we use  $\mathbb{E}_l$  and  $\Pr_l$  to denote the expectation and probability, respectively, under the hypothesis  $H_l$ .

Define  $\theta = 4\delta$ . Define

$$t_e^* = \frac{1}{16\Delta_e^2} \log \left( \frac{1}{\theta} \right). \quad (31)$$

Recall that  $T_e$  denotes the total number of samples of arm  $e$ . Define the event  $\mathcal{A} = \{T_e \leq 4t_e^*\}$ .

First, we show that  $\Pr_0[\mathcal{A}] \geq 3/4$ . This can be proved by Markov inequality as follows.

$$\Pr_0[T_e > 4t_e^*] \leq \frac{\mathbb{E}_0[T_e]}{4t_e^*}$$

$$= \frac{t_e^*}{4t_e^*} = \frac{1}{4}.$$

Let  $X_1, \dots, X_{T_e}$  denote the sequence of reward outcomes of arm  $e$ . We define  $K_t(e)$  as the sum of outcomes of arm  $e$  up to round  $t$ , i.e.  $K_t(e) = \sum_{i \in [t]} X_i$ . Next, we define the event

$$\mathcal{C} = \left\{ \max_{1 \leq t \leq 4t_e^*} |K_t(e) - t \cdot w(e)| < \sqrt{t_e^* \log(1/\theta)} \right\}.$$

We now show that  $\Pr_0[\mathcal{C}] \geq 3/4$ . First, notice that  $K_t(e) - p_e t$  is a martingale under  $H_0$ . Then, by Kolmogorov's inequality, we have

$$\begin{aligned} \Pr_0 \left[ \max_{1 \leq t \leq 4t_e^*} |K_t(e) - t \cdot w(e)| \geq \sqrt{t_e^* \log(1/\theta)} \right] &\leq \frac{\mathbb{E}_0[(K_{4t_e^*}(e) - 4w(e)t_e^*)^2]}{t_e^* \log(1/\theta)} \\ &= \frac{4t_e^*}{t_e^* \log(1/\theta)} \\ &< \frac{1}{4}, \end{aligned}$$

where the second inequality follows from the fact that  $\mathbb{E}_0[(K_{4t_e^*}(e) - 4w(e)t_e^*)^2] = 4t_e^*$ ; the last inequality follows since  $\theta < e^{-16}$ .

Then, we define the event  $\mathcal{B}$  as the event that the algorithm eventually returns  $M_*$ , i.e.

$$\mathcal{B} = \{\text{Out} = M_*\}.$$

Since the probability of error of the algorithm is smaller than  $\delta < 1/4$ , we have  $\Pr_0[\mathcal{B}] \geq 3/4$ . Define  $\mathcal{S}$  be  $\mathcal{S} = \mathcal{A} \cap \mathcal{B} \cap \mathcal{C}$ . Then, by union bound, we have  $\Pr_0[\mathcal{S}] \geq 1/4$ .

Now, we show that if  $\mathbb{E}_0[T_e] \leq t_e^*$ , then  $\Pr_1[\mathcal{B}] \geq \delta$ . Let  $W$  be the history of the sampling process until the algorithm stops (including the sequence of arms chosen at each time and the sequence of observed outcomes). Define the likelihood function  $L_l$  as

$$L_l(w) = p_l(W = w),$$

where  $p_l$  is the probability density function under hypothesis  $H_l$ . Let  $K$  be the shorthand of  $K_e(T_e)$ .

Assume that the event  $\mathcal{S}$  occurred. We will bound the likelihood ratio  $L_1(W)/L_0(W)$  under this assumption. To do this, we divide our analysis into two different cases.

**Case (1):**  $e \notin M_*$ . In this case, the reward distribution of arm  $e$  under  $H_1$  is a Gaussian distribution with mean  $p_e + 2\Delta_e$  and variance 1. Recall that the probability density function of a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$  is given by  $\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ . Hence, we have

$$\begin{aligned} \frac{L_1(W)}{L_0(W)} &= \prod_{i=1}^{T_e} \exp\left(\frac{-(X_i - w(e) - 2\Delta_e)^2 + (X_i - w(e))^2}{2}\right) \\ &= \prod_{i=1}^{T_e} \exp(\Delta_e(2X_i - 2w(e)) - 2\Delta_e^2) \\ &= \exp(\Delta_e(2K - 2w(e)T_e) - 2\Delta_e^2 T_e) \\ &= \exp(\Delta_e(2K - 2w(e)T_e)) \exp(-2\Delta_e^2 T_e). \end{aligned} \tag{32}$$

Next, we bound each individual term on the right-hand side of Eq. (32). We begin with bounding the second term of Eq. (32)

$$\exp(-2\Delta_e^2 T_e) \geq \exp(-8\Delta_e^2 t_e^*) \tag{33}$$

$$= \exp\left(-\frac{8}{16}\log(1/\theta)\right) \quad (34)$$

$$= \theta^{1/2}, \quad (35)$$

where Eq. (103) follows from the assumption that event  $\mathcal{S}$  occurred, which implies that event  $\mathcal{A}$  occurred and therefore  $T_e \leq 4t_e^*$ ; Eq. (104) follows from the definition of  $t_e^*$ .

Then, we bound the first term on the right-hand side of Eq. (32) as follows

$$\exp(\Delta_e(2K - 2w(e)T_e)) \geq \exp\left(-2\Delta_e\sqrt{t_e^*\log(1/\theta)}\right) \quad (36)$$

$$= \exp\left(-\frac{2}{\sqrt{4}}\log(1/\theta)\right) \quad (37)$$

$$= \theta^{1/2}, \quad (38)$$

where Eq. (106) follows from the assumption that event  $\mathcal{S}$  occurred, which implies that event  $\mathcal{C}$  and therefore  $|2K - 2w(e)T_e| \leq \sqrt{t_e^*\log(1/\theta)}$ ; Eq. (107) follows from the definition of  $t_e^*$ .

Combining Eq. (105) and Eq. (108), we can bound  $L_1(W)/L_0(W)$  for this case as follows

$$\frac{L_1(W)}{L_0(W)} \geq \theta. \quad (39)$$

(End of Case (1).)

**Case (2):**  $e \in M_*$ . In this case, we know that the mean reward of arm  $e$  under  $H_1$  is  $p_e - 2\Delta$ . Therefore, the likelihood ratio  $L_1(W)/L_0(W)$  is given by

$$\begin{aligned} \frac{L_1(W)}{L_0(W)} &= \prod_{i=1}^{T_e} \exp\left(\frac{-(X_i - w(e) + 2\Delta_e)^2 + (X_i - w(e))^2}{2}\right) \\ &= \prod_{i=1}^{T_e} \exp(\Delta_e(2w(e) - 2X_i) - 2\Delta_e^2) \\ &= \exp(\Delta_e(2w(e)T_e - 2K)) \exp(-2\Delta_e^2 T_e). \end{aligned} \quad (40)$$

Notice that the right-hand side of Eq. (40) differs from Eq. (32) only in its first term. Now, we bound the first term as follows

$$\exp(\Delta_e(2K - 2w(e)T_e)) \geq \exp\left(-2\Delta_e\sqrt{t_e^*\log(1/\theta)}\right) \quad (41)$$

$$= \exp\left(-\frac{2}{4}\log(1/\theta)\right) \quad (42)$$

$$= \theta^{1/2}, \quad (43)$$

where the inequalities hold due to reasons similar to Case (1): Eq. (41) follows from the assumption that event  $\mathcal{S}$  occurred, which implies that event  $\mathcal{C}$  and therefore  $|2K - 2w(e)T_e| \leq \sqrt{t_e^*\log(1/\theta)}$ ; Eq. (42) follows from the definition of  $t_e^*$ .

Combining Eq. (105) and Eq. (108), we can obtain the same bound of  $L_1(W)/L_0(W)$  as in Eq. (39), i.e.  $L_1(W)/L_0(W) \geq \theta$ .

(End of Case (2).)

At this point, we have proved that, if the event  $\mathcal{S}$  occurred, then the bound of likelihood ratio Eq. (39) holds,

i.e.  $\frac{L_1(W)}{L_0(W)} \geq \theta$ . Hence, we have

$$\begin{aligned} \frac{L_1(W)}{L_0(W)} &\geq \theta \\ &= 4\delta. \end{aligned} \tag{44}$$

Define  $1_S$  as the indicator variable of event  $\mathcal{S}$ , i.e.  $1_S = 1$  if and only if  $\mathcal{S}$  occurs and otherwise  $1_S = 0$ . Then, we have

$$\frac{L_1(W)}{L_0(W)} 1_S \geq 4\delta 1_S$$

holds regardless the occurrence of event  $\mathcal{S}$ . Therefore, we can obtain

$$\begin{aligned} \Pr_1[\mathcal{B}] &\geq \Pr_1[\mathcal{S}] = \mathbb{E}_1[1_S] \\ &= \mathbb{E}_0 \left[ \frac{L_1(W)}{L_0(W)} 1_S \right] \\ &\geq 4\delta \mathbb{E}_0[1_S] \\ &= 4\delta \Pr_0[\mathcal{S}] > \delta. \end{aligned}$$

Now we have proved that, if  $\mathbb{E}_0[T_e] \leq t_e^*$ , then  $\Pr_1[\mathcal{B}] > \delta$ . This means that, if  $\mathbb{E}_0[T_e] \leq t_e^*$ , algorithm  $\mathbb{A}$  will choose  $M_*$  as the output with probability at least  $\delta$ , under hypothesis  $H_1$ . However, under  $H_1$ , we have shown that  $M_*$  is not the optimal set since  $w_1(M_e) > w_1(M_*)$ . Therefore, algorithm  $\mathbb{A}$  has a probability of error larger than  $\delta$  under  $H_1$ . This contradicts to the assumption that algorithm  $\mathbb{A}$  is a  $\delta$ -correct algorithm. Hence, we must have  $\mathbb{E}_0[T_e] > t_e^* = \frac{1}{16\Delta_e^2} \log(1/4\delta)$ .  $\square$

*Proof.* Fix  $\delta > 0$ ,  $\mathbf{w} \in \mathbb{R}^n$ , diff-set  $b = (b_+, b_-)$  and a  $\delta$ -correct algorithm  $\mathbb{A}$ . Assume that  $\varphi_e(e) = \mathcal{N}(w(e), 1)$  for all  $e \in [n]$ .

