Pure Exploration of Combinatorial Bandits

Shouyuan Chen

May 4, 2014

1 Pure Exploration of Combinatorial Bandits

Notations.
`ExpCMAB`: **problem formulation.** Let n denote the number of arms and suppose that the arms are numbered $1, 2, \ldots, n$. Each arm $e \in [n]$ is associated with a reward distribution φ_e. We assume that all reward distributions are R-sub-Gaussian []. Notice that all distributions that are supported on $[0, R]$ are R-sub-Gaussian distributions [] and therefore our model subsumes the cases of bounded rewards. Let $w(e)$ denote the expected reward of arm e, i.e. $w(e) = \mathbb{E}_{X \sim \varphi_e}[X]$. Further, we define $\boldsymbol{w} = \big(w(1), \ldots, w(n)\big)^T$ as the vector of expected rewards.

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 family 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 φ_{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 $\mathsf{Out} \in \mathcal{M}$ which maximizes the sum of expected weight, i.e. $\mathsf{Out} = M_*$, 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.

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 δ, the learner need to guarantee that $\Pr[\mathsf{Out} = M_*] \geq 1 - \delta$. 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 $\Pr[\mathsf{Out} \neq M_*]$ 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

In this section, we present `CGapExp`, a learning algorithm for the `ExpCMAB` problem in the fixed confidence setting, and analyze its sample complexity. The `CGapExp` algorithm can be extended to the fixed budget and

PAC learning settings. We will discuss this extension in Section X.

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 $\boldsymbol{v} = \big(v(1),\ldots,v(n)\big)^T$ as input and returns a set $\text{Oracle}(\boldsymbol{v}) = \arg\max_{M\in\mathcal{M}} v(M)$. It is clear that a large class of combinatorial problems admit efficient maximization oracles.

Algorithm. The `CGapExp` 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, `CGapExp` accesses the 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 $\bar{\boldsymbol{w}}_t$. Then, `CGapExp` explores possible refinements of M_t. In particular, `CGapExp` uses the confidence radius to compute an adjusted expectation vector $\tilde{\boldsymbol{w}}_t$ 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 $\tilde{\boldsymbol{w}}_t$ 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 $\tilde{\boldsymbol{w}}_t$ as input, which returns another set $\tilde{M}_t = \arg\max_{M\in\mathcal{M}} \tilde{w}_t(M)$. The `CGapExp` algorithm stops if $\tilde{w}_t(\tilde{M}_t) = \tilde{w}_t(M_t)$ and outputs $\text{Out} = M_t$. Otherwise, `CGapExp` pulls the arm belonging to the symmetric difference $(\tilde{M}_t\backslash M_t) \cup (M_t\backslash\tilde{M}_t)$ between M_t and $\tilde{M}_t$ with the largest confidence radius in the end of round t. The pseudo-code of `CGapExp` is shown in Algorithm 1.

Algorithm 1 `CGapExp`: Combinatorial Gap Exploration

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

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

1: **for** $t = n, n+1, \ldots$ **do**
2:     $M_t \leftarrow \text{Oracle}(\bar{\boldsymbol{w}}_t)$
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:    $\tilde{M}_t \leftarrow \text{Oracle}(\tilde{\boldsymbol{w}}_t)$
11:    **if** $\tilde{w}_t(\tilde{M}_t) = \tilde{w}_t(M_t)$ **then**
12:      $\text{Out} \leftarrow M_t$
13:      **return** Out
14:    **end if**
15:    $p_t \leftarrow \arg\max_{e\in(\tilde{M}_t\backslash M_t)\cup(M_t\backslash\tilde{M}_t)} \text{rad}_t(e)$
16:    Pull arm p_t and observe the reward
17:    Update empirical means $\bar{\boldsymbol{w}}_{t+1}$ using the observed reward
18:    Update number of pulls: $T_{t+1}(p_t) \leftarrow T_t(p_t) + 1$ and $T_{t+1}(e) \leftarrow T_t(e)$ for all $e \neq p_t$
19: **end for**

2.1 Analysis

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

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

where we use the convention that the maximum value of an empty set is $-\infty$. We also define $\mathbf{H}$ as the sum of inverse squared gaps

$$\mathbf{H} = \sum_{e \in [n]} \Delta_e^{-2}.$$

From Eq. (1), we see that, for each arm $e \notin M_*$, Δ_e represents the gap between the optimal set M_* and the best set that includes arm e; and, for each arm $e \in M_*$, Δ_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 capture these properties, we introduce notions of *exchange set* and *exchange class* as tools for our 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 $M \oplus b \triangleq M \backslash b_- \cup b_+$. Similarly, we also define operator $\ominus$ such that $M \ominus b \triangleq M \backslash b_+ \cup b_-$. 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 $e \in (M \backslash M')$, there exists an exchange set $(b_+, b_-) \in \mathcal{B}$ which satisfies $e \in b_-$, $b_+ \subseteq M' \backslash M$, $b_- \subseteq M \backslash M'$, $(M \oplus b) \in \mathcal{M}$ and $(M' \ominus b) \in \mathcal{M}$. Finally, we define the *width* of exchange class $\mathcal{B}$ as follows

$$\text{width}(\mathcal{B}) = \max_{(b_+, b_-) \in \mathcal{B}} |b_+| + |b_-|. \tag{2}$$

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 $b_1, \ldots, b_k$ of $\mathcal{B}$ such that $M' = M \oplus b_1 \ldots \oplus b_k$.

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

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

    1. $\mathcal{B}_{\textbf{Multi}} = \big\{(\{i\}, \{j\}) \mid \forall i \in [n], j \in [n]\big\}$.

    2. $\mathcal{B}_{\textbf{Matroid}} = \big\{(\{i\}, \{j\}) \mid \forall i \in [n], j \in [n]\big\}$.

    3. $\mathcal{B}_{\textbf{Match}} = \big\{(C_+, C_-) \mid C_+ \cup C_-$ *is a cycle of* $G\big\}$.

    4. $\mathcal{B}_{\textbf{Path}} = \big\{(P_1, P_2) \mid P_1, P_2$ *are two disjoint paths of* G *with same endpoints*$\big\}$.

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

The construction for **Multi** problem is straightforward. For **Matroid** problem, we leverage the basis exchange property of matroids (see Lemma 15 in the appendix). And for **Match** and **Path** problems, we appeal to graph-theoretical properties of matchings and paths. We illustrate these exchanges classes in Figure 1. A detailed proof of Lemma 1 is deferred to the supplementary material.
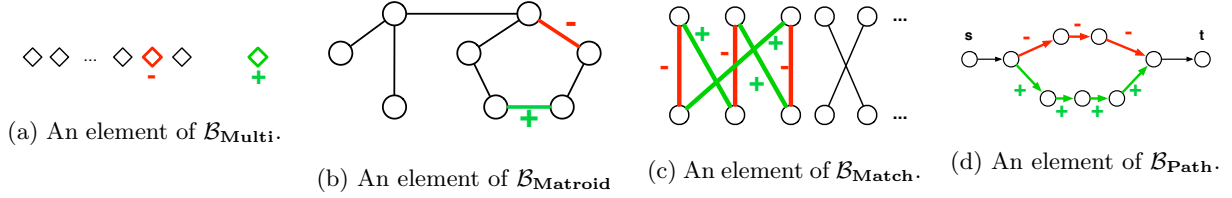
(a) An element of $\mathcal{B}_{\mathbf{Multi}}$.

(b) An element of $\mathcal{B}_{\mathbf{Matroid}}$

(c) An element of $\mathcal{B}_{\mathbf{Match}}$.

(d) An element of $\mathcal{B}_{\mathbf{Path}}$.

Figure 1: Elements of exchange classes $\mathcal{B}_{\mathbf{Multi}}$, $\mathcal{B}_{\mathbf{Matroid}}$, $\mathcal{B}_{\mathbf{Match}}$ and $\mathcal{B}_{\mathbf{Path}}$ for running examples. (In Figure 1b, we use spanning tree as a specific instance for the **Matroid** problem.)

Main result. Our main result is a problem-dependent sample complexity bound of `CGapExp` algorithm. In particular, we show that `CGapExp` returns the optimal set with high probability and uses at most $\tilde{O}\big(\text{width}(\mathcal{B})^2\mathbf{H}\big)$ samples.

Theorem 1. *Given any $\delta \in (0,1)$, any combinatorial problem $\mathcal{M} \subseteq 2^{[n]}$ and any vector $\boldsymbol{w} \in \mathbb{R}^n$. Assume that the reward distribution φ_e for each arm $e \in [n]$ is R-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).*

Set $\text{rad}_t(e) = R\sqrt{\frac{2\log\left(\frac{4nt^2}{\delta}\right)}{T_e(t)}}$ for all $t > 0$ and $e \in [n]$. Then, with probability at least $1 - \delta$, the `CGapExp` algorithm (Algorithm 1) returns the optimal set $\mathsf{Out} = M_*$ and

$$T \leq O\left(R^2 \,\text{width}(\mathcal{B})^2\mathbf{H}\log\left(R^2\,\text{width}(\mathcal{B})^2\mathbf{H}n/\delta\right)\right),$$

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 $\text{width}(\mathcal{B}_{\mathbf{Multi}}) = O(1)$. Therefore, the sample complexity bound of our algorithm for the **Multi** problem is $O\big(\mathbf{H}\log(n\delta^{-1}\mathbf{H})\big)$. This matches the best known problem-dependent bounds for the **Multi** problem due to XXX [], within logarithmic factors.

3 Lower Bounds

In this section, 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 *δ-correct algorithm* as follows. For any $\delta \in (0,1)$, we call an algorithm $\mathbb{A}$ a δ-correct algorithm if, for any expected reward $\boldsymbol{w} \in \mathbb{R}^n$, the probability of error of $\mathbb{A}$ is at most δ, i.e $\Pr[M_* \neq \mathsf{Out}] \leq \delta$, where Out is the output of algorithm $\mathbb{A}$.

We show that, for any combinatorial problem $\mathcal{M}$ and any expected rewards $\boldsymbol{w}$, any δ-correct algorithm $\mathbb{A}$ must use at least $\tilde{\Omega}(\mathbf{H})$ samples in expectation.

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

$$\mathbb{E}[T] \geq \frac{1}{16}\mathbf{H}\log(\frac{1}{4\delta}),$$

where T denote the number of total samples used by algorithm $\mathbb{A}$ and Δ_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**. We see that `CGapExp` algorithm uses at most $\tilde{O}(\text{width}(\mathcal{B})^2\mathbf{H})$ 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 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 $\boldsymbol{w}$. This means that the sample complexity of Algorithm 1 has an optimal dependency on the gaps $\{\Delta_e\}_{e \in [n]}$.

