# Pure Exploration of Combinatorial Bandits

**Anonymous Author(s)**
Affiliation
Address
email

**Abstract**

The abstract paragraph should be indented 1/2 inch (3 picas) on both left and right-hand margins. Use 10 point type, with a vertical spacing of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

## 1 Introduction

### 1.1 Related Work

**Notations.**

## 2 Pure Exploration of Combinatorial Bandits

**ExpCMAB: problem formulation.** Suppose that there are $n$ arms and the arms are numbered $1, 2, \ldots, n$. Each arm $e \in [n]$ is associated with a reward distribution $\varphi_e$ and define $w(e) = \mathbb{E}_{X \sim \varphi_e}[X]$ be the expected reward. Let $\boldsymbol{w} = \big(w(1), \ldots, w(n)\big)^T$ denote the vector of expected rewards.

Let $\mathcal{M} \