We define three hypotheses  $H_0$ ,  $H_1$  and  $H_2$ . Under hypothesis  $H_0$ , the reward distribution

$$H_0 : \varphi_l = \mathcal{N}(w(l), 1) \quad \text{for all } l \in [n].$$

Under hypothesis  $H_1$ , the mean reward of each arm is given by

$$H_1 : \varphi_e = \begin{cases} \mathcal{N}\left(w(e) + 2\frac{w(b)}{|b_-|}, 1\right) & \text{if } e \in b_-, \\ \mathcal{N}(w(e), 1) & \text{if } e \notin b_-. \end{cases}$$

And under hypothesis  $H_2$ , the mean reward of each arm is given by

$$H_2 : \varphi_e = \begin{cases} \mathcal{N}\left(w(e) - 2\frac{w(b)}{|b_-|}, 1\right) & \text{if } e \in b_+, \\ \mathcal{N}(w(e), 1) & \text{if } e \notin b_+. \end{cases}$$

Since  $b \in \mathcal{B}_{\text{opt}}$ , it is clear that  $\neg b \prec M_*$ . Hence we define  $M = M_* \ominus b$ . Let  $w_0, w_1$  and  $w_2$  be the expected reward vectors under  $H_0, H_1$  and  $H_2$  respectively. It is easy to check that  $w_1(M_*) - w_1(M) = -w(b) < 0$  and  $w_2(M_*) - w_2(M) = -w(b) < 0$ . This means that under  $H_1$  or  $H_2$ ,  $M_*$  is not the optimal set. Further, for  $l \in \{0, 1, 2\}$ , we use  $\mathbb{E}_l$  and  $\Pr_l$  to denote the expectation and probability, respectively, under the hypothesis  $H_l$ . In addition, let  $W$  be the history of the sampling process until algorithm  $\mathbb{A}$  stops. Define the likelihood function  $L_l$  as

$$L_l(w) = p_l(W = w),$$

where  $p_l$  is the probability density function under  $H_l$ .

Define  $\theta = 4\delta$ . Let  $T_{b_-}$  and  $T_{b_+}$  denote the number of trials of arms belonging to  $b_-$  and  $b_+$ , respectively. In the rest of the proof, we will bound  $\mathbb{E}_0[T_{b_-}]$  and  $\mathbb{E}_0[T_{b_+}]$  individually.

**Part (1): Lower bound of  $\mathbb{E}_0[T_{b_-}]$ .** In this part, we will show that  $\mathbb{E}_0[T_{b_-}] \geq t_{b_-}^*$ , where we define  $t_{b_-}^* = \frac{|b_-|^2}{16w(b)^2} \log(1/\theta)$ .

Consider the complete sequence of sampling process by algorithm A. Formally, let  $W = \{(\tilde{I}_1, \tilde{X}_1), \dots, (\tilde{I}_T, \tilde{X}_T)\}$  be the sequence of all trials by algorithm A, where  $\tilde{I}_i$  denotes the arm played in  $i$ -th trial and  $\tilde{X}_i$  be the reward outcome of  $i$ -th trial. Then, consider the subsequence  $W_1$  of  $W$  which consists all the trials of arms in  $b_-$ . Specifically, we write  $W = \{(I_1, X_1), \dots, (I_{T_{b_-}}, X_{T_{b_-}})\}$  such that  $W_1$  is a subsequence of  $W$  and  $I_i \in b_-$  for all  $i$ .

Next, we define several random events in a way similar to the proof of Theorem 2. Define event  $\mathcal{A}_1 = \{T_{b_-} \leq 4t_{b_-}^*\}$ . Define event

$$\mathcal{C}_1 = \left\{ \max_{1 \leq t \leq 4t_{b_-}^*} \left| \sum_{i=1}^t X_i - \sum_{i=1}^t w(I_i) \right| < \sqrt{t_{b_-}^* \log(1/\theta)} \right\}.$$

Define event

$$\mathcal{B} = \{\text{Out} = M_*\}. \quad (45)$$

Define event  $\mathcal{S}_1 = \mathcal{A}_1 \cap \mathcal{B} \cap \mathcal{C}_1$ . Then, we bound the probability of events  $\mathcal{A}_1$ ,  $\mathcal{B}$ ,  $\mathcal{C}_1$  and  $\mathcal{S}_1$  under  $H_0$  using methods similar to Theorem 2. First, we show that  $\Pr_0[\mathcal{A}_1] \geq 3/4$ . This can be proved by Markov inequality as follows.

$$\begin{aligned} \Pr_0[T_{b_-} > 4t_{b_-}^*] &\leq \frac{\mathbb{E}_0[T_{b_-}]}{4t_{b_-}^*} \\ &= \frac{t_{b_-}^*}{4t_{b_-}^*} = \frac{1}{4}. \end{aligned}$$

Next, we show that  $\Pr_0[\mathcal{C}_1] \geq 3/4$ . Notice that the sequence  $\left\{ \sum_{i=1}^t X_i - \sum_{i=1}^t p_{I_i} \right\}_{t \in [4t_{b_-}^*]}$  is a martingale.

Hence, by Kolmogorov's inequality, we have

$$\begin{aligned} \Pr_0 \left[ \max_{1 \leq t \leq 4t_{b_-}^*} \left| \sum_{i=1}^t X_i - \sum_{i=1}^t w(I_i) \right| \geq \sqrt{t_{b_-}^* \log(1/\theta)} \right] &\leq \frac{\mathbb{E}_0 \left[ \left( \sum_{i=1}^{4t_{b_-}^*} X_i - \sum_{i=1}^{4t_{b_-}^*} w(I_i) \right)^2 \right]}{t_{b_-}^* \log(1/\theta)} \\ &= \frac{4t_{b_-}^*}{t_{b_-}^* \log(1/\theta)} \\ &< \frac{1}{4}, \end{aligned}$$

where the second inequality follows from the fact that all reward distributions have unit variance and hence  $\mathbb{E}_0 \left[ \left( \sum_{i=1}^{4t_{b_-}^*} X_i - \sum_{i=1}^{4t_{b_-}^*} p_{I_i} \right)^2 \right] = 4t_{b_-}^*$ ; the last inequality follows since  $\theta < e^{-16}$ . Last, since algorithm A is a  $\delta$ -correct algorithm with  $\delta < 1/4$ . Therefore, it is easy to see that  $\Pr_0[\mathcal{B}] \geq 3/4$ . And by union bound, we have

$$\Pr_0[\mathcal{S}_1] \geq 1/4.$$

Now, we show that if  $\mathbb{E}_0[T_{b_-}] \leq t_{b_-}^*$ , then  $\Pr_1[\mathcal{B}] \geq \delta$ . Assume that the event  $\mathcal{S}_1$  occurred. We bound the



likelihood ratio  $L_1(W)/L_0(W)$  under this assumption as follows

$$\begin{aligned}
\frac{L_1(W)}{L_0(W)} &= \prod_{i=1}^{T_{b_-}} \exp \left( \frac{-\left(X_i - w(I_i) - \frac{2w(b)}{|b_-|}\right)^2 + (X_i - w(I_i))^2}{2} \right) \\
&= \prod_{i=1}^{T_{b_-}} \exp \left( \frac{w(b)}{|b_-|} (2X_i - 2w(I_i)) - \frac{2w(b)^2}{|b_-|^2} \right) \\
&= \exp \left( \frac{w(b)}{|b_-|} \left( \sum_{i=1}^{T_{b_-}} 2X_i - 2w(I_i) \right) - \frac{2w(b)^2}{|b_-|^2} T_{b_-} \right) \\
&= \exp \left( \frac{w(b)}{|b_-|} \left( \sum_{i=1}^{T_{b_-}} 2X_i - 2w(I_i) \right) \right) \exp \left( -\frac{2w(b)^2}{|b_-|^2} T_{b_-} \right). \tag{46}
\end{aligned}$$

Then, we bound each term on the right-hand side of Eq. (46). First, we bound the second term of Eq. (46).

$$\exp \left( -\frac{2w(b)^2}{|b_-|^2} T_{b_-} \right) \geq \exp \left( -\frac{2w(b)^2}{|b_-|^2} 4t_b^* \right) \tag{47}$$

$$= \exp \left( -\frac{8}{16} \log(1/\theta) \right) \tag{48}$$

$$= \theta^{1/2}, \tag{49}$$

where Eq. (47) follows from the assumption that events  $\mathcal{S}_1$  and  $\mathcal{A}_1$  occurred and therefore  $T_{b_-} \leq 4t_b^*$ ; Eq. (48) follows from the definition of  $t_b^*$ . Next, we bound the first term of Eq. (46) as follows

$$\exp \left( \frac{w(b)}{|b_-|} \left( \sum_{i=1}^{T_{b_-}} 2X_i - 2w(I_i) \right) \right) \geq \exp \left( -\frac{2w(b)}{|b_-|} \sqrt{t_b^* \log(1/\theta)} \right) \tag{50}$$

$$= \exp \left( -\frac{2}{4} \log(1/\theta) \right) \tag{51}$$

$$= \theta^{1/2}, \tag{52}$$

where Eq. (50) follows since event  $\mathcal{S}_1$  and  $\mathcal{C}_1$  occurred and therefore  $|2K - 2p_e T_e| \leq \sqrt{t_e^* \log(1/\theta)}$ ; Eq. (51) follows from the definition of  $t_b^*$ .