Next, we argue that the dependency on width$(\mathcal{B})$ on the sample complexity might be intrinsic. To this end, we provide evidence showing that the sample complexity of any δ-correct algorithm should be related to size of exchange sets. In fact, we show that, for any optimal exchange set $b \in \mathcal{B}_{\mathsf{opt}}$ and any δ-correct algorithm, the algorithm must spend $\tilde{\Omega}\left(|b|^2/w(b)^2\right)$ 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 $\boldsymbol{w} \in \mathbb{R}^n$. Suppose that, for each arm $e \in [n]$, the reward distribution φ_e is given by $\varphi_e = \mathcal{N}(w(e), 1)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2. Fix any $\delta \in (0, e^{-16}/4)$ and any δ-correct algorithm $\mathbb{A}$.*
Then, for any $b \in \mathcal{B}_{\mathsf{opt}}$, 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

4   Extensions

4.1   Fixed Budget Setting

We can extend the `CGapExp` algorithm to the fixed budget setting through two simple modifications: (1) requiring `CGapExp` to terminate after T rounds; and (2) using a different construction of confidence intervals. The first modification ensures that `CGapExp` use at most T samples, which meets the requirement of the fixed budget setting. And the second modification guarantees that the confidence intervals are valid for all arms in T rounds. The following theorem shows that the probability of error of the modified `CGapExp` is bounded by $O\left(Tn \exp\left(\frac{-T}{\text{width}(\mathcal{B})^2 \mathbf{H}}\right)\right)$.

Theorem 4. *Use the same notations as in Theorem 1. Given $T > 0$ and parameter $\alpha > 0$, set the confidence radius $\text{rad}_t(e) = R\sqrt{\frac{\alpha}{T_e(t)}}$ for all arms $e \in [n]$ and all $t > 0$. Run `CGapExp` algorithm for at most T rounds. Then, we have*

$$\Pr\left[\mathsf{Out} \neq M_*\right] \leq 2Tn \exp\left(-2\alpha\right), \tag{3}$$

as long as $\alpha < 9T\left(R^2 \text{width}(\mathcal{B})^2 \mathbf{H}\right)^{-1}$.

The right-hand side of Eq. (3) equals to $O\left(Tn \exp\left(\frac{-T}{\text{width}(\mathcal{B})^2 \mathbf{H}}\right)\right)$ when parameter $\alpha = O(T\mathbf{H}^{-1} \text{width}(\mathcal{B})^{-2})$. For **Multi** problem, we see that this matches the guarantees of previous fixed budget algorithm [], up to logarithmic factors.

4.2   PAC Learning

Now we consider a setting where the learner is only required to report an approximately optimal set of arms. More specifically, we consider the notion of (ϵ, δ)-PAC algorithm. Formally, an algorithm $\mathbb{A}$ is called an (ϵ, δ)-PAC algorithm if its output Out satisfies $\Pr\left[w(M_*) - w(\mathsf{Out}) > \epsilon\right] \leq \delta$.

We show that a simple modification on the `CGapExp` algorithm gives an (ϵ, δ)-PAC algorithm, with guarantees similar to Theorem 1. In fact, the only modification needed is to change the stopping condition from $\tilde{w}_t(\tilde{M}_t) \leq \tilde{w}_t(M_t)$ to $w(\tilde{M}_t) - w(M_t) \leq \epsilon$ on line XXX of Algorithm 1. We denote the modified algorithm as

---

`CGapExpPAC`. In the following theorem, we show that `CGapExpPAC` is indeed an (ϵ, δ)-PAC algorithm and has sample complexity similar to Theorem 1.

Theorem 5. *Use the same notations as in Theorem 1. Fix $\delta \in (0,1)$ and $\epsilon \geq 0$. Then, with probability at least $1 - \delta$, the output* Out *of **CGapExpPAC** satisfies $w(M_*) - w(\textsf{Out}) \leq \epsilon$. In addition, the number of samples T used by the algorithm satisfies*

$$T \leq O\left(R^2 \sum_{e \in [n]} \min\left\{ \frac{\text{width}(\mathcal{B})^2}{\Delta_e^2}, \frac{K^2}{\epsilon^2} \right\} \log\left(\frac{R^2 n}{\delta} \sum_{e \in [n]} \min\left\{ \frac{\text{width}(\mathcal{B})^2}{\Delta_e^2}, \frac{K^2}{\epsilon^2} \right\} \right) \right), \tag{4}$$

where $K = \max_{M \in \mathcal{M}} |M|$ is the size of largest feasible solution.

We see that the sample complexity of `CGapExpPAC` decreases when ϵ increases. And if $\epsilon = 0$, the sample complexity Eq. (4) of `CGapExpPAC` equals to that of `CGapExp`. (difference??)

5   Proof of Main Result

In this section, we prove our main result: Theorem 1.

Notations. We need some additional notations for our analysis. For any set $a \subseteq [n]$, let $\boldsymbol{\chi}_a \in \{0,1\}^n$ 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 $\boldsymbol{\chi}_b \triangleq \boldsymbol{\chi}_{b_+} - \boldsymbol{\chi}_{b_-}$ as the incidence vector of b. We notice that $\boldsymbol{\chi}_b \in \{-1, 0, 1\}^n$.

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

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

5.1   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*

$$\boldsymbol{\chi}_{M_1} + \boldsymbol{\chi}_b = \boldsymbol{\chi}_{M_2}.$$

Proof. Recall that $M_2 = M_1 \backslash b_- \oplus b_+$ and $b_+ \cap b_- = \emptyset$. Therefore we see that $M_2 \backslash M_1 = b_+$ and $M_1 \backslash M_2 = b_-$. Then, we decompose $\boldsymbol{\chi}_{M_1}$ as $\boldsymbol{\chi}_{M_1} = \boldsymbol{\chi}_{M_1 \backslash M_2} + \boldsymbol{\chi}_{M_1 \cap M_2}$. Hence, we have

$$\begin{aligned} \boldsymbol{\chi}_{M_1} + \boldsymbol{\chi}_b &= \boldsymbol{\chi}_{M_1 \backslash M_2} + \boldsymbol{\chi}_{M_1 \cap M_2} + \boldsymbol{\chi}_{b_+} - \boldsymbol{\chi}_{b_-} \\ &= \boldsymbol{\chi}_{M_1 \cap M_2} + \boldsymbol{\chi}_{M_2 \backslash M_1} \\ &= \boldsymbol{\chi}_{M_2}. \end{aligned}$$

$\square$

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 $e \in (M \backslash M') \cup (M' \backslash M)$, there exists an exchange set $b = (b_+, b_-) \in \mathcal{B}$ such that $e \in (b_+ \cup b_-)$, $b_- \subseteq (M \backslash M')$, $b_+ \subseteq (M' \backslash M)$, $(M \oplus b) \in \mathcal{M}$ and $(M' \ominus b) \in \mathcal{M}$. Moreover, if $M' = M_*$, then we have $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e > 0$, where Δ_e is the gap defined in Eq. (1).*

Proof. We decompose our proof into two cases.

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

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

$$w(M_*) - w(M_0) = \langle \boldsymbol{w}, \boldsymbol{\chi}_{M_*} \rangle - \langle \boldsymbol{w}, \boldsymbol{\chi}_{M_0} \rangle = \langle \boldsymbol{w}, \boldsymbol{\chi}_{M_*} - \boldsymbol{\chi}_{M_0} \rangle = \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle.$$

Therefore, we obtain $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e$.

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

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

We construct $b = (b_+, b_-)$ by setting $b_+ = c_-$ and $b_- = c_+$. Notice that, by the construction of b, we have $M \oplus b = M \ominus c$ and $M' \ominus b = M' \oplus c$. 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 $M_2 = M_* \ominus b$ 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

$$\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle = w(M_*) - w(M_2) \geq \Delta_e.$$

$\square$

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

$$\max_{e \in (M \backslash M') \cup (M' \backslash M)} \mathrm{rad}_t(e) = \big\| \mathbf{rad}_t \circ |\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_M| \big\|_\infty.$$

Proof. Notice that $\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_M = \boldsymbol{\chi}_{M' \backslash M} - \boldsymbol{\chi}_{M \backslash M'}$. In addition, since $(M' \backslash M) \cap (M \backslash M') = \emptyset$, we have $\boldsymbol{\chi}_{M' \backslash M} \circ \boldsymbol{\chi}_{M \backslash M'} = \mathbf{0}_n$. Also notice that $\boldsymbol{\chi}_{M' \backslash M} - \boldsymbol{\chi}_{M \backslash M'} \in \{-1, 0, 1\}^n$. Therefore, we have

$$\begin{aligned}
|\boldsymbol{\chi}_{M' \backslash M} - \boldsymbol{\chi}_{M \backslash M'}| &= (\boldsymbol{\chi}_{M' \backslash M} - \boldsymbol{\chi}_{M \backslash M'})^2 \\
&= \boldsymbol{\chi}_{M' \backslash M}^2 + \boldsymbol{\chi}_{M \backslash M'}^2 + 2\boldsymbol{\chi}_{M' \backslash M} \circ \boldsymbol{\chi}_{M \backslash M'} \\
&= \boldsymbol{\chi}_{M' \backslash M} + \boldsymbol{\chi}_{M \backslash M'} \\
&= \boldsymbol{\chi}_{(M' \backslash M) \cup (M \backslash M')},
\end{aligned}$$

where the third equation follows from the fact that $\boldsymbol{\chi}_{M \backslash M'} \in \{0, 1\}^n$ and $\boldsymbol{\chi}_{M' \backslash M} \in \{0, 1\}^n$. The lemma follows immediately from the fact that $\mathrm{rad}_t(e) \geq 0$ and $\boldsymbol{\chi}_{(M \backslash M') \cup (M' \backslash M)} \in \{0, 1\}^n$. $\square$

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

Proof. We have

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

$\square$

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

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

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

$$
\begin{aligned}
\langle \tilde{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \rangle &= \langle \bar{\boldsymbol{w}}_t + \mathbf{rad}_t \circ (1 - 2\boldsymbol{\chi}_{M_t}), \, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \rangle \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, (\mathbf{1}_n - 2\boldsymbol{\chi}_{M_t}) \circ (\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t}) \rangle && (5) \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} - 2\boldsymbol{\chi}_{M_t} \circ \boldsymbol{\chi}_{M'} + 2\boldsymbol{\chi}_{M_t}^2 \rangle \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, \boldsymbol{\chi}_{M'}^2 - \boldsymbol{\chi}_{M_t}^2 - 2\boldsymbol{\chi}_{M_t} \circ \boldsymbol{\chi}_{M'} + 2\boldsymbol{\chi}_{M_t}^2 \rangle && (6) \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, (\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t})^2 \rangle \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t}| \rangle , && (7)
\end{aligned}
$$

where Eq. (5) follows from Lemma 5; Eq. (6) holds since $\boldsymbol{\chi}_{M'} \in \{0,1\}^n$ and $\boldsymbol{\chi}_{M_t} \in \{0,1\}^n$ and therefore $\boldsymbol{\chi}_{M'} = \boldsymbol{\chi}_{M'}^2$ and $\boldsymbol{\chi}_{M_t} = \boldsymbol{\chi}_{M_t}^2$; and Eq. (7) follows since $\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \in \{-1, 0, 1\}^n$. □

5.2 Confidence Intervals

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

$$\xi_t = \left\{ \forall i \in [n], \quad |w(i) - \bar{w}_t(i)| \le \mathrm{rad}_t(i) \right\}. \tag{8}$$

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

If the confidence bounds are valid, we can generalize Eq. (8) to inner products as follows.

Lemma 7. *Given any $t > 0$, assume that event ξ_t as defined in Eq. (8) occurs. Then, for any vector $\boldsymbol{a} \in \mathbb{R}^n$, we have*

$$\left| \langle \boldsymbol{w}, \boldsymbol{a} \rangle - \langle \bar{\boldsymbol{w}}_t, \boldsymbol{a} \rangle \right| \le \langle \mathbf{rad}_t, |\boldsymbol{a}| \rangle .$$

Proof. Suppose that ξ occurs. Then, we have

$$
\begin{aligned}
\left| \langle \boldsymbol{w}, \boldsymbol{a} \rangle - \langle \bar{\boldsymbol{w}}_t, \boldsymbol{a} \rangle \right| &= \left| \langle \boldsymbol{w} - \bar{\boldsymbol{w}}_t, \boldsymbol{a} \rangle \right| \\
&= \left| \sum_{i=1}^n \big(w(i) - \bar{w}_t(i) \big) a(i) \right| \\
&\le \sum_{i=1}^n \big| w(i) - \bar{w}_t(i) \big| |a(i)| \\
&\le \sum_{i=1}^n \mathrm{rad}_t(i) \cdot |a(i)| && (9) \\
&= \langle \mathbf{rad}_t, |\boldsymbol{a}| \rangle ,
\end{aligned}
$$

where Eq. (9) follows the definition of event ξ_t in Eq. (8) and the assumption that it occurs. □

Next, we construct the high probability confidence intervals for the fixed confidence setting.

Lemma 8. *Suppose that the reward distribution φ_e is a R-sub-Gaussian distribution for all $e \in [n]$. And if, for all $t > 0$ and all $e \in [n]$, the confidence radius $\mathrm{rad}_t(e)$ is given by*

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

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 φ_e is a R-sub-Gaussian distribution with mean $w(e)$ and $w_t(e)$ is the empirical mean of φ_e for $T_e(t)$ samples. Using Hoeffding's inequality (see Lemma 16 in Section 8), we obtain

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

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

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

$$\geq 1 - \sum_{t=1}^{\infty} \frac{\delta}{2t^2}$$

$$= 1 - \frac{\pi^2}{12}\delta \geq 1 - \delta.$$

$\square$

5.3 Main Lemmas

Lemma 9. *Given any $t > 0$, assume that event ξ_t (defined in Eq. (8)) 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 $\langle \boldsymbol{w}, \boldsymbol{\chi}_{M_*}\rangle > \langle \boldsymbol{w}, \boldsymbol{\chi}_{M_t}\rangle$.
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

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

$$= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'_t} - \boldsymbol{\chi}_{M_t}\right\rangle + \left\langle \mathbf{rad}_t, |\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t}|\right\rangle \tag{10}$$

$$\geq \left\langle \boldsymbol{w}, \boldsymbol{\chi}_{M'_t} - \boldsymbol{\chi}_{M_t}\right\rangle \tag{11}$$

$$= w(M'_t) - w(M_t) > 0, \tag{12}$$

where Eq. (10) follows from Lemma 6; and Eq. (11) follows the assumption that event ξ_t occurs and Lemma 7; Therefore Eq. (12) shows that $\tilde{w}_t(\tilde{M}_t) > \tilde{w}_t(M_t)$. However, this contradicts to the stopping condition of `CGapExp`: $\tilde{w}_t(\tilde{M}_t) \leq \tilde{w}_t(M_t)$ and the assumption that the algorithm terminates on round t. $\square$

Lemma 10. *Given any $t > 0$ and suppose that event ξ_t (defined in Eq. (8)) occurs. For any $e \in [n]$, if $\mathrm{rad}_t(e) < \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}$, 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 $e \in (c_+ \cup c_-)$, $c_- \subseteq (M_t \backslash \tilde{M}_t)$, $c_+ \subseteq (\tilde{M}_t \backslash M_t)$, $(M_t \oplus c) \in \mathcal{M}$ 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 $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e > 0$.
Now, we define vectors $\boldsymbol{d} = \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t}$, $\boldsymbol{d}_1 = \boldsymbol{\chi}_{M_t'} - \boldsymbol{\chi}_{M_t}$ and $\boldsymbol{d}_2 = \boldsymbol{\chi}_{M_t' \oplus b} - \boldsymbol{\chi}_{M_t}$. By the definition of M_t' and Lemma 3, we see that $\boldsymbol{d}_1 = \boldsymbol{d} - \boldsymbol{\chi}_c$ and $\boldsymbol{d}_2 = \boldsymbol{d}_1 + \boldsymbol{\chi}_b = \boldsymbol{d} - \boldsymbol{\chi}_c + \boldsymbol{\chi}_b$.
Then, we claim that $\|\mathbf{rad}_t \circ (\boldsymbol{d} - \boldsymbol{\chi}_c)\|_\infty < \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}$. Since $c_- \subseteq M_t$ and $c_+ \cap M_t = \emptyset$, using standard set theoretical manipulations, we can show that $M_t \backslash \tilde{M}_t = (M_t \backslash M_t') \cup c_-$. Similarly, one can show that $\tilde{M}_t \backslash M_t = (M_t' \backslash M_t) \cup c_+$. This means that $\big((M_t \backslash M_t') \cup (M_t' \backslash M_t)\big) \subseteq \big((M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)\big)$. Then, applying Lemma Y, we obtain

$$
\begin{aligned}
\|\mathbf{rad}_t \circ (\boldsymbol{d} - \boldsymbol{\chi}_c)\|_\infty &= \left\|\mathbf{rad}_t \circ (\boldsymbol{\chi}_{M_t'} - \boldsymbol{\chi}_{M_t})\right\|_\infty \\
&= \max_{i \in (M_t \backslash M_t') \cup (M_t' \backslash M_t)} \mathrm{rad}_t(i) \\
&\leq \max_{i \in (M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)} \mathrm{rad}_t(i) \\
&= \mathrm{rad}_t(e) < \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}.
\end{aligned} \tag{13}
$$

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

$$
\begin{aligned}
\|\mathbf{rad}_t \circ |\boldsymbol{\chi}_c|\|_\infty &= \max_{i \in c_+ \cup c_-} \mathrm{rad}_t(i) \\
&\leq \max_{i \in (\tilde{M}_t \backslash M_t) \cup (M_t \backslash \tilde{M}_t)} \mathrm{rad}_t(i) \\
&= \mathrm{rad}_t(e) < \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}.
\end{aligned} \tag{14}
$$

Next, we claim that $\boldsymbol{d} \circ \boldsymbol{\chi}_c = |\boldsymbol{\chi}_c|$. Recall that $\boldsymbol{\chi}_c = \boldsymbol{\chi}_{c_+} - \boldsymbol{\chi}_{c_-}$ and $\boldsymbol{d} = \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t} = \boldsymbol{\chi}_{\tilde{M}_t \backslash M_t} - \boldsymbol{\chi}_{M_t \backslash \tilde{M}_t}$. We also notice that $c_+ \subseteq (\tilde{M}_t \backslash M_t)$ and $c_- \subseteq (M_t \backslash \tilde{M}_t)$. This implies that $c_+ \cap (M_t \backslash \tilde{M}_t) = \emptyset$ and $c_- \cap (\tilde{M}_t \backslash M_t) = \emptyset$. Therefore, we have

$$
\begin{aligned}
\boldsymbol{d} \circ \boldsymbol{\chi}_c &= (\boldsymbol{\chi}_{\tilde{M}_t \backslash M_t} - \boldsymbol{\chi}_{M_t \backslash \tilde{M}_t}) \circ (\boldsymbol{\chi}_{c_+} - \boldsymbol{\chi}_{c_-}) \\
&= \boldsymbol{\chi}_{\tilde{M}_t \backslash M_t} \circ \boldsymbol{\chi}_{c_+} + \boldsymbol{\chi}_{M_t \backslash \tilde{M}_t} \circ \boldsymbol{\chi}_{c_-} - \boldsymbol{\chi}_{\tilde{M}_t \backslash M_t} \circ \boldsymbol{\chi}_{c_-} - \boldsymbol{\chi}_{M_t \backslash \tilde{M}_t} \circ \boldsymbol{\chi}_{c_+} \\
&= \boldsymbol{\chi}_{\tilde{M}_t \backslash M_t} \circ \boldsymbol{\chi}_{c_+} + \boldsymbol{\chi}_{M_t \backslash \tilde{M}_t} \circ \boldsymbol{\chi}_{c_-} \\
&= \boldsymbol{\chi}_{c_+} + \boldsymbol{\chi}_{c_-} = |\boldsymbol{\chi}_c|.
\end{aligned}
$$