Hence, if event  $\mathcal{S}_1$  occurred, we can bound the likelihood ratio as follows

$$\frac{L_1(W)}{L_0(W)} \geq \theta = 4\delta. \tag{53}$$

Let  $1_{\mathcal{S}_1}$  denote the indicator variable of event  $\mathcal{S}_1$ . Then, we have  $\frac{L_1(W)}{L_0(W)} 1_{\mathcal{S}_1} \geq 4\delta 1_{\mathcal{S}_1}$ . Therefore, we can bound  $\Pr_1[\mathcal{B}]$  as follows

$$\begin{aligned}
\Pr_1[\mathcal{B}] &\geq \Pr_1[\mathcal{S}_1] = \mathbb{E}_1[1_{\mathcal{S}_1}] \\
&= \mathbb{E}_0 \left[ \frac{L_1(W)}{L_0(W)} 1_{\mathcal{S}_1} \right] \\
&\geq 4\delta \mathbb{E}_0[1_{\mathcal{S}_1}] \\
&= 4\delta \Pr_0[\mathcal{S}_1] > \delta. \tag{54}
\end{aligned}$$

This means that, if  $\mathbb{E}_0[T_{b_-}] \leq t_{b_-}^*$ , then, under  $H_1$ , the probability of algorithm  $\mathbb{A}$  returning  $M_*$  as output is at least  $\delta$ . But  $M_*$  is not the optimal set under  $H_1$ . Hence this contradicts to the assumption that  $\mathbb{A}$  is a  $\delta$ -correct algorithm. Hence we have proved that

$$\mathbb{E}_0[T_{b_-}] \geq t_{b_-}^* = \frac{|b_-|^2}{16w(b)^2} \log(1/4\delta). \quad (55)$$

(End of Part (1).)

**Part (2): Lower bound of  $\mathbb{E}_0[T_{b_+}]$ .** In this part, we will show that  $\mathbb{E}_0[T_{b_+}] \geq t_{b_+}^*$ , where we define  $t_{b_+}^* = \frac{|b_+|^2}{16w(b)^2} \log(1/\theta)$ . The arguments used in this part are similar to that of Part (1). Hence, we will omit the redundant parts and highlight the differences.

Recall that we have defined that  $W$  to be the history of all trials by algorithm  $\mathbb{A}$ . We define  $W$  be the subsequence of  $\tilde{S}$  which contains the trials of arms belonging to  $b_+$ . We write  $S_2 = \{(J_1, Y_1), \dots, (J_{T_{b_+}}, Y_{T_{b_+}})\}$ , where  $J_i$  is  $i$ -th played arm in sequence  $S_2$  and  $Y_i$  is the associated reward outcome.

We define the random events  $\mathcal{A}_2$  and  $\mathcal{C}_2$  similar to Part (1). Specifically, we define

$$\mathcal{A}_2 = \{T_{b_+} \leq 4t_{b_+}^*\} \quad \text{and} \quad \mathcal{C}_2 = \left\{ \max_{1 \leq t \leq 4t_{b_+}^*} \left| \sum_{i=1}^t Y_i - \sum_{i=1}^t w(J_i) \right| < \sqrt{t_{b_+}^* \log(1/\theta)} \right\}.$$

Using the similar arguments, we can show that  $\Pr_0[\mathcal{A}_2] \geq 3/4$  and  $\Pr_0[\mathcal{C}_2] \geq 3/4$ . Define event  $\mathcal{S}_2 = \mathcal{A}_2 \cap \mathcal{B} \cap \mathcal{C}_2$ , where  $\mathcal{B}$  is defined in Eq. (45). By union bound, we see that

$$\Pr_0[\mathcal{S}_2] \geq 1/4.$$

Then, we show that if  $\mathbb{E}_0[T_{b_+}] \leq t_{b_+}^*$ , then  $\Pr_2[\mathcal{B}] \geq \delta$ . We bound likelihood ratio  $L_2(W)/L_0(W)$  under the assumption that  $\mathcal{S}_2$  occurred as follows

$$\begin{aligned} \frac{L_2(W)}{L_0(W)} &= \prod_{i=1}^{T_{b_+}} \exp \left( \frac{-\left(Y_i - w(J_i)\right) + \frac{2w(b)}{|b_-|}}{2} + \frac{(Y_i - w(J_i))^2}{2} \right) \\ &= \prod_{i=1}^{T_{b_+}} \exp \left( \frac{w(b)}{|b_+|} (2w(J_i) - 2Y_i) - \frac{2w(b)^2}{|b_+|^2} \right) \\ &= \exp \left( \frac{w(b)}{|b_+|} \left( \sum_{i=1}^{T_{b_+}} 2w(J_i) - 2Y_i \right) - \frac{2w(b)^2}{|b_+|^2} T_{b_+} \right) \\ &= \exp \left( \frac{w(b)}{|b_+|} \left( \sum_{i=1}^{T_{b_+}} 2w(J_i) - 2Y_i \right) \right) \exp \left( -\frac{2w(b)^2}{|b_+|^2} T_{b_+} \right) \\ &\geq \theta \\ &= 4\delta, \end{aligned} \quad (56)$$

where Eq. (56) can be obtained using same method as in Part (1) as well as the assumption that  $\mathcal{S}_2$  occurred. Next, similar to the derivation in Eq. (54), we see that

$$\Pr_2[\mathcal{B}] \geq \Pr_2[\mathcal{S}_2] = \mathbb{E}_2[1_{\mathcal{S}_2}] = \mathbb{E}_0 \left[ \frac{L_2(W)}{L_0(W)} 1_{\mathcal{S}_2} \right] \geq 4\delta \mathbb{E}_0[1_{\mathcal{S}_2}] > \delta,$$

where  $1_{\mathcal{S}_2}$  is the indicator variable of event  $\mathcal{S}_2$ . Therefore, we see that if  $\mathbb{E}_0[T_{b_+}] \leq t_{b_+}^*$ , then, under  $H_2$ , the probability of algorithm  $\mathbb{A}$  returning  $M_*$  as output is at least  $\delta$ , which is not the optimal set under  $H_2$ . This

contradicts to the assumption that algorithm  $\mathbb{A}$  is a  $\delta$ -correct algorithm. In sum, we have proved that

$$\mathbb{E}_0[T_{b_+}] \geq t_{b_+}^* = \frac{|b_+|^2}{16w(b)^2} \log(1/4\delta). \quad (57)$$

(End of Part (2))

Finally, we combine the results from both parts, i.e. Eq. (55) and Eq. (57). We obtain

$$\begin{aligned} \mathbb{E}_0[T_b] &= \mathbb{E}_0[T_{b_-}] + \mathbb{E}_0[T_{b_+}] \\ &\geq \frac{|b_+|^2 + |b_-|^2}{16w(b)^2} \log(1/4\delta) \\ &\geq \frac{|b|^2}{32w(b)^2} \log(1/4\delta). \end{aligned}$$

□

Now we prove a lower bound on the probability of error in the fixed budget setting. In particular, we show that for any expected rewards  $\{w(1), \dots, w(n)\}$  and any fixed budget algorithm  $\mathbb{A}$  for pure exploration combinatorial bandit problem with feasible sets  $\mathcal{M}$ . We show that one can slightly modify the vector  $\{w(e)\}_{e \in [n]}$  to construct another vector  $\{\tilde{w}(1), \dots, \tilde{w}(n)\}$  such that the probability of error of the fixed algorithm on a **ExpCMAB** problem with expected rewards given by  $\tilde{w}$  is at least  $\Omega(\exp(n/\mathbf{H}(w)))$ .