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

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

$$\langle \mathbf{rad}_t, |\boldsymbol{d}_2| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle = \langle \mathbf{rad}_t, |\boldsymbol{d}_2| - |\boldsymbol{d}| \rangle = \langle \mathbf{rad}_t, \boldsymbol{d}_2^2 - \boldsymbol{d}^2 \rangle \tag{15}$$

$$= \langle \mathbf{rad}_t, (\boldsymbol{d} - \boldsymbol{\chi}_c + \boldsymbol{\chi}_b)^2 - \boldsymbol{d}^2 \rangle$$

$$= \langle \mathbf{rad}_t, \boldsymbol{\chi}_b^2 + \boldsymbol{\chi}_c^2 - 2\boldsymbol{\chi}_b \circ \boldsymbol{\chi}_c - 2\boldsymbol{d} \circ \boldsymbol{\chi}_c + 2\boldsymbol{d} \circ \boldsymbol{\chi}_b \rangle$$

$$= \langle \mathbf{rad}_t, \boldsymbol{\chi}_b^2 - \boldsymbol{\chi}_c^2 + 2\boldsymbol{\chi}_b \circ (\boldsymbol{d} - \boldsymbol{\chi}_c) \rangle \tag{16}$$

$$= \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - 2 \langle \mathbf{rad}_t, \boldsymbol{\chi}_b \circ (\boldsymbol{d} - \boldsymbol{\chi}_c) \rangle$$

$$= \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - 2 \langle \mathbf{rad}_t \circ (\boldsymbol{d} - \boldsymbol{\chi}_c), \boldsymbol{\chi}_b \rangle \tag{17}$$

$$\geq \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - 2 \|\mathbf{rad}_t \circ (\boldsymbol{d} - \boldsymbol{\chi}_c)\|_\infty \|\boldsymbol{\chi}_b\|_1 \tag{18}$$

$$> \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - \frac{2\Delta_e}{3 \operatorname{width}(\mathcal{B})} \|\boldsymbol{\chi}_b\|_1 \tag{19}$$

$$\geq \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - \frac{2\Delta_e}{3}, \tag{20}$$

where Eq. (15) holds since $\boldsymbol{d} \in \{-1, 0, 1\}^n$ and $\boldsymbol{d}_2 \in \{-1, 0, 1\}^n$; Eq. (16) follows from the claim that $\boldsymbol{d} \circ \boldsymbol{\chi}_c = |\boldsymbol{\chi}_c| = \boldsymbol{\chi}_c^2$; Eq. (17) and Eq. (18) follow from Lemma 5 and Hölder's inequality; Eq. (19) follows from Eq. (13); and Eq. (20) holds since $b \in \mathcal{B}$ and $\|\boldsymbol{\chi}_b\|_1 = |b_+| + |b_-| \leq \operatorname{width}(\mathcal{B})$.

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

$$\langle \bar{\boldsymbol{w}}_t, \boldsymbol{d} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle = \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t}| \rangle$$

$$= \tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t)$$

$$\geq \tilde{w}_t(M_t') - \tilde{w}_t(M_t)$$

$$= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t'} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_{M_t'} - \boldsymbol{\chi}_{M_t}| \rangle$$

$$= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{d}_2 \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_2| \rangle$$

$$= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{d} \rangle - \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_c \rangle + \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_2| \rangle,$$

where the last equality follows from the fact that $\boldsymbol{d}_2 = \boldsymbol{d} - \boldsymbol{\chi}_c + \boldsymbol{\chi}_b$. Rearranging the above inequality, we obtain

$$\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_c \rangle \geq \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_2| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle$$

$$\geq \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - \frac{2\Delta_e}{3} \tag{21}$$

$$> \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \langle \mathbf{rad}_t, \boldsymbol{\chi}_c \rangle - \frac{2\Delta_e}{3} \tag{22}$$

$$> \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{\Delta_e}{3} - \frac{2\Delta}{3} \tag{23}$$

$$= \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \Delta_e \geq 0, \tag{24}$$

where Eq. (21) uses Eq. (20); Eq. (22) follows from the assumption that event ξ occurs and Lemma X; and Eq. (22) holds since Eq. (14).

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

$$\bar{w}_t(M_t') = \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t'} \rangle = \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} + \boldsymbol{\chi}_c \rangle = \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \rangle + \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_c \rangle > \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \rangle = w_t(M_t).$$

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 $e \in (b_+ \cup b_-)$, $b_- \subseteq (\tilde{M}_t \backslash M_*)$, $b_+ \subseteq (M_* \backslash \tilde{M}_t)$ and $(\tilde{M}_t \oplus b) \in \mathcal{M}$. Lemma 3 also indicates that $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e > 0$.

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

$$\|\mathbf{rad}_t \circ \boldsymbol{d}\|_\infty \leq \max_{i : (\tilde{M}_t \backslash M_t) \cup (M_t \backslash \tilde{M}_t)} \mathrm{rad}_t(i) = \mathrm{rad}_t(e) < \frac{\Delta_e}{3}. \tag{25}$$

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

$$\begin{aligned}
\langle \bar{\boldsymbol{w}}_t, \boldsymbol{d}_1 \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_1| \rangle - \langle \bar{\boldsymbol{w}}_t, \boldsymbol{d} \rangle - \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle &= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_1| - |\boldsymbol{d}| \rangle \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, \boldsymbol{d}_1^2 - \boldsymbol{d}^2 \rangle \tag{26} \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, 2\boldsymbol{d} \circ \boldsymbol{\chi}_b + \boldsymbol{\chi}_b^2 \rangle \tag{27} \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, \boldsymbol{\chi}_b^2 \rangle + 2 \langle \mathbf{rad}_t \circ \boldsymbol{d}, \boldsymbol{\chi}_b \rangle \\
&\geq \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - 2 \langle \mathbf{rad}_t \circ \boldsymbol{d}, \boldsymbol{\chi}_b \rangle \tag{28} \\
&\geq \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - 2 \|\mathbf{rad}_t \circ \boldsymbol{d}\|_\infty \|\boldsymbol{\chi}_b\|_1 \tag{29} \\
&> \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{2\Delta_e}{3} \tag{30} \\
&\geq 0, \tag{31}
\end{aligned}$$

where Eq. (26) follows from the fact that $\boldsymbol{d}_1 \in \{-1, 0, 1\}^n$ and $\boldsymbol{d} \in \{-1, 0, 1\}^n$; Eq. (27) holds since $\boldsymbol{d}_1 = \boldsymbol{d} + \boldsymbol{\chi}_b$; Eq. (28) follows from the assumption that ξ occurs and Lemma X; Eq. (29) follows from Lemma 5 and Hölder's inequality; and Eq. (30) is due to Eq. (25).

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

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

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

$\square$

5.4 Proof of Theorem 1

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

Proof. Lemma 8 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 ξ, we see that $\mathsf{Out} = M_*$. Next, we focus on bounding the total number T of samples.

Fix any arm $e \in [n]$. Let T_e denote the total number of pull of arm $e \in [n]$. Let t_e be the last round which arm e is pulled, i.e. $p_{t_e} = e$. It is easy to see that $T_e(t_e) = T_e - 1$. By Lemma 10, we see that

$\mathrm{rad}_{t_e}(e) \geq \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}$. Plugging in the construction radius rad, we have

$$\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})} \leq R\sqrt{\frac{2\log\left(4nt_e^2/\delta\right)}{T_e - 1}} \leq R\sqrt{\frac{2\log\left(4nT^2/\delta\right)}{T_e - 1}}. \tag{32}$$

Solving Eq. (32) for T_e, we obtain

$$T_e \leq \frac{18\,\mathrm{width}(\mathcal{B})^2 R^2}{\Delta_e^2}\log(4nT^2/\delta) + 1. \tag{33}$$

Notice that $T = \sum_{i\in[n]} T_i$. Hence the theorem follows by summing up Eq. (33) for all $e \in [n]$ and solving for T. $\qquad\square$

6   Proof of Lower Bounds

Lemma 11.
$$\Delta_e = \min_{b:e\in b, b\in\mathcal{B}_{\mathsf{opt}}} w(b).$$

Proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof. Fix $\delta > 0$, $\boldsymbol{w} = \{w(1), \ldots, w(n)\}$ and a δ-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

$$\mathbb{E}[T_e] \geq \frac{1}{16\Delta_e^2}\log(1/4\delta). \tag{34}$$

Notice that the theorem follows immediately by summing up Eq. (34) for all $e \in [n]$.

Fix an arm $e \in [n]$. We now focus on proving Eq. (34). 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 Δ_e, we know that $w(M_*) - w(M_e) = \Delta_e$.

Let $\boldsymbol{w}_0$ and $\boldsymbol{w}_1$ 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, $w_1(M_*) - w_1(M_e) = -\Delta_e < 0$. 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

$$t_e^* = \frac{1}{16\Delta_e^2}\log\left(\frac{1}{\theta}\right). \tag{35}$$

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 $\text{Pr}_0[\mathcal{A}] \geq 3/4$. This can be proved by Markov inequality as follows.

$$\text{Pr}_0[T_e > 4t_e^*] \leq \frac{\mathbb{E}_0[T_e]}{4t_e^*}$$
$$= \frac{t_e^*}{4t_e^*} = \frac{1}{4}.$$

Let $X_1, \ldots, X_{T_e}$ 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 $\text{Pr}_0[\mathcal{C}] \geq 3/4$. First, notice that $K_t(e) - p_e t$ is a martingale under H_0. Then, by Kolmogorov's inequality, we have

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

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.

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

Since the probability of error of the algorithm is smaller than $\delta < 1/4$, we have $\text{Pr}_0[\mathcal{B}] \geq 3/4$. Define $\mathcal{S}$ be $\mathcal{S} = \mathcal{A} \cap \mathcal{B} \cap \mathcal{C}$. Then, by union bound, we have $\text{Pr}_0[\mathcal{S}] \geq 1/4$.
Now, we show that if $\mathbb{E}_0[T_e] \leq t_e^*$, then $\text{Pr}_1[\mathcal{B}] \geq \delta$. 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 μ and variance σ^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

$$\frac{L_1(W)}{L_0(W)} = \prod_{i=1}^{T_e} \exp\left(\frac{-(X_i - w(e) - 2\Delta_e)^2 + (X_i - w(e))^2}{2}\right)$$
$$= \prod_{i=1}^{T_e} \exp\left(\Delta_e(2X_i - 2w(e)) - 2\Delta_e^2\right)$$
$$= \exp\left(\Delta_e(2K - 2w(e)T_e) - 2\Delta_e^2 T_e\right)$$
$$= \exp\left(\Delta_e(2K - 2w(e)T_e)\right) \exp(-2\Delta_e^2 T_e). \tag{36}$$

Next, we bound each individual term on the right-hand side of Eq. (36). We begin with bounding the second term of Eq. (36)

$$\exp(-2\Delta_e^2 T_e) \geq \exp(-8\Delta_e^2 t_e^*) \tag{37}$$

$$= \exp\left(-\frac{8}{16}\log(1/\theta)\right) \tag{38}$$

$$= \theta^{1/2}, \tag{39}$$

where Eq. (117) follows from the assumption that event $\mathcal{S}$ occurred, which implies that event $\mathcal{A}$ occurred and therefore $T_e \leq 4t_e^*$; Eq. (118) follows from the definition of t_e^*.

Then, we bound the first term on the right-hand side of Eq. (36) as follows

$$\exp\left(\Delta_e(2K - 2w(e)T_e)\right) \geq \exp\left(-2\Delta_e\sqrt{t_e^*\log(1/\theta)}\right) \tag{40}$$

$$= \exp\left(-\frac{2}{\sqrt{4}}\log(1/\theta)\right) \tag{41}$$

$$= \theta^{1/2}, \tag{42}$$

where Eq. (120) 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. (121) follows from the definition of t_e^*.

Combining Eq. (119) and Eq. (122), we can bound $L_1(W)/L_0(W)$ for this case as follows

$$\frac{L_1(W)}{L_0(W)} \geq \theta. \tag{43}$$

(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

$$\frac{L_1(W)}{L_0(W)} = \prod_{i=1}^{T_e} \exp\left(\frac{-(X_i - w(e) + 2\Delta_e)^2 + (X_i - w(e))^2}{2}\right)$$

$$= \prod_{i=1}^{T_e} \exp\left(\Delta_e(2w(e) - 2X_i) - 2\Delta_e^2\right)$$

$$= \exp\left(\Delta_e(2w(e)T_e - 2K)\right)\exp(-2\Delta_e^2 T_e). \tag{44}$$

Notice that the right-hand side of Eq. (44) differs from Eq. (36) only in its first term. Now, we bound the first term as follows

$$\exp\left(\Delta_e(2K - 2w(e)T_e)\right) \geq \exp\left(-2\Delta_e\sqrt{t_e^*\log(1/\theta)}\right) \tag{45}$$

$$= \exp\left(-\frac{2}{4}\log(1/\theta)\right) \tag{46}$$

$$= \theta^{1/2}, \tag{47}$$

where the inequalities hold due to reasons similar to Case (1): Eq. (45) 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. (46) follows from the definition of t_e^*.

Combining Eq. (119) and Eq. (122), we can obtain the same bound of $L_1(W)/L_0(W)$ as in Eq. (43), 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. (43) holds, i.e. $\frac{L_1(W)}{L_0(W)} \geq \theta$. Hence, we have

$$\frac{L_1(W)}{L_0(W)} \geq \theta$$
$$= 4\delta. \tag{48}$$

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

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

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 δ, 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 at least δ under H_1. This contradicts to the assumption that algorithm $\mathbb{A}$ is a δ-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 $\delta > 0$, $\boldsymbol{w} \in \mathbb{R}^n$, diff-set $b = (b_+, b_-)$ and a δ-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 $b \in \mathcal{B}_{\mathsf{opt}}$, 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 $t_{b_-}^* = \frac{|b_-|^2}{16 w(b)^2} \log(1/\theta)$.

Consider the complete sequence of sampling process by algorithm $\mathbb{A}$. Formally, let $W = \{(\tilde{I}_1, \tilde{X}_1), \ldots, (\tilde{I}_T, \tilde{X}_T)\}$ be the sequence of all trials by algorithm $\mathbb{A}$, 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 $W = \{(I_1, X_1), \ldots, (I_{T_{b_-}}, X_{T_{b_-}})\}$ 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

$$\mathcal{B} = \{\mathsf{Out} = M_*\}. \tag{49}$$

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 $\mathrm{Pr}_0[\mathcal{A}_1] \geq 3/4$. This can be proved by Markov inequality as follows.

$$\mathrm{Pr}_0[T_{b_-} > 4t_{b_-}^*] \leq \frac{\mathbb{E}_0[T_{b_-}]}{4t_{b_-}^*}$$

$$= \frac{t_{b_-}^*}{4t_{b_-}^*} = \frac{1}{4}.$$

Next, we show that $\mathrm{Pr}_0[\mathcal{C}_1] \geq 3/4$. 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

$$\mathrm{Pr}_0\left[\max_{1 \leq t \leq 4t_{b_-}^*} \left| \sum_{i=1}^t X_i - \sum_{i=1}^t w(I_i) \right| \geq \sqrt{t_e^* \log(1/\theta)} \right] \leq \frac{\mathbb{E}_0\left[\left(\sum_{i=1}^{4t_{b_-}^*} X_i - \sum_{i=1}^{4t_{b_-}^*} w(I_i) \right)^2 \right]}{t_e^* \log(1/\theta)}$$

$$= \frac{4t_{b_-}^*}{t_{b_-}^* \log(1/\theta)}$$

$$< \frac{1}{4},$$

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 $\mathbb{A}$ is a δ-correct algorithm with $\delta < 1/4$. Therefore, it is easy to see that $\mathrm{Pr}_0[\mathcal{B}] \geq 3/4$. And by union bound, we have

$$\mathrm{Pr}_0[\mathcal{S}_1] \geq 1/4.$$

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

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

$$
\frac{L_1(W)}{L_0(W)} = \prod_{i=1}^{T_{b_-}} \exp\left(\frac{-\left(X_i - w(I_i) - \frac{2w(b)}{|b_-|}\right)^2 + (X_i - w(I_i))^2}{2} \right)
$$

$$
= \prod_{i=1}^{T_{b_-}} \exp\left(\frac{w(b)}{|b_-|}(2X_i - 2w(I_i)) - \frac{2w(b)^2}{|b_-|^2} \right)
$$

$$
= \exp\left(\frac{w(b)}{|b_-|}\left(\sum_{i=1}^{T_{b_-}} 2X_i - 2w(I_i) \right) - \frac{2w(b)^2}{|b_-|^2}T_{b_-} \right)
$$

$$
= \exp\left(\frac{w(b)}{|b_-|}\left(\sum_{i=1}^{T_{b_-}} 2X_i - 2w(I_i) \right) \right) \exp\left(-\frac{2w(b)^2}{|b_-|^2}T_{b_-} \right). \tag{50}
$$

Then, we bound each term on the right-hand side of Eq. (50). First, we bound the second term of Eq. (50).

$$
\exp\left(-\frac{2w(b)^2}{|b_-|^2}T_{b_-} \right) \geq \exp\left(-\frac{2w(b)^2}{|b_-|^2}4t_b^* \right) \tag{51}
$$

$$
= \exp\left(-\frac{8}{16}\log(1/\theta) \right) \tag{52}
$$

$$
= \theta^{1/2}, \tag{53}
$$

where Eq. (51) follows from the assumption that events $\mathcal{S}_1$ and $\mathcal{A}_1$ occurred and therefore $T_{b_-} \leq 4t_{b_-}^*$; Eq. (52) follows from the definition of $t_{b_-}^*$. Next, we bound the first term of Eq. (50) as follows

$$
\exp\left(\frac{w(b)}{|b_-|}\left(\sum_{i=1}^{T_{b_-}} 2X_i - 2w(I_i) \right) \right) \geq \exp\left(-\frac{2w(b)}{|b_-|}\sqrt{t_b^*\log(1/\theta)} \right) \tag{54}
$$

$$
= \exp\left(-\frac{2}{4}\log(1/\theta) \right) \tag{55}
$$

$$
= \theta^{1/2}, \tag{56}
$$

where Eq. (54) follows since event $\mathcal{S}_1$ and $\mathcal{C}_1$ occurred and therefore $|2K - 2p_eT_e| \leq \sqrt{t_e^*\log(1/\theta)}$; Eq. (55) follows from the definition of $t_{b_-}^*$.

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

$$
\frac{L_1(W)}{L_0(W)} \geq \theta = 4\delta. \tag{57}
$$

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

$$
\Pr_1[\mathcal{B}] \geq \Pr_1[\mathcal{S}_1] = \mathbb{E}_1[1_{\mathcal{S}_1}]
$$

$$
= \mathbb{E}_0\left[\frac{L_1(W)}{L_0(W)}1_{\mathcal{S}_1} \right]
$$

$$
\geq 4\delta\mathbb{E}_0[1_{\mathcal{S}_1}]
$$

$$
= 4\delta\Pr_0[\mathcal{S}_1] > \delta. \tag{58}
$$

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 δ. But M_* is not the optimal set under H_1. Hence this contradicts to the assumption that $\mathbb{A}$ is a δ-correct algorithm. Hence we have proved that

$$\mathbb{E}_0[T_{b_-}] \geq t_{b_-}^* = \frac{|b_-|^2}{16w(b)^2} \log(1/4\delta). \tag{59}$$

(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 $S_2 = \{(J_1, Y_1), \ldots, (J_{T_{b_+}}, Y_{T_{b_+}})\}$, 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. (49). 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

$$
\begin{aligned}
\frac{L_2(W)}{L_0(W)} &= \prod_{i=1}^{T_{b_+}} \exp\left(\frac{-\left(Y_i - w(J_i)\right) + \frac{2w(b)}{|b_-|}\right)^2 + (Y_i - w(J_i))^2}{2} \right) \\
&= \prod_{i=1}^{T_{b_+}} \exp\left(\frac{w(b)}{|b_+|}(2w(J_i) - 2Y_i) - \frac{2w(b)^2}{|b_+|^2} \right) \\
&= \exp\left(\frac{w(b)}{|b_+|} \left(\sum_{i=1}^{T_{b_+}} 2w(J_i) - 2Y_i \right) - \frac{2w(b)^2}{|b_+|^2} T_{b_+} \right) \\
&= \exp\left(\frac{w(b)}{|b_+|} \left(\sum_{i=1}^{T_{b_+}} 2w(J_i) - 2Y_i \right) \right) \exp\left(-\frac{2w(b)^2}{|b_+|^2} T_{b_+} \right) \\
&\geq \theta \\
&= 4\delta,
\end{aligned}
\tag{60}
$$

where Eq. (60) 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. (58), we see that

$$\Pr_2[\mathcal{B}] \geq \Pr_2[\mathcal{S}_2] = \mathbb{E}_2[1_{S_2}] = \mathbb{E}_0\left[\frac{L_2(W)}{L_0(W)} 1_{S_2}\right] \geq 4\delta \mathbb{E}_0[1_{S_2}] > \delta,$$

where 1_{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 δ, which is not the optimal set under H_2. This

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

$$\mathbb{E}_0[T_{b_+}] \geq t^*_{b_+} = \frac{|b_+|^2}{16w(b)^2} \log(1/4\delta). \tag{61}$$

(End of Part (2))
Finally, we combine the results from both parts, i.e. Eq. (59) and Eq. (61). We obtain

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

$\square$

7   Proof of Extension Results

7.1   Fixed Budget Setting

In this part, we analyze the probability of error of the modified `CGapExp` algorithm in the fixed budget setting and prove Theorem 4. First, we prove a lemma which characterizes the confidence intervals constructed in Theorem 4.