## 5 Technical Lemmas

**Lemma 12** (Basis exchange property). *AA*

## 6 Trash

**Definition 1** (Optimal diff-sets). *Given a diff-set class  $\mathcal{B}$  and the optimal set  $M_*$ . We define  $\mathcal{B}_{\text{opt}}$  as a subset of  $\mathcal{B}$ , and for all  $b \in \mathcal{B}$ ,  $b \in \mathcal{B}_{\text{opt}}$  if and only if, there exists  $M \neq M_*$  and  $M_* \ominus M$  can be decomposed as  $b, b_1, \dots, b_k$  on  $\mathcal{B}$ .*

**Definition 2** (Hardness  $\Delta_e$  of base arm  $e$ ). *For each  $e \in [n]$ , we define its hardness  $\Delta_e$  as follows*

$$\Delta_e = \min_{b \in \mathcal{B}_{\text{opt}}, e \in b} \frac{1}{\text{width}(\mathcal{B})} w(b).$$

**Definition 3** (Sufficient exploration). *For all  $t > 0$ , we define  $E_t^3 \subseteq [n]$ , such that, for all  $e \in [n]$   $e \in E_t^3$  if and only if  $\text{rad}_t(e) < \frac{1}{3}\Delta_e$ .*

**Corollary 1.** *For all  $t > 0$  and  $e \in [n]$*

$$n_t(e) \geq O\left(\frac{1}{\Delta_e^2} \log(\Delta_e n / \delta)\right) \implies e \in E_t^3.$$

**Theorem 4.** *With probability at least  $1 - \delta$ , the algorithm returns  $M_*$ , and the number of samples used by the algorithm are at most*

$$\sum_{e \in [n]} \Delta_e^{-2} \log(\Delta_e n / \delta).$$

**Theorem 5.** *Given confidence parameter  $\delta \in (0, 1)$ , tolerance parameter  $\epsilon \geq 0$ , number of arms  $n$  and a combinatorial problem instance  $\mathcal{M} \subseteq 2^{[n]}$ . Let oracle  $\text{Oracle}(w)$  be a maximization oracle associated with  $\mathcal{M}$  such that  $\text{Oracle}(w) = \arg \max_{M \in \mathcal{M}} w(M)$ , where  $w : 2^{[n]} \rightarrow R$  is a weight function.*

Then, with probability at least  $1 - \delta$ , the output  $\text{Out}$  of Algorithm 1 satisfies  $w(M_*) - w(\text{Out}) \leq \epsilon$ , where  $M_* = \arg \max_{M \in \mathcal{M}} w(M)$  is the optimal set. In addition, the number of samples  $T$  used by the algorithm satisfies

$$T \leq \mathbf{H}_\epsilon \log \left( \frac{n}{\delta} \mathbf{H}_\epsilon \right),$$

where

$$\mathbf{H}_\epsilon = \sum_{e \in [n]} \min \left\{ \frac{\text{width}(\mathcal{B})^2}{\Delta_e^2}, \frac{n^2}{\epsilon^2} \right\}.$$

**Lemma 13.** For any arm  $e \in [n]$  and any round  $t > n$  after initialization, if  $\text{rad}_t(e) \leq \max \left\{ \frac{\Delta_\epsilon}{3 \text{width}(\mathcal{B})}, \frac{\epsilon}{n} \right\}$ , then arm  $e$  will not be played on round  $t$ , i.e.  $p_t \neq e$ .

*Proof.* If  $\text{rad}_t(e) \leq \frac{\Delta_\epsilon}{3 \text{width}(\mathcal{B})}$ , then we can apply Lemma 10 which immediately gives that  $p_t \neq e$ . Hence, we only need to prove the case that  $\text{rad}_t(e) \leq \frac{\epsilon}{n}$ . By the definition of  $p_t$ , we know that for each  $i \in D_t$ , we have  $\text{rad}_t(i) \leq \text{rad}_t(e) \leq \frac{\epsilon}{n}$ . Summing up all  $i \in D_t$ , we obtain

$$\text{rad}_t(D_t) \leq \epsilon. \quad (58)$$

Next, we notice that the definition of  $M_t$  gives that  $\bar{w}_t(M_t) = \max_{M \in \mathcal{M}} \bar{w}_t(M) \geq \bar{w}_t(M_t \oplus D_t)$ . This means that

$$\bar{w}_t(D_t) = \bar{w}_t(M_t \oplus D_t) - \bar{w}_t(M_t) \leq 0. \quad (59)$$

Using the above inequalities, we have.

$$w_t^+(D_t) = \bar{w}_t(D_t) + \text{rad}_t(D_t) \quad (60)$$

$$\leq \bar{w}_t(D_t) + \epsilon \quad (61)$$

$$\leq \epsilon, \quad (62)$$

where Eq. (60) follows from the definition of  $w_t^+(\cdot)$ ; Eq. (61) follows from Eq. (58); Eq. (62) holds since Eq. (59).  $\square$

## 7 Preliminaries

### 7.1 Problems

Let  $n$  be the number of base arms. Let  $\mathcal{M} \subseteq 2^{[n]}$  be the set of super arms.

In this note, we consider the following cases of  $\mathcal{M}$ .

**Example 1** (Explore- $m$ ).  $\mathcal{M}_{\text{TOP}_m}(n) = \{M \subseteq [n] \mid |M| = m\}$ . This corresponds to finding the top  $m$  arms from  $[n]$ .

**Example 2** (Explore- $m$ -bandits). Suppose  $n = mk$ . Then  $\mathcal{M}_{\text{BANDIT}_m}(n)$  contains all subsets  $M \subseteq [n]$  with size  $m$ , such that

$$M \cap \{ik + 1, \dots, (i+1)k\} = 1, \quad \text{for all } i \in \{0, \dots, m-1\}.$$

This corresponds to finding the top arms from  $m$  bandits, where each bandit has  $k$  arms.

**Example 3** (Perfect Matching). Let  $G = (V, E)$  be a bipartite graph and  $|E| = n$ . For simplicity, let each edge  $e \in E$  corresponds to a unique integer  $i \in [n]$ , and vice versa. Then  $\mathcal{M}_{\text{MATCH}}(n, G)$  contains all subsets  $M \subseteq [n]$  such that  $M$  corresponds to a perfect matching in  $G$ .

## 7.2 Diff-Sets

**Definition 4** (Diff-set). An  $n$ -diff-set (or diff-set in short) is a pair of sets  $c = (c_+, c_-)$ , where  $c_+ \subseteq [n]$ ,  $c_- \subseteq [n]$  and  $c_+ \cap c_- = \emptyset$ .

**Definition 5** (Difference of sets). Given any  $M_1 \subseteq [n], M_2 \subseteq [n]$ . We define  $M_1 \ominus M_2 \triangleq C$ , where  $C = (C_+, C_-)$  is a diff-set and  $C_+ = M_1 \setminus M_2$  and  $C_- = M_2 \setminus M_1$ .

**Definition 6.** Denote  $\text{diff}[n]$  be the set of all possible  $n$ -diff-sets.

**Definition 7** (Set operations of diff-sets). Let  $C = (C_+, C_-), D = (D_+, D_-)$  be two diff-sets. We define  $C \cap D \triangleq (C_+ \cap D_+, C_- \cap D_-)$  and  $C \setminus D \triangleq (C_+ \setminus D_+, C_- \setminus D_-)$ .

Further, for all  $e \in [n]$ ,  $e \in C \Leftrightarrow (e \in C_+) \vee (e \in C_-)$ . And  $|C| \triangleq |C_+| + |C_-|$ .

**Definition 8** (Valid diff-set). Given a set  $M \subseteq [n]$  and a diff-set  $C = (C_+, C_-)$ , we call  $C$  a valid diff-set for  $M$ , iff  $C_+ \cap M = \emptyset$  and  $C_- \subseteq M$ . In this case, we denote  $C \prec M$ .

**Definition 9** (Negative diff-set). Given a diff-set  $A = (A_+, A_-)$ , we define  $\neg A = (A_-, A_+)$ .

### 7.2.1 diff-set operations

**Definition 10** (Operators  $\oplus$  and  $\ominus$ ). Given any  $M \subseteq [n]$  and  $C \in \text{diff}[n]$ . If  $C \prec M$ , we define operator  $\oplus$  such that  $M \oplus C \triangleq M \setminus C_- \cup C_+$ . On the other hand if  $\neg C \prec M$ , we define operator  $\ominus$  such that  $M \ominus C \triangleq M \oplus (\neg C) = M \setminus C_+ \cup C_-$ .

**Definition 11.** Given two diff-sets  $A = (A_+, A_-)$  and  $B = (B_+, B_-)$ . We denote  $B \prec A$ , if and only if  $B_+ \cap A_+ = \emptyset$  and  $A_+ \cap A_- = \emptyset$ .

**Definition 12.** Given two diff-sets  $A = (A_+, A_-)$  and  $B = (B_+, B_-)$ . If  $B \prec A$ , we define  $A \oplus B = ((A_+ \cup B_+) \setminus (A_- \cup B_-), (A_- \cup B_-) \setminus (A_+ \cup B_+))$ .

**Lemma 14.** Given two diff-sets  $A = (A_+, A_-)$  and  $B = (B_+, B_-)$ . If  $B \prec A$ , then  $A \oplus B$  is a diff-set.

*Proof.* Let  $C = A \oplus B$ . By definition, we have  $C_+ = (A_+ \cup B_+) \setminus (A_- \cup B_-)$  and  $C_- = (A_- \cup B_-) \setminus (A_+ \cup B_+)$ . We only need to show that  $C_+ \cap C_- = \emptyset$ .

$$\begin{aligned} C_+ \cap C_- &= ((A_+ \cup B_+) \setminus (A_- \cup B_-)) \cap ((A_- \cup B_-) \setminus (A_+ \cup B_+)) \\ &= (A_+ \cup B_+) \cap ((A_- \cup B_-) \setminus (A_+ \cup B_+)) \setminus (A_- \cup B_-) \\ &= \emptyset. \end{aligned}$$

□

**Lemma 15.** Given two diff-sets  $A = (A_+, A_-)$  and  $B = (B_+, B_-)$ . If there exists  $M \subseteq [n]$  such that  $A \prec M$ , and  $B \prec (M \oplus A)$ , then  $B \prec A$  and  $(M \oplus A \oplus B) \ominus M = A \oplus B$ .

*Proof.* We first show that  $B \prec A$ . Since  $B \prec (M \oplus A)$ , we know that  $B_+ \cap (M \setminus A_- \cup A_+) = \emptyset$ . Therefore, we have

$$\begin{aligned} \emptyset &= B_+ \cap (M \setminus A_- \cup A_+) \\ &= (B_+ \cap (M \setminus A_-)) \cup (B_+ \cap A_+) \end{aligned}$$

We see that  $B_+ \cap A_+ = \emptyset$ .