Lemma 12. *Fix parameter $\alpha > 0$ and the number of rounds $T > 0$. Assume that the reward distribution φ_e is a R-sub-Gaussian distribution for all $e \in [n]$. Let the confidence radius $\mathrm{rad}_t(e)$ of arm $e \in [n]$ and round $t > 0$ be $\mathrm{rad}_t(e) = R\sqrt{\frac{\alpha}{T_e(t)}}$. Then, we have*

$$\Pr\left[\bigcap_{t=1}^{T} \xi_t\right] \geq 1 - 2nT \exp\left(-2\alpha\right).$$

Proof. For any $t > 0$ and $e \in [n]$, using Hoeffding's inequality, we have

$$\Pr\left[\left|\bar{w}_t(e) - w(e)\right| \geq \mathrm{rad}_t(e)\right] \leq 2\exp(-2\alpha).$$

By a union bound over all arms $e \in [n]$, we see that $\Pr[\xi_t] \geq 1 - 2n\exp(-2\alpha)$. The lemma follows immediately by using union bound again over all round $t \in [T]$. $\square$

Then, Theorem 4 can be obtained from the key lemmas (Lemma 9 and Lemma 10) and Lemma 12 as follows.

Proof. Define random event $\xi = \bigcap_{t=1}^{T} \xi_t$. By Lemma 12, we see that $\Pr[\xi] \geq 1 - 2nT \exp(-2\alpha)$. In the rest of the proof, we assume that ξ happens.
We claim that Algorithm 1 stops before round T. If the claim is true, then there exists a round $t^* < T$ such that $\tilde{M}_{t^*} = M_{t^*}$ and $\mathsf{Out} = M_{t^*}$. By assumption on ξ and Lemma 9, we know that $M_{t^*} = M_*$. Therefore the theorem follows immediately from this claim and the bound of $\Pr[\xi]$.
Next, we show that this claim is true. For any round $t \in [T]$ and any arm $e \in [n]$, by Lemma 10, we see that if $\mathrm{rad}_t(e) < \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}$, then arm e will no longer be pulled. Also notice that, by construction in the fixed budget setting, the radius $\mathrm{rad}_t(e)$ is monotone decreasing with respect to $T_e(t)$. Therefore, plugging the definition of $\mathrm{rad}_t(e)$, we have

$$T_e \leq \frac{9R^2 \,\mathrm{width}(\mathcal{B})^2}{\Delta_e^2} \cdot \alpha.$$

Summing up T_e for all $e \in [n]$, we have

$$t^* = \sum_{e \in [n]} T_e \leq \alpha \cdot 9R^2 \operatorname{width}(\mathcal{B})^2 \left(\sum_{e \in [n]} \Delta_e^{-2} \right) < T,$$

where we have used the assumption that $\alpha < 9T \cdot \left(R^2 \operatorname{width}(\mathcal{B})^2 \left(\sum_{e \in [n]} \Delta_e^{-2} \right) \right)^{-1}$.

□

7.2 PAC Learning

First, we prove a (ϵ, δ)-PAC counterpart of Lemma 9.

Lemma 13. *If* `CGapExpPAC` *stops on round t and suppose that event ξ_t occurs. Then, we have $w(M_*) - w(\mathsf{Out}) \leq \epsilon$.*

Proof. By definition, we know that $\mathsf{Out} = M_t$. Notice that the stopping condition of `CGapExpPAC` ensures that $\tilde{w}_t(\tilde{M}_t) \leq \epsilon$. Therefore, we have

$$\epsilon \geq \tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t)$$
$$\geq \tilde{w}_t(M_*) - \tilde{w}_t(M_t) \tag{62}$$
$$= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_*} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_{M_*} - \boldsymbol{\chi}_{M_t}| \rangle \tag{63}$$
$$\geq \langle \boldsymbol{w}, \boldsymbol{\chi}_{M_*} - \boldsymbol{\chi}_{M_t} \rangle \tag{64}$$
$$= w(M_*) - w(M_t),$$

where Eq. (62) follows from the definition of $\tilde{M}_t \triangleq \arg\max_{M \in \mathcal{M}} \tilde{w}_t(M)$; Eq. (63) follows from Lemma 6; Eq. (64) follows from the assumption that ξ_t occurs and Lemma 7. □

The next lemma generalizes Lemma 10 and shows that, with high probability, each arm $e \in [n]$ will not be played if .

Lemma 14. *For any arm $e \in [n]$ and any round $t > n$ after initialization, if $\operatorname{rad}_t(e) \leq \max \left\{ \frac{\Delta_e}{3 \operatorname{width}(\mathcal{B})}, \frac{\epsilon}{2K} \right\}$, then arm e will not be played on round t, i.e. $p_t \neq e$.*

Proof. If $\operatorname{rad}_t(e) \leq \frac{\Delta_e}{3 \operatorname{width}(\mathcal{B})}$, then we can apply Lemma 10 which immediately gives that $p_t \neq e$. Hence, we only need to prove the case that $\frac{\Delta_e}{3 \operatorname{width}(\mathcal{B})} \leq \operatorname{rad}_t(e) \leq \frac{\epsilon}{2K}$.

Now suppose that $p_t = e$. By the choice of p_t, we know that for each $i \in (M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)$, we have $\operatorname{rad}_t(i) \leq \operatorname{rad}_t(e) \leq \frac{\epsilon}{2K}$. By summing up this inequality for all $i \in (M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)$, we have

$$\epsilon \geq \sum_{i \in (M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)} \operatorname{rad}_t(i) \tag{65}$$
$$= \langle \mathbf{rad}_t, |\chi_{M_t} - \chi_{\tilde{M}_t}| \rangle, \tag{66}$$

where Eq. (74) follows from the fact that $|(M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)| \leq |M_t| + |\tilde{M}_t| \leq 2K$; and Eq. (75) uses the fact that $\boldsymbol{\chi}_{(M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)} = |\boldsymbol{\chi}_{M_t} - \boldsymbol{\chi}_{\tilde{M}_t}|$.

Then, using the fact that $\bar{w}_t(M_t) \geq \bar{w}_t(\tilde{M}_t)$, we have

$$\tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) = \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t}| \rangle \tag{67}$$
$$\leq \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t} \rangle + \epsilon \tag{68}$$
$$= \bar{w}_t(\tilde{M}_t) - \bar{w}_t(M_t) + \epsilon$$

$$\leq \epsilon, \tag{69}$$

where Eq. (76) follows from Lemma 6; Eq. (68) uses Eq. (75); and Eq. (69) follows from $\bar{w}_t(M_t) \geq \bar{w}_t(\tilde{M}_t)$. Therefore, we see that $\tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) \leq \epsilon$. By the stopping condition of `CGapExpPAC`, the algorithm must terminate on round t. This contradicts to the assumption that $p_t = e$. $\qquad\square$

Proof. Similar to the proof of Theorem 1, we appeal to Lemma 8, which shows that the event $\xi \triangleq \bigcap_{t=1}^{\infty} \xi_t$ occurs with probability at least $1 - \delta$. And we shall assume that ξ occurs in the rest of the proof.

By the assumption of ξ and Lemma 25, we know that $\mathsf{Out} = M_*$. Therefore, we only remain to bound the number of samples T.

Consider an arbitrary arm $e \in [n]$. Let T_e denote the total number of pull of arm $e \in [n]$. Let t_e be the last round which arm e is pulled, i.e. $p_{t_e} = e$. Hence $T_e(t_e) = T_e - 1$. By Lemma 14, we see that $\mathrm{rad}_{t_e}(e) \geq \min\{\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}, \frac{\epsilon}{2K}\}$. Then, by the construction of $\mathrm{rad}_{t_e}(e)$, we have

$$\min\left\{\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}, \frac{\epsilon}{2K}\right\} \leq R\sqrt{\frac{2\log\left(4nt_e^2/\delta\right)}{T_e - 1}} \leq R\sqrt{\frac{2\log\left(4nT^2/\delta\right)}{T_e - 1}}. \tag{70}$$

Solving Eq. (70) for T_e, we obtain

$$T_e \leq R^2 \min\left\{\frac{18\,\mathrm{width}(\mathcal{B})^2}{\Delta_e^2}, \frac{16K^2}{\epsilon^2}\right\} \log(4nT^2/\delta) + 1. \tag{71}$$

Notice that $T = \sum_{i \in [n]} T_i$. Hence the theorem follows by summing up Eq. (71) for all $e \in [n]$ and solving for T. $\qquad\square$

8 Technical Lemmas

Lemma 15 (Basis exchange property). *AA*

Lemma 16 (Hoeffding's inequality). *Let $X_1, \ldots, X_n$ be n independent R-sub-Gaussian random variables. Let $\bar{X} = \frac{1}{n}\sum X_i$ be the average of these random variables. Then, we have*

$$\Pr\left[\left|\bar{X} - \mathbb{E}[\bar{X}]\right| \geq t\right] \leq 2\exp\left(-\frac{2nt^2}{R^2}\right).$$

9 Trash

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

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

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

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 $\mathrm{rad}_t(e) < \frac{1}{3}\Delta_e$.*

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

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

Theorem 6. *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 7. *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 $\mathrm{Oracle}(w)$ be a maximization oracle associated with $\mathcal{M}$ such that $\mathrm{Oracle}(w) = \arg\max_{M \in \mathcal{M}} w(M)$, where $w : 2^{[n]} \to R$ is a weight function.*
Then, with probability at least $1 - \delta$, the output Out of Algorithm 1 satisfies $w(M_*) - w(\mathsf{Out}) \leq \epsilon$, 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

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

Lemma 17. *For any arm $e \in [n]$ and any round $t > n$ after initialization, if $\mathrm{rad}_t(e) \leq \max\left\{\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}, \frac{\epsilon}{n}\right\}$, then arm e will not be played on round t, i.e. $p_t \neq e$.*

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

$$\mathrm{rad}_t(D_t) \leq \epsilon. \tag{72}$$

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

$$\bar{w}_t(D_t) = \bar{w}_t(M_t \oplus D_t) - \bar{w}_t(M_t) \leq 0. \tag{73}$$

Using the above inequalities, we have.

$$w_t^+(D_t) = \bar{w}_t(D_t) + \mathrm{rad}_t(D_t) \tag{74}$$
$$\leq \bar{w}_t(D_t) + \epsilon \tag{75}$$
$$\leq \epsilon, \tag{76}$$

where Eq. (74) follows from the definition of $w_t^+(\cdot)$; Eq. (75) follows from Eq. (72); Eq. (76) holds since Eq. (73). $\qquad\square$

10 Preliminaries

10.1 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). $\mathcal{M}_{\mathsf{TOP}m}(n) = \{M \subseteq [n] \mid |M| = m\}$. *This corresponds to finding the top m arms from $[n]$.*

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

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

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

23

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 $\mathcal{M}_{\mathsf{MATCH}}(n, G)$ contains all subsets $M \subseteq [n]$ such that M corresponds to a perfect matching in G.*

10.2 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 $C_+ = M_1 \backslash M_2$ and $C_- = M_2 \backslash M_1$.*

Definition 6. *Denote $\mathsf{diff}[n]$ 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 $C \backslash D \triangleq (C_+ \backslash D_+, C_- \backslash D_-)$.*
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 $\neg A = (A_-, A_+)$.*

10.2.1 diff-set operations

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

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 $A \oplus B = ((A_+ \cup B_+) \backslash (A_- \cup B_-), (A_- \cup B_-) \backslash (A_+ \cup B_+))$.*

Lemma 18. *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 $C_+ = (A_+ \cup B_+) \backslash (A_- \cup B_-)$ and $C_- = (A_- \cup B_-) \backslash (A_+ \cup B_+)$. We only need to show that $C_+ \cap C_- = \emptyset$.