On the other hand, we have  $B_- \subseteq (M \setminus A_- \cup A_+)$ , therefore

$$B_- \cap A_- \subseteq (M \setminus A_- \cup A_+) \cap A_-$$

$$\begin{aligned}
&= (M \setminus A_- \cap A_-) \cup (A_+ \cap A_-) \\
&= \emptyset.
\end{aligned}$$

Hence we proved that  $B \prec A$ .

Define  $D = (M \oplus A \oplus B) \ominus M$  and write  $D = (D_+, D_-)$ . Then,

$$\begin{aligned}
D_+ &= (M \oplus A \oplus B) \setminus M \\
&= (M \setminus A_- \cup A_+ \setminus B_- \cup B_+) \setminus M \\
&= (A_+ \cup B_+) \setminus (A_- \cup B_-).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
D_- &= M \setminus (M \oplus A \oplus B) \\
&= M \setminus (M \setminus A_- \cup A_+ \setminus B_- \cup B_+) \\
&= (A_- \cup B_-) \setminus (A_+ \cup B_+).
\end{aligned}$$

□

### 7.2.2 Diff-set class

**Definition 13** (Decomposition of diff-set). *Given  $\mathcal{B} \subseteq \text{diff}[n]$  and  $D \in \text{diff}[n]$ , a decomposition of  $D$  on  $\mathcal{B}$  is a set  $\{b_1, \dots, b_k\} \subseteq \mathcal{B}$  satisfying the following*

1. For all  $i \in [k]$  and  $j \in [k]$ , we write  $b_i = (b_i^+, b_i^-)$  and  $b_j = (b_j^+, b_j^-)$ . Then, the following holds  
 $b_i^+ \cap b_j^+ = \emptyset$ ,  $b_i^+ \cap b_j^- = \emptyset$ ,  $b_i^- \cap b_j^+ = \emptyset$  and  $b_i^- \cap b_j^- = \emptyset$ .
2.  $D = b_1 \oplus b_2 \oplus \dots \oplus b_k$ .

**Lemma 16.** *Given  $\mathcal{B} \subseteq \text{diff}[n]$  and  $D \in \text{diff}[n]$ . Let  $\{b_1, \dots, b_k\} \subseteq \mathcal{B}$  be a decomposition of  $D$  on  $\mathcal{B}$ . Then,*

1. Let  $D = (D_+, D_-)$  and for all  $i \in [k]$ , we write  $b_i = (b_i^+, b_i^-)$ . Then  $D_+ = b_1^+ \cup \dots \cup b_k^+$  and  $D_- = b_1^- \cup \dots \cup b_k^-$ .
2. For all  $M \subseteq [n]$ , if  $D \prec M$ , then, for all  $i \in [k]$ , we have  $b_i \prec M$ .

*Proof.* We prove (1) by induction. Let  $D_i = b_1 \oplus \dots \oplus b_i$  and write  $D_i = (D_i^+, D_i^-)$ . We show that  $D_i^+ = \bigcup_{j=1}^i b_j^+$  and  $D_i^- = \bigcup_{j=1}^i b_j^-$  for all  $i \in [k]$ . For  $i = 1$ , this is trivially true. Then, assume that this is true for some  $i > 1$ . By definition  $D_{i+1} = D_i \oplus b_{i+1}$ , hence  $D_{i+1}^+ = (D_i^+ \cup b_{i+1}^+) \setminus (D_i^- \cup b_{i+1}^-)$ . Note that

$$\begin{aligned}
(D_i^- \cup b_{i+1}^-) \cap (D_i^+ \cup b_{i+1}^+) &= (D_i^- \cap D_i^+) \cup (D_i^- \cap b_{i+1}^+) \cup (b_{i+1}^- \cap D_i^+) \cup (b_{i+1}^- \cap b_{i+1}^+) \\
&= (D_i^- \cap b_{i+1}^+) \cup (b_{i+1}^- \cap D_i^+) \\
&= \left( \left( \bigcup_{j=1}^i b_j^- \right) \cap b_{i+1}^+ \right) \cup \left( \left( \bigcup_{j=1}^i b_j^+ \right) \cap b_{i+1}^- \right) \\
&= \emptyset.
\end{aligned}$$

Hence  $D_{i+1}^+ = D_i^+ \cup b_{i+1}^+$ . We can use the same method to show that  $D_{i+1}^- = D_i^- \cup b_{i+1}^-$ .

Next, we prove (2) using (1). To show that  $b_i \prec M$ , we only need to show that  $b_i^+ \cap M = \emptyset$  and  $b_i^- \subseteq M$ . Since  $D \prec M$ , we know that  $D_+ \cap M = \emptyset$  and  $D_- \subseteq M$ . By (1), we see that  $b_i^+ \subseteq D_+$  and  $b_i^- \subseteq D_-$ . Therefore, we have  $(b_i^+ \cap M) \subseteq (D_+ \cap M) = \emptyset$  and  $b_i^- \subseteq D_- \subseteq M$ .

□

**Definition 14** (diff-set class). Given  $\mathcal{M} \subseteq 2^{[n]}$ .  $\mathcal{B} \subseteq \text{diff}[n]$  is a diff-set class for  $\mathcal{M}$ , if the following hold.

1.  $(\emptyset, \emptyset) \notin \mathcal{B}$ .
2. For all  $M \in \mathcal{M}$  and for all  $b \in \mathcal{B}$ , if  $b \prec M$ , then  $M \oplus b \in \mathcal{M}$ .
3. For all  $M_1 \in \mathcal{M}$  and  $M_2 \in \mathcal{M}$ , where  $M_1 \neq M_2$ . Let  $D = M_1 \ominus M_2$ . Then, there exists a decomposition of  $D$  on  $\mathcal{B}$ .

**Definition 15** (Rank of diff-set class). Let  $\mathcal{B} \subseteq [n]$  be a diff-set class for some  $\mathcal{M}$ . We define

$$\text{width}(\mathcal{B}) \triangleq \max_{b \in \mathcal{B}} |b|.$$

**Example 4** (diff-set class for Explore- $m$ ). One diff-set class  $\mathcal{B}$  for  $\mathcal{M}_{\text{TOP}_m(n)}$  is given by

$$\mathcal{B} = \{(\{b_1\}, \{b_2\}) \mid b_1 \neq b_2, b_1 \in [n], b_2 \in [n]\}.$$

*Proof omitted. Further, we see that  $\text{width}(\mathcal{B}) = 2$ .*

**Example 5** (diff-set class for Explore- $m$ -badit). Let  $n = mk$ . One diff-set class  $\mathcal{B}$  for  $\mathcal{M}_{\text{BANDIT}_m(n)}$  is given by

$$\mathcal{B} = \{(\{b_1\}, \{b_2\}) \mid b_1 \neq b_2, \exists i \in \{0, \dots, k-1\}, b_1 \in \{ik+1, \dots, (i+1)k\}, b_2 \in \{ik+1, \dots, (i+1)k\}\}.$$

*Proof omitted. Further, we see that  $\text{width}(\mathcal{B}) = 2$ .*

**Example 6** (diff-set class for Perfect Matching). One diff-set class  $\mathcal{B}$  for  $\mathcal{M}_{\text{MATCH}}(n, G)$  is the set of all augmenting cycles of  $G$ . More specifically,

$$\mathcal{B} = \{(b_+, b_-) \mid b_+ \cup b_- \text{ is a cycle of } G\}.$$

Note  $\text{width}(\mathcal{B}) \leq n$ .

### 7.3 Weights and confidence bounds

**Definition 16** (Weight functions). Define function  $w : [n] \rightarrow \mathbb{R}^+$  which represents the weight of each base arm. Further, we slight abuse the notations, and extend the definition of  $w$  to diff-sets and sets as follows.

1. For all  $M \subseteq [n]$ , we denote  $w(M) = \sum_{e \in M} w(e)$ .
2. For all  $b = (b_+, b_-) \in \text{diff}[n]$ , we denote  $w(b) = \sum_{e \in b_+} w(e) - \sum_{e \in b_-} w(e)$ .

**Lemma 17.** Let  $c \in \text{diff}[n], d \in \text{diff}[n]$ . Let  $w$  be a weight function. Then,

$$w(c \cup d) = w(c) + w(d) - w(c \cap d). \quad (63)$$

*Proof.* Let  $c = (c_+, c_-)$  and  $d = (d_+, d_-)$ . We have

$$w(c \cup d) = w(c_+ \cup d_+) - w(c_- \cup d_-) \quad (64)$$

$$= w(c_+) + w(d_+) - w(c_+ \cap d_+) - w(c_-) - w(d_-) + w(c_- \cap d_-) \quad (65)$$

$$= w(c) + w(d) - (w(c_+ \cap d_+) - w(c_- \cap d_-)) \quad (66)$$

$$= w(c) + w(d) - w(c \cap d). \quad (67)$$

□

**Definition 17** (Mean weight  $\bar{w}_t$ , sample size  $n_t$ ). Given  $t > 0$ . Define  $\bar{w}_t$  be a weight function such that, for all  $e \in [n]$ ,  $\bar{w}_t(e)$  equals to the empirical mean of  $e$  up to round  $t$ . Let  $n_t : [n] \rightarrow \mathbb{N}$ , such that  $n_t(e)$  equals to number of plays of base arm  $e$  up to round  $t$ .

**Definition 18** (Confidence radius  $\text{rad}_t$ ). Given  $n$  and  $t > 0$ . Define  $\text{rad}_t : [n] \rightarrow \mathbb{R}^+$  satisfying, for all  $e \in [n]$ ,

$$\text{rad}_t(e) = c_{\text{rad}} \log \left( \frac{c_\delta n t^2}{\delta} \right) \frac{1}{\sqrt{n_t(e)}}, \quad (68)$$

where  $c_{\text{rad}} > 0$  and  $c_\delta > 0$  are some universal constants (specify later) and  $\delta > 0$  is a parameter. We extend the notation of  $\text{rad}_t$  to diff-sets and sets as follows.

1. For all  $M \subseteq [n]$ ,  $\text{rad}_t(M) \triangleq \sum_{e \in M} \text{rad}_t(e)$ .
2. For all  $b = (b_+, b_-) \in \text{diff}[n]$ ,  $\text{rad}_t(b) \triangleq \text{rad}_t(b_+) + \text{rad}_t(b_-)$ .

**Definition 19** (UCB  $w_t^+$ ). Define  $w_t^+ : [n] \rightarrow \mathbb{R}^+$ , s.t., for all  $e \in [n]$ ,

$$w_t^+(e) = \bar{w}_t(e) + \text{rad}_t(e).$$

We extend the notation of  $w_t^+$  to diff-sets and sets as follows.

1. For all  $M \subseteq [n]$ ,  $w_t^+(M) \triangleq \bar{w}_t(M) + \text{rad}_t(M)$ .
2. For all  $b = (b_+, b_-) \in \text{diff}[n]$ ,  $w_t^+(b) \triangleq \bar{w}_t(b) + \text{rad}_t(b)$ .

**Lemma 18.** Define random event

$$\xi = \{ \forall e \in [n] \forall t > 0, |\bar{w}_t(e) - w(e)| \leq \text{rad}_t(e) \}.$$

Then, there exist constants  $c_{\text{rad}}$  and  $c_\delta$ ,

$$\Pr[\xi] \geq 1 - \delta.$$

*Proof.* Hoeffding inequality and union bound. □

**Corollary 2.**

$$\begin{aligned} \xi &\implies \forall t, \forall e \in [n] \ w_t^+(e) \geq w(e). \\ \xi &\implies \forall t, \forall M \subseteq [n], \ w_t^+(M) \geq w(M). \\ \xi &\implies \forall t, \forall b \in \text{diff}[n] \ w_t^+(b) \geq w(b). \end{aligned}$$

## 7.4 Properties of $\text{rad}_t$

**Lemma 19.** Let  $c \in \text{diff}[n]$ ,  $d \in \text{diff}[n]$ . Then

$$\text{rad}_t(c \setminus d) = \text{rad}_t(c) - \text{rad}_t(c \cap d). \quad (69)$$

*Proof.* Let  $c = (c_+, c_-)$  and  $d = (d_+, d_-)$ . We have

$$\begin{aligned} \text{rad}_t(c \setminus d) &= \text{rad}_t(c_+ \setminus d_+) + \text{rad}_t(c_- \setminus d_-) \\ &= \text{rad}_t(c_+) - \text{rad}_t(c_+ \cap d_+) + \text{rad}_t(c_-) - \text{rad}_t(c_- \cap d_-) \\ &= \text{rad}_t(c) - \text{rad}_t(c \cap d). \end{aligned}$$

□



**Lemma 20.** Let  $C = (C_+, C_-)$  and  $D = (D_+, D_-)$  be two diff-sets. If  $D \prec C$ , then

$$\text{rad}_t(C \oplus D) = \text{rad}_t(C) + \text{rad}_t(D) - 2\text{rad}_t(C_+ \cap D_-) - 2\text{rad}_t(C_- \cap D_+).$$

In addition, if  $\neg D \prec C$ , then

$$\text{rad}_t(C \ominus D) = \text{rad}_t(C) + \text{rad}_t(D) - 2\text{rad}_t(C_+ \cap D_+) - 2\text{rad}_t(C_- \cap D_-).$$

*Proof.* We prove the first part of the lemma. The second part follows from the first part and the definition of  $\neg D$ .

By definition, we have  $C \oplus D = ((C_+ \cup D_+) \setminus (C_- \cup D_-), (C_- \cup D_-) \setminus (C_+ \cup D_+))$ . Hence, we have

$$\text{rad}_t((C_+ \cup D_+) \setminus (C_- \cup D_-)) = \text{rad}_t(C_+ \cup D_+) - \text{rad}_t((C_+ \cup D_+) \cap (C_- \cup D_-)) \quad (70)$$

$$= \text{rad}_t(C_+) + \text{rad}_t(D_+) - \text{rad}_t((C_+ \cup D_+) \cap (C_- \cup D_-)), \quad (71)$$

where the second equality holds due to  $C_+ \cap D_+ = \emptyset$  by the definition of  $D \prec C$ .

Similarly, we have

$$\text{rad}_t((C_- \cup D_-) \setminus (C_+ \cup D_+)) = \text{rad}_t(C_-) + \text{rad}_t(D_-) - \text{rad}_t((C_+ \cup D_+) \cap (C_- \cup D_-)).$$

Combine both equalities, we have

$$\text{rad}_t(C \oplus D) = \text{rad}_t((C_+ \cup D_+) \setminus (C_- \cup D_-)) + \text{rad}_t((C_- \cup D_-) \setminus (C_+ \cup D_+)) \quad (72)$$

$$= \text{rad}_t(C_+) + \text{rad}_t(D_+) + \text{rad}_t(C_-) + \text{rad}_t(D_-) - 2\text{rad}_t((C_+ \cup D_+) \cap (C_- \cup D_-)) \quad (73)$$

$$= \text{rad}_t(C) + \text{rad}_t(D) - 2\text{rad}_t((C_+ \cup D_+) \cap (C_- \cup D_-)). \quad (74)$$

□

*Proof.* Suppose that  $p_t = e$ . Since  $\mathcal{B}$  is an exchange class for  $\mathcal{M}$ , there exists an exchange set  $c = (c_+, c_-)$  of  $\mathcal{B}$  such that  $e \in c_-$ ,  $c_- \subseteq M_t$  and  $c_+ \subseteq \tilde{M}_t$ .

We decompose the proof into three cases.

**Case (1).** Suppose that  $c \in \mathcal{B}_{\text{opt}}$ . Then  $w(c) > 0$ . We have  $\text{rad}_t(e) \leq \frac{1}{3}\Delta_e \leq \frac{1}{3K}w(c)$ .

In addition, by the choice of  $p_t$ , we have  $\forall g \in c_t, g \neq e$ ,  $\text{rad}_t(g) \leq \text{rad}_t(e) \leq \frac{1}{3K}w(c)$ . Hence,  $\text{rad}_t(c) = \sum_{g \in c_t} \text{rad}_t(g) \leq \frac{|c_t|}{3K}w(c) \leq \frac{1}{3}w(c)$ .

Hence,  $w_t(c) \geq w(c) - \text{rad}_t(c) \geq \frac{2}{3}w(c) > 0$ . This means that  $w_t(M_t \oplus c) = w_t(M_t) + w_t(c) > w_t(M_t)$ . Therefore,  $M_t \neq \max_{M \in \mathcal{M}} \tilde{w}_t(M)$ . This contradicts to the definition of  $M_t$ .

**Case (2).** Suppose that  $c_t \notin \mathcal{B}_{\text{opt}}$ . Then, one of the following mutually exclusive cases must hold.

**Case (2.1).**  $(e \in M_* \wedge e \in c_+)$  or  $(e \notin M_* \wedge e \in c_-)$ .

Let the decomposition of  $M_* \ominus (M_t \oplus D \ominus c)$  on  $\mathcal{B}$  be  $b, b_1, \dots, b_l$ , which exists due to  $\mathcal{B}$  is a diff-set class.

Assume wlog that  $e \in b$ . We write  $b = (b_+, b_-)$ . It is easy to see that  $b \in \mathcal{B}_{\text{opt}}$ .

Define  $\tilde{D} = (M_t \oplus D \ominus c) \ominus M_t$  and  $D' = (M_t \oplus \tilde{D} \oplus b) \ominus M_t$ . By Lemma 15, we know that  $\tilde{D} = D \ominus c$  and  $D' = \tilde{D} \oplus b$ . We also write  $\tilde{D} = (\tilde{D}_+, \tilde{D}_-)$  and  $D' = (D'_+, D'_-)$ . By definition, we have

$$\begin{aligned} \tilde{D}_+ &= (D_+ \cup c_-) \setminus (D_- \cup c_+) \\ &= (D_+ \cup c_- \setminus D_-) \cap (D_+ \cup c_- \setminus c_+) \\ &= D_+ \cap (D_+ \setminus c_-) \\ &= D_+ \setminus c_+. \end{aligned}$$

By the same method, we are able to show that  $\tilde{D}_- = D_- \setminus c_-$ . Therefore we have

$$\tilde{D}_+ \subseteq D_+ \quad \text{and} \quad \tilde{D}_- \subseteq D_- . \quad (75)$$

First, we show that  $\text{rad}_t(c) \leq \frac{1}{3}w(b)$ . Since  $e \in E_t^3$ ,  $e \in b$  and  $b \in \mathcal{B}_{\text{opt}}$ , we have  $\text{rad}_t(e) \leq \frac{1}{3}\Delta_e \leq \frac{1}{3K}w(b)$ . In addition,  $\forall g \in c, g \neq e$ ,  $\text{rad}_t(g) \leq \text{rad}_t(e) \leq \frac{1}{3K}w(b)$ . Hence,

$$\begin{aligned} \text{rad}_t(c) &= \sum_{g \in c} \text{rad}_t(g) \\ &\leq \frac{|c|}{3K}w(b) \\ &\leq \frac{1}{3}w(b). \end{aligned} \quad (76)$$