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

$\square$

Lemma 19. *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 $B_+ \cap (M \backslash A_- \cup A_+) = \emptyset$. Therefore, we have

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

We see that $B_+ \cap A_+ = \emptyset$.
On the other hand, we have $B_- \subseteq (M \backslash A_- \cup A_+)$, therefore

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

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

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

Similarly, we have

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

$\square$

10.2.2 Diff-set class

Definition 13 (Decomposition of diff-set). *Given $\mathcal{B} \subseteq \mathsf{diff}[n]$ and $D \in \mathsf{diff}[n]$, a decomposition of D on $\mathcal{B}$ is a set $\{b_1, \ldots, b_k\} \subseteq \mathcal{B}$ 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. *$D = b_1 \oplus b_2 \oplus \ldots b_k$.*

Lemma 20. *Given $\mathcal{B} \subseteq \mathsf{diff}[n]$ and $D \in \mathsf{diff}[n]$. Let $\{b_1, \ldots, b_k\} \subseteq \mathcal{B}$ 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 $D_+ = b_1^+ \cup \ldots \cup b_k^+$ and $D_- = b_1^- \cup \ldots \cup b_k^-$.*

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 $D_i = b_1 \oplus \ldots \oplus b_i$ and write $D_i = (D_i^+, D_i^-)$. We show that $D_i^+ = \bigcup_{j=1}^i b_i^+$ and $D_{i-} = \bigcup_{j=1}^i b_i^-$ 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 $D_{i+1}^+ = (D_i^+ \cup b_{i+1}^+) \backslash (D_i^- \cup b_{i+1}^-)$. 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 $b_i^- \subseteq D_- \subseteq M$.

$\square$

Definition 14 (diff-set class). *Given $\mathcal{M} \subseteq 2^{[n]}$. $\mathcal{B} \subseteq \mathrm{diff}[n]$ 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*

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

Example 4 (diff-set class for Explore-m). *One diff-set class $\mathcal{B}$ for $\mathcal{M}_{\mathsf{TOP}m}(n)$ 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 $\mathrm{width}(\mathcal{B}) = 2$.

Example 5 (diff-set class for Explore-m-badit). *Let $n = mk$. One diff-set class $\mathcal{B}$ for $\mathcal{M}_{\mathsf{BANDIT}m}(n)$ 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 $\mathrm{width}(\mathcal{B}) = 2$.

Example 6 (diff-set class for Perfect Matching). *One diff-set class $\mathcal{B}$ for $\mathcal{M}_{\mathsf{MATCH}}(n, G)$ 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 $\mathrm{width}(\mathcal{B}) \leq n$.

10.3 Weights and confidence bounds

Definition 16 (Weight functions). *Define function $w : [n] \to \mathbb{R}^+$ 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 $b = (b_+, b_-) \in \mathrm{diff}[n]$, we denote $w(b) = \sum_{e \in b_+} w(e) - \sum_{e \in b_-} w(e)$.*

Lemma 21. *Let $c \in \mathrm{diff}[n], d \in \mathrm{diff}[n]$. Let w be a weight function. Then,*

$$w(c \cup d) = w(c) + w(d) - w(c \cap d). \tag{77}$$

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

$$w(c \cup d) = w(c_+ \cup d_+) - w(c_- \cup d_-) \tag{78}$$

$$= w(c_+) + w(d_+) - w(c_+ \cap d_+) - w(c_-) - w(d_-) + w(c_- \cap d_-) \tag{79}$$

26

$$= w(c) + w(d) - (w(c_+ \cap d_+) - w(c_- \cap d_-)) \tag{80}$$

$$= w(c) + w(d) - w(c \cap d). \tag{81}$$

$$\square$$

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 $n_t : [n] \to \mathbb{N}$, such that $n_t(e)$ equals to number of plays of base arm e up to round t.*

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

$$\mathrm{rad}_t(e) = c_{\mathrm{rad}} \log\left(\frac{c_\delta n t^2}{\delta}\right) \frac{1}{\sqrt{n_t(e)}}, \tag{82}$$

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

1. *For all $M \subseteq [n]$, $\mathrm{rad}_t(M) \triangleq \sum_{e \in M} \mathrm{rad}_t(e)$.*

2. *For all $b = (b_+, b_-) \in \mathsf{diff}[n]$, $\mathrm{rad}_t(b) \triangleq \mathrm{rad}_t(b_+) + \mathrm{rad}_t(b_-)$.*

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

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

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

1. *For all $M \subseteq [n]$, $w_t^+(M) \triangleq \bar{w}_t(M) + \mathrm{rad}_t(M)$.*

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

Lemma 22. *Define random event*

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

Then, there exist constants c_{rad} and c_δ,

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

Proof. Hoeffding inequality and union bound. $\square$

Corollary 2.

$$\xi \implies \forall t, \forall e \in [n] \; w_t^+(e) \geq w(e).$$

$$\xi \implies \forall t, \forall M \subseteq [n], \; w_t^+(M) \geq w(M).$$

$$\xi \implies \forall t, \forall b \in \mathsf{diff}[n] \; w_t^+(b) \geq w(b).$$

10.4   Properties of rad_t

Lemma 23. *Let $c \in \mathsf{diff}[n], d \in \mathsf{diff}[n]$. Then*

$$\mathrm{rad}_t(c \backslash d) = \mathrm{rad}_t(c) - \mathrm{rad}_t(c \cap d). \tag{83}$$

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

$$\mathrm{rad}_t(c \backslash d) = \mathrm{rad}_t(c_+ \backslash d_+) + \mathrm{rad}_t(c_- \backslash d_-)$$

$$= \mathrm{rad}_t(c_+) - \mathrm{rad}_t(c_+ \cap d_+) + \mathrm{rad}_t(c_-) - \mathrm{rad}_t(c_- \cap d_-)$$
$$= \mathrm{rad}_t(c) - \mathrm{rad}_t(c \cap d).$$

$\square$

Lemma 24. *Let $C = (C_+, C_-)$ and $D = (D_+, D_-)$ be two diff-sets. If $D \prec C$, then*

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

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

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

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

$$\mathrm{rad}_t((C_+ \cup D_+)\backslash(C_- \cup D_-)) = \mathrm{rad}_t(C_+ \cup D_+) - \mathrm{rad}_t((C_+ \cup D_+) \cap (C_- \cup D_-)) \tag{84}$$
$$= \mathrm{rad}_t(C_+) + \mathrm{rad}_t(D_+) - \mathrm{rad}_t((C_+ \cup D_+) \cap (C_- \cup D_-)), \tag{85}$$

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

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

Combine both equalities, we have

$$\mathrm{rad}_t(C \oplus D) = \mathrm{rad}_t((C_+ \cup D_+)\backslash(C_- \cup D_)) + \mathrm{rad}_t((C_- \cup D_-)\backslash(C_- \cap D_+)) \tag{86}$$
$$= \mathrm{rad}_t(C_+) + \mathrm{rad}_t(D_+) + \mathrm{rad}_t(C_-) + \mathrm{rad}_t(D_-) - 2\,\mathrm{rad}_t((C_+ \cup D_+) \cap (C_- \cup D_-)) \tag{87}$$
$$= \mathrm{rad}_t(C) + \mathrm{rad}_t(D) - 2\,\mathrm{rad}_t((C_+ \cup D_+) \cap (C_- \cup D_-)). \tag{88}$$

$\square$

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 $c \in \mathcal{B}_{\mathsf{opt}}$. Then $w(c) > 0$. We have $\mathrm{rad}_t(e) \leq \frac{1}{3}\Delta_e \leq \frac{1}{3K}w(c)$.
In addition, by the choice of p_t, we have $\forall g \in c_t, g \neq e$, $\mathrm{rad}_t(g) \leq \mathrm{rad}_t(e) \leq \frac{1}{3K}w(c)$. Hence, $\mathrm{rad}_t(c) = \sum_{g \in c_t} \mathrm{rad}_t(g) \leq \frac{|c_t|}{3K}w(c) \leq \frac{1}{3}w(c)$.
Hence, $w_t(c) \geq w(c) - \mathrm{rad}_t(c) \geq \frac{2}{3}w(c) > 0$. This means that $w_t(M_t \oplus c) = w_t(M_t) + w_t(c) > w_t(M_t)$. Therefore, $M_t \neq \max_{M \in \mathcal{M}} \bar{w}_t(M)$. This contradicts to the definition of M_t.
Case (2). Suppose that $c_t \notin \mathcal{B}_{\mathsf{opt}}$. 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 $b, b_1, \ldots, b_l$, 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 $b \in \mathcal{B}_{\mathsf{opt}}$.
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 19, 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

$$\tilde{D}_+ = (D_+ \cup c_-)\backslash(D_- \cup c_+)$$
$$= (D_+ \cup c_-\backslash D_-) \cap (D_+ \cup c_-\backslash c_+)$$

$$= D_+ \cap (D_+ \backslash c_-)$$
$$= D_+ \backslash c_+.$$

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

$$\tilde{D}_+ \subseteq D_+ \quad \text{and} \quad \tilde{D}_- \subseteq D_-. \tag{89}$$

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

$$
\begin{aligned}
\mathrm{rad}_t(c) &= \sum_{g \in c} \mathrm{rad}_t(g) \\
&\le \frac{|c|}{3K} w(b) \\
&\le \frac{1}{3} w(b). \tag{90}
\end{aligned}
$$