Now, we show that  $\text{rad}_t(\tilde{D}_+ \cap b_-) + \text{rad}_t(\tilde{D}_- \cap b_+) \leq \frac{1}{3}w(b)$ . Since Eq. (75), we have  $\forall g \in (\tilde{D}_+ \cap b_-) \cup (\tilde{D}_- \cap b_+), g \neq e$ ,  $\text{rad}_t(g) \leq \text{rad}_t(e) \leq \frac{1}{3K}w(b)$ . Note that  $|\tilde{D}_+ \cap b_-| + |\tilde{D}_- \cap b_+| \leq |b_+| + |b_-| \leq K$ . Hence,

$$\begin{aligned} \text{rad}_t(\tilde{D}_+ \cap b_-) + \text{rad}_t(\tilde{D}_- \cap b_+) &= \sum_{g \in (\tilde{D}_+ \cap b_-) \cup (\tilde{D}_- \cap b_+)} \text{rad}_t(g) \\ &\leq \frac{K}{3K}w(b) \\ &\leq \frac{1}{3}w(b). \end{aligned} \quad (77)$$

Then, we have

$$\text{rad}_t(D') - \text{rad}_t(D) = \text{rad}_t(\tilde{D} \oplus b) - \text{rad}_t(D) \quad (78)$$

$$= \text{rad}_t(\tilde{D}) + \text{rad}_t(b) - 2\text{rad}_t(\tilde{D}_+ \cap b_-) - 2\text{rad}_t(\tilde{D}_- \cap b_+) - \text{rad}_t(D) \quad (79)$$

$$= \text{rad}_t(D \ominus c) + \text{rad}_t(b) - 2\text{rad}_t(\tilde{D}_+ \cap b_-) - 2\text{rad}_t(\tilde{D}_- \cap b_+) - \text{rad}_t(D) \quad (80)$$

$$\begin{aligned} &= \text{rad}_t(D) + \text{rad}_t(c) + \text{rad}_t(b) - 2\text{rad}_t(D_+ \cap c_+) - 2\text{rad}_t(D_- \cap c_-) \\ &\quad - 2\text{rad}_t(\tilde{D}_+ \cap b_-) - 2\text{rad}_t(\tilde{D}_- \cap b_+) - \text{rad}_t(D) \end{aligned} \quad (81)$$

$$\begin{aligned} &= \text{rad}_t(D) + \text{rad}_t(c) + \text{rad}_t(b) - 2\text{rad}_t(c_+) - 2\text{rad}_t(c_-) \\ &\quad - 2\text{rad}_t(\tilde{D}_+ \cap b_-) - 2\text{rad}_t(\tilde{D}_- \cap b_+) - \text{rad}_t(D) \end{aligned} \quad (82)$$

$$= \text{rad}_t(b) - \text{rad}_t(c) - 2\text{rad}_t(\tilde{D}_+ \cap b_-) - 2\text{rad}_t(\tilde{D}_- \cap b_+), \quad (83)$$

where Eq. (79) and Eq. (81) follow from Lemma 20, and Eq. (82) follows from Eq. (??).

By the definition of  $D$ , we have that  $w_t^+(D) \geq w_t^+(D')$ . This means that

$$\bar{w}_t(D) + \text{rad}_t(D) \geq \bar{w}_t(D') + \text{rad}_t(D') \quad (84)$$

$$= \bar{w}_t(D) - \bar{w}_t(c) + \bar{w}_t(b) + \text{rad}_t(D'). \quad (85)$$

By regrouping the above inequality, we have

$$\bar{w}_t(c) \geq \bar{w}_t(b) + \text{rad}_t(D') - \text{rad}_t(D) \quad (86)$$

$$= \bar{w}_t(b) + \text{rad}_t(b) - \text{rad}_t(c) - 2\text{rad}_t(\tilde{D}_+ \cap b_-) - 2\text{rad}_t(\tilde{D}_- \cap b_+) \quad (87)$$

$$\geq w(b) - \text{rad}_t(c) - 2\text{rad}_t(\tilde{D}_+ \cap b_-) - 2\text{rad}_t(\tilde{D}_- \cap b_+) \quad (88)$$

$$> w(b) - \frac{1}{3}w(b) - \frac{2}{3}w(b) \quad (89)$$

$$= 0, \quad (90)$$

where Eq. (89) follows from Eq. (76) and Eq. (77).

This contradicts to the definition of  $M_t$ .

**Case (2.2).**  $(e \in M_* \wedge e \in c_-)$  or  $(e \notin M_* \wedge e \in c_+)$ .

Let the decomposition of  $M_* \ominus (M_t \oplus D)$  on  $\mathcal{B}$  be  $b, b_1, \dots, b_l$ . Assume wlog that  $e \in b$ . We write that  $b = (b_+, b_-)$ . Note that  $b \in \mathcal{B}_{\text{opt}}$  and hence  $w(b) > 0$ .

Define  $D' = (M_t \oplus D \oplus b) \ominus M_t$ . By Lemma 15, we know that  $D' = D \oplus b$ .

First, we show that  $|D \setminus D'| \leq |b|$ . Let  $C = D \setminus D'$  and write  $C = (C_+, C_-)$ . We can bound  $|C_+|$  as follows.

$$\begin{aligned} C_+ &= D_+ \setminus D'_+ \\ &= D_+ \setminus ((D_+ \cup b_+) \setminus (D_- \cup b_-)) \\ &= (D_+ \cap (D_- \cup b_-)) \cup (D_+ \setminus (D_+ \cup b_+)) \\ &= D_+ \cap b_-. \end{aligned}$$

Hence, we have  $|C_+| \leq |b_-|$ . Then, we move to bounding  $|C_-|$

$$\begin{aligned} C_- &= D_- \setminus D'_- \\ &= D_- \setminus ((D_- \cup b_-) \setminus (D_+ \cup b_+)) \\ &= (D_- \cap (D_+ \cup b_+)) \cup (D_- \setminus (D_- \cup b_-)) \\ &= D_- \cap b_+. \end{aligned}$$

Thus  $|C_-| \leq |b_+|$  and we proved that  $|D \setminus D'| \leq |b|$ .

Next, we show that  $\text{rad}_t(D \setminus D') \leq \frac{1}{3}w(b)$ . Since  $e \in E_t^3$ ,  $e \in b$  and  $b \in \mathcal{B}_{\text{opt}}$ , we have  $\text{rad}_t(e) \leq \frac{1}{3}\Delta_e \leq \frac{1}{3K}w(b)$ . In addition,  $\forall g \in (D \setminus D'), g \neq e$ ,  $\text{rad}_t(g) \leq \text{rad}_t(e) \leq \frac{1}{3K}w(b)$ . Note that  $|D \setminus D'| \leq |b| \leq K$ . Hence,  $\text{rad}_t(D \setminus D') = \sum_{g \in (D \setminus D')} \text{rad}_t(g) \leq \frac{K}{3K}w(b) \leq \frac{1}{3}w(b)$ .

We also note that

$$w(D' \setminus D) - w(D \setminus D') = w(D' \setminus D) + w(D' \cap D) - w(D \cap D') - w(D \setminus D') \quad (91)$$

$$= w(D') - w(D) \quad (92)$$

$$= w(b), \quad (93)$$

where we have repeatedly applied Lemma 17.

Then, we show that  $w_t^+(D') > w_t^+(D)$ .

$$w_t^+(D') - w_t^+(D) = \bar{w}_t(D') - \bar{w}_t(D) + \text{rad}_t(D') - \text{rad}_t(D) \quad (94)$$

$$= \bar{w}_t(D' \setminus D) - \bar{w}_t(D \setminus D') + \text{rad}_t(D' \setminus D) - \text{rad}_t(D \setminus D') \quad (95)$$

$$\geq w(D' \setminus D) - w(D \setminus D') - 2\text{rad}_t(D \setminus D') \quad (96)$$

$$= w(b) - 2\text{rad}_t(D \setminus D') \quad (97)$$

$$> w(b) - \frac{2}{3}w(b) \quad (98)$$

$$= \frac{1}{3}w(b) > 0, \quad (99)$$

where Eq. (95) follows from Lemma 19 and Eq. (96) follows from the fact that  $\bar{w}_t(D' \setminus D) + \text{rad}_t(D' \setminus D) \geq w(D' \setminus D)$  and that  $\bar{w}_t(D \setminus D') + \text{rad}_t(D \setminus D') \geq w(D \setminus D')$ , under the random event  $\xi$ .

This contradicts to the fact that  $D$  is chosen on round  $t$ .  $\square$

**Theorem 6.** Given a vector  $\{w(1), \dots, w(n)\}$ , a budget  $T > 0$  and a collection of feasible sets  $\mathcal{M} \subseteq 2^{[n]}$ . Let  $\mathbb{A}$  be an arbitrary algorithm for  $\mathcal{M}$ -**ExpCMB** problem which uses at most  $T$  samples. There exists a vector  $\{\tilde{w}(1), \dots, \tilde{w}(n)\}$  such that  $\mathbf{H}(\tilde{w}) \leq 2\mathbf{H}(w)$  and satisfies the following property. Consider the bandit problem

with reward distributions defined by  $\varphi_e = \mathcal{N}(\tilde{w}(e), 1)$  for all  $e \in [n]$ , where  $\mathcal{N}(\mu, \sigma^2)$  denotes Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ . The probability of error of  $\mathbb{A}$  on this bandit problem satisfies

$$\Pr[\text{Out} \neq M_*] \geq \exp\left(-\frac{T}{\mathbf{H}(w)}\right),$$

where  $\text{Out}$  is the output of  $\mathbb{A}$  and  $M_* = \arg \max_{M \in \mathcal{M}} w(M)$  is the optimal set. In addition, vector  $\{\tilde{w}(1), \dots, \tilde{w}(n)\}$  differs from vector  $\{w(1), \dots, w(n)\}$  on exactly one index.

*Proof.* Fix  $\mathcal{M} \subseteq 2^{[n]}$ ,  $w(e)$  for all  $e \in [n]$  and a fixed budget algorithm  $\mathbb{A}$  for  $\mathcal{M}$ -ExpCMAB problem. Let  $\sigma(1), \dots, \sigma(n)$  be a permutation of  $1, 2, \dots, n$  such that  $\Delta_{\sigma(1)} \leq \Delta_{\sigma(2)} \leq \dots \leq \Delta_{\sigma(n)}$ . Define  $L' = \arg \max_{i \in [n]} i / \Delta_{\sigma(i)}^2$  and  $L = \sigma(L')$ .

Then, we construct hypothesis  $H_0$  as follows

$$H_0 : \varphi_e = \mathcal{N}(w(e), 1) \quad \text{for all } e \in [n].$$

We define random event  $\mathcal{C}$  as follows.

We show that  $\Pr_0[\mathcal{C}] \geq 1/2$ .

We define random variables  $X, Y, Z$  as follows

$$X = \arg \min_{i \in [L] \setminus \text{Out}} T_i, \quad Y = \arg \min_{i \in [L] \cap \text{Out}} T_i \quad \text{and} \quad Z = \arg \min_{i \in [L]} T_i,$$

where, for convenience, if  $[L] \setminus \text{Out} = \emptyset$ , we set  $X = 0$ ; and if  $[L] \cap \text{Out} = \emptyset$ , we set  $Y = 0$ . By definition, we see that  $X \neq Y$ ,  $Z \in \{X, Y\}$  and  $Z \neq 0$ . Now, by summing up all possible values of  $X, Y$  and  $Z$ , we have

$$1/2 < \Pr_0[\mathcal{C}] = \sum_{\substack{x \in \{0, \dots, L\} \\ y \in \{0, \dots, L\} \\ x \neq y, z \in \{x, y\}}} \Pr_0[\mathcal{C} \cap \{X = x, Y = y, Z = z\}].$$

Since a maximal is larger than an average, we see that there exists  $x, y, z$  such that  $x \neq y, z \in \{x, y\}$  and

$$\Pr_0[\mathcal{C} \cap \{X = x, Y = y, Z = z\}] \geq \frac{1}{4L(L+1)}. \quad (100)$$

We point out that  $x, y$  and  $z$  are deterministic and only depends  $\mathbb{A}$ ,  $w$  and  $\mathcal{M}$ . Now, depending on the value of  $x, y$  and  $z$ , we divide our analysis into two cases.

**Case (1):** ( $z = x \wedge x \in M_*$ ) **or** ( $z = y \wedge y \notin M_*$ ). Eq. (100) implies that

$$\begin{aligned} \Pr_0[\{X = x, Y = y, Z = z\}] &\geq \Pr_0[\mathcal{C} \cap \{X = x, Y = y, Z = z\}] \\ &\geq \frac{1}{4L(L+1)} \geq G. \end{aligned}$$

First, let us assume that  $z = x$  and  $x \in M_*$ . By definition, we have  $X \notin \text{Out}$ . Notice that  $x$  belong  $M_*$ . Therefore the event that  $X = x$  and the assumption that  $x \in M_*$  imply that  $\text{Out} \neq M_*$ . This means that, if  $z = x$  and  $x \in M_*$ , then  $\Pr_0[\text{Out} \neq M_*] \geq \Pr_0[X = x] \geq G$ .

Next, we assume that  $z = y$  and  $y \notin M_*$ . Notice that  $Y \in \text{Out}$  and  $y \notin M_*$ . Hence, the event  $Y = y$  and the assumption that  $y \notin M_*$  imply that  $\text{Out} \neq M_*$ . Therefore, if  $z = y$  and  $y \notin M_*$ , then  $\Pr_0[\text{Out} \neq M_*] \geq \Pr_0[Y = y] \geq G$ .

Therefore, we proved that, in Case (1), the probability of error of algorithm  $\mathbb{A}$  is larger than  $G$  under  $H_0$ .

**Case (2):** ( $z = x \wedge x \notin M_*$ ) **or** ( $z = y \wedge y \in M_*$ ). By definition,  $Z$  is the arm with smallest number of

samples among arms in  $[L]$  and algorithm  $\mathbb{A}$  uses at most  $T$  samples. Therefore, we have

$$T_Z \leq \frac{T}{L}. \quad (101)$$

Then, we consider two cases separately.

**Case (2.1):**  $(z = x \wedge x \notin M_*)$ . We construct hypothesis  $H_1$  as follows

$$H_1 : \varphi_x = \mathcal{N}(w(x) + \Delta_L + \varepsilon, 1) \quad \text{and} \quad \varphi_e(e) = \mathcal{N}(w(e), 1) \quad \text{for all } e \neq y.$$

Notice that, by the choice of  $L$ , we have  $\Delta_x \geq \Delta_L$ . Hence we see that  $w_1(M_x) = w_0(M_x) + \Delta_L \geq w_0(M_x) + \Delta_x = w_0(M_*) = w_1(M_*)$ . Therefore, under  $H_1$ ,  $M_*$  is not the optimal set. Now we bound the likelihood ratio  $L_1(W)/L_0(W)$  as follows

$$\begin{aligned} \frac{L_1(W)}{L_0(W)} &= \prod_{i=1}^{T_Z} \exp \left( \frac{-(X_i - w(x) - \Delta_L)^2 + (X_i - w(x))^2}{2} \right) \\ &= \prod_{i=1}^{T_Z} \exp (\Delta_L(X_i - w(x)) - \Delta_L^2) \\ &= \exp (\Delta_L(K - T_Z w(x)) - \Delta_L^2 T_Z) \\ &= \exp (\Delta_L(K - T_Z w(x))) \exp (-\Delta_L^2 T_Z). \end{aligned} \quad (102)$$

Then we analyze the right-hand side of Eq. (102) as follows

$$\exp(-\Delta_L^2 T_Z) \geq \exp(-\Delta_L^2 T/L) \quad (103)$$

$$= \exp \left( -\frac{8}{16} \log(1/\theta) \right) \quad (104)$$

$$= \theta^{1/2}, \quad (105)$$

where Eq. (103) follows from the assumption that event  $\mathcal{S}$  occurred, which implies that event  $\mathcal{A}$  occurred and therefore  $T_e \leq 4t_e^*$ ; Eq. (104) follows from the definition of  $t_e^*$ .

Then, we bound the first term on the right-hand side of Eq. (32) as follows

$$\exp (\Delta_e(2K - 2p_e T_e)) \geq \exp \left( -2\Delta_e \sqrt{t_e^* \log(1/\theta)} \right) \quad (106)$$

$$= \exp \left( -\frac{2}{\sqrt{4}} \log(1/\theta) \right) \quad (107)$$

$$= \theta^{1/2}, \quad (108)$$

where Eq. (106) follows from the assumption that event  $\mathcal{S}$  occurred, which implies that event  $\mathcal{C}$  and therefore  $|2K - 2p_e T_e| \leq \sqrt{t_e^* \log(1/\theta)}$ ; Eq. (107) follows from the definition of  $t_e^*$ .

**Case (2.2):**  $y \notin M_*$ . By definition of  $y$ , we see that the event  $Y = y$  implies that  $y \notin \text{Out}$  and therefore  $\text{Out} \neq M_*$ . On the other hand, using Eq. (100), we have

$$\Pr_0[Y = y] \geq \Pr_0[\mathcal{C} \cap \{Y = y, Z = 0\}] = \frac{1}{2(L+1)^2}.$$

This gives that  $\Pr_0[\text{Out} \neq M_*] \geq \frac{1}{2(L+1)^2} \geq A$ . □

**Lemma 21.** *If Algorithm 1 stops, then  $w(M_*) - w(\text{Out}) \leq \epsilon$ .*

*Proof.* Suppose that  $\text{Out} \neq M_*$ . Suppose that the algorithm stops on round  $T$ , we know that  $\text{Out} = M_T$ . Consider the diff-set  $D = M_* \ominus M_T$  and the diff-set  $D_T$  as defined in Step 15 of Algorithm 1. By Lemma Z, we see that

$$w_T^+(D_T) = \max_{C: C \prec M_T} w_T^+(C) \geq w_T^+(D). \quad (109)$$

On the other hand, the stopping condition of Algorithm 1 gives that

$$\begin{aligned} \epsilon &\geq \tilde{w}_T(\tilde{M}_T) - \tilde{w}_T(M_T) \\ &= w_T^+(D_T) \geq w_T^+(D) \end{aligned} \quad (110)$$

$$\geq w(D) = w(M_*) - w(M_T), \quad (111)$$

where Eq. (110) follows from Eq. (109); Eq. (111) follows from the assumption that event  $\xi$  occurred.  $\square$

## References