Now, we show that $\mathrm{rad}_t(\tilde{D}_+ \cap b_-) + \mathrm{rad}_t(\tilde{D}_- \cap b_+) + \le \frac{1}{3} w(b)$. Since Eq. (89), we have $\forall g \in (\tilde{D}_+ \cap b_-) \cup (\tilde{D}_- \cap b_+), g \ne e$, $\mathrm{rad}_t(g) \le \mathrm{rad}_t(e) \le \frac{1}{3K} w(b)$. Note that $|\tilde{D}_+ \cap b_-| + |\tilde{D}_- \cap b_+| \le |b_+| + |b_-| \le K$. Hence,

$$
\begin{aligned}
\mathrm{rad}_t(\tilde{D}_+ \cap b_-) + \mathrm{rad}_t(\tilde{D}_- \cap b_+) &= \sum_{g \in (\tilde{D}_+ \cap b_-) \cup (\tilde{D}_- \cap b_+)} \mathrm{rad}_t(g) \\
&\le \frac{K}{3K} w(b) \\
&\le \frac{1}{3} w(b). \tag{91}
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
\mathrm{rad}_t(D') - \mathrm{rad}_t(D) &= \mathrm{rad}_t(\tilde{D} \oplus b) - \mathrm{rad}_t(D) \tag{92} \\
&= \mathrm{rad}_t(\tilde{D}) + \mathrm{rad}_t(b) - 2\,\mathrm{rad}_t(\tilde{D}_+ \cap b_-) - 2\,\mathrm{rad}_t(\tilde{D}_- \cap b_+) - \mathrm{rad}_t(D) \tag{93} \\
&= \mathrm{rad}_t(D \ominus c) + \mathrm{rad}_t(b) - 2\,\mathrm{rad}_t(\tilde{D}_+ \cap b_-) - 2\,\mathrm{rad}_t(\tilde{D}_- \cap b_+) - \mathrm{rad}_t(D) \tag{94} \\
&= \mathrm{rad}_t(D) + \mathrm{rad}_t(c) + \mathrm{rad}_t(b) - 2\,\mathrm{rad}_t(D_+ \cap c_+) - 2\,\mathrm{rad}_t(D_- \cap c_-) \\
&\quad - 2\,\mathrm{rad}_t(\tilde{D}_+ \cap b_-) - 2\,\mathrm{rad}_t(\tilde{D}_- \cap b_+) - \mathrm{rad}_t(D) \tag{95} \\
&= \mathrm{rad}_t(D) + \mathrm{rad}_t(c) + \mathrm{rad}_t(b) - 2\,\mathrm{rad}_t(c_+) - 2\,\mathrm{rad}_t(c_-) \\
&\quad - 2\,\mathrm{rad}_t(\tilde{D}_+ \cap b_-) - 2\,\mathrm{rad}_t(\tilde{D}_- \cap b_+) - \mathrm{rad}_t(D) \tag{96} \\
&= \mathrm{rad}_t(b) - \mathrm{rad}_t(c) - 2\,\mathrm{rad}_t(\tilde{D}_+ \cap b_-) - 2\,\mathrm{rad}_t(\tilde{D}_- \cap b_+), \tag{97}
\end{aligned}
$$

where Eq. (93) and Eq. (95) follow from Lemma 24, and Eq. (96) follows from Eq. (**??**).
By the definition of D, we have that $w_t^+(D) \ge w_t^+(D')$. This means that

$$
\begin{aligned}
\bar{w}_t(D) + \mathrm{rad}_t(D) &\ge \bar{w}_t(D') + \mathrm{rad}_t(D') \tag{98} \\
&= \bar{w}_t(D) - \bar{w}_t(c) + \bar{w}_t(b) + \mathrm{rad}_t(D'). \tag{99}
\end{aligned}
$$

By regrouping the above inequality, we have

$$
\begin{aligned}
\bar{w}_t(c) &\ge \bar{w}_t(b) + \mathrm{rad}_t(D') - \mathrm{rad}_t(D) \tag{100} \\
&= \bar{w}_t(b) + \mathrm{rad}_t(b) - \mathrm{rad}_t(c) - 2\,\mathrm{rad}_t(\tilde{D}_+ \cap b_-) - 2\,\mathrm{rad}_t(\tilde{D}_- \cap b_+) \tag{101} \\
&\ge w(b) - \mathrm{rad}_t(c) - 2\,\mathrm{rad}_t(\tilde{D}_+ \cap b_-) - 2\,\mathrm{rad}_t(\tilde{D}_- \cap b_+) \tag{102}
\end{aligned}
$$

$$> w(b) - \frac{1}{3}w(b) - \frac{2}{3}w(b) \tag{103}$$

$$= 0, \tag{104}$$

where Eq. (103) follows from Eq. (90) and Eq. (91).

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 $b, b_1, \ldots, b_l$. Assume wlog that $e \in b$. We write that $b = (b_+, b_-)$. Note that $b \in \mathcal{B}_{\mathsf{opt}}$ and hence $w(b) > 0$.

Define $D' = (M_t \oplus D \oplus b) \ominus M_t$. By Lemma 19, we know that $D' = D \oplus b$.

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

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

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

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

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

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

We also note that

$$w(D'\backslash D) - w(D\backslash D') = w(D'\backslash D) + w(D' \cap D) - w(D \cap D') - w(D\backslash D') \tag{105}$$

$$= w(D') - w(D) \tag{106}$$

$$= w(b), \tag{107}$$

where we have repeatedly applied Lemma 21.

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

$$w_t^+(D') - w_t^+(D) = \bar{w}_t(D') - \bar{w}_t(D) + \mathrm{rad}_t(D') - \mathrm{rad}_t(D) \tag{108}$$

$$= \bar{w}_t(D'\backslash D) - \bar{w}_t(D\backslash D') + \mathrm{rad}_t(D'\backslash D) - \mathrm{rad}_t(D\backslash D') \tag{109}$$

$$\geq w(D'\backslash D) - w(D\backslash D') - 2\,\mathrm{rad}_t(D\backslash D') \tag{110}$$

$$= w(b) - 2\,\mathrm{rad}_t(D\backslash D') \tag{111}$$

$$> w(b) - \frac{2}{3}w(b) \tag{112}$$

$$= \frac{1}{3}w(b) > 0, \tag{113}$$

where Eq. (109) follows from Lemma 23 and Eq. (110) follows from the fact that $\bar{w}_t(D'\backslash D) + \mathrm{rad}_t(D'\backslash D) \geq w(D'\backslash D)$ and that $\bar{w}_t(D\backslash D') + \mathrm{rad}_t(D\backslash D') \geq w(D\backslash D')$, under the random event ξ.

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

Theorem 8. *Given a vector $\{w(1), \ldots, w(n)\}$, a budget $T > 0$ and a collection of feasible sets $\mathcal{M} \subseteq 2^{[n]}$. Let $\mathbb{A}$ be an arbitrary algorithm for $\mathcal{M}$-**ExpCMAB** problem which uses at most T samples. There exists a vector $\{\tilde{w}(1), \ldots, \tilde{w}(n)\}$ 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 μ and variance σ^2. The probability of error of $\mathbb{A}$ on this bandit problem satisfies*

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

where Out *is the output of $\mathbb{A}$ and $M_* = \arg\max_{M \in \mathcal{M}} w(M)$ is the optimal set. In addition, vector $\{\tilde{w}(1), \ldots, \tilde{w}(n)\}$ differs from vector $\{w(1), \ldots, w(n)\}$ 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 $\sigma(1), \ldots, \sigma(n)$ be a permutation of $1, 2, \ldots, n$ such that $\Delta_{\sigma(1)} \leq \Delta_{\sigma(2)} \ldots \leq \Delta_{\sigma(n)}$. 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

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

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

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

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

$$\Pr_0[\mathcal{C} \cap \{X = x, Y = y, Z = z\}] \geq \frac{1}{4L(L+1)}. \tag{114}$$

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. (114) implies that

$$\Pr_0[\{X = x, Y = y, Z = z\}] \geq \Pr_0[\mathcal{C} \cap \{X = x, Y = y, Z = z\}]$$

$$\geq \frac{1}{4L(L+1)} \geq G.$$

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

$$T_Z \leq \frac{T}{L}. \tag{115}$$

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

$$\begin{aligned}
\frac{L_1(W)}{L_0(W)} &= \prod_{i=1}^{T_Z} \exp\left(\frac{-(X_i - w(x) - \Delta_L)^2 + (X_i - w(x))^2}{2}\right) \\
&= \prod_{i=1}^{T_Z} \exp\left(\Delta_L(X_i - w(x)) - \Delta_L^2\right) \\
&= \exp\left(\Delta_L(K - T_Z w(x)) - \Delta_L^2 T_Z\right) \\
&= \exp\left(\Delta_L(K - T_Z w(x))\right) \exp(-\Delta_L^2 T_Z).
\end{aligned} \tag{116}$$

Then we analyze the right-hand side of Eq. (116) as follows

$$\exp(-\Delta_Z^2 T_Z) \geq \exp(-\Delta_Z^2 T/L) \tag{117}$$

$$= \exp\left(-\frac{8}{16}\log(1/\theta)\right) \tag{118}$$

$$= \theta^{1/2}, \tag{119}$$

where Eq. (117) follows from the assumption that event $\mathcal{S}$ occurred, which implies that event $\mathcal{A}$ occurred and therefore $T_e \leq 4t_e^*$; Eq. (118) follows from the definition of t_e^*.

Then, we bound the first term on the right-hand side of Eq. (36) as follows

$$\exp\left(\Delta_e(2K - 2p_e T_e)\right) \geq \exp\left(-2\Delta_e \sqrt{t_e^* \log(1/\theta)}\right) \tag{120}$$

$$= \exp\left(-\frac{2}{\sqrt{4}}\log(1/\theta)\right) \tag{121}$$

$$= \theta^{1/2}, \tag{122}$$

where Eq. (120) 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. (121) 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 $y \notin \mathsf{Out}$ and therefore $\mathsf{Out} \neq M_*$. On the other hand, using Eq. (114), 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 $\Pr_0[\mathsf{Out} \neq M_*] \geq \frac{1}{2(L+1)^2} \geq A$.

$\square$

Lemma 25. *If Algorithm 1 stops, then* $w(M_*) - w(\mathsf{Out}) \leq \epsilon$.

Proof. Suppose that $\mathsf{Out} \neq M_*$. Suppose that the algorithm stops on round T, we know that $\mathsf{Out} = M_T$. 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

$$w_T^+(D_T) = \max_{C:C \prec M_T} w_T^+(C) \geq w_T^+(D). \tag{123}$$

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

$$\epsilon \geq \tilde{w}_T(\tilde{M}_T) - \tilde{w}_T(M_T)$$
$$= w_T^+(D_T) \geq w_T^+(D) \tag{124}$$
$$\geq w(D) = w(M_*) - w(M_T), \tag{125}$$

where Eq. (124) follows from Eq. (123); Eq. (125) follows from the assumption that event ξ occurred.

$\square$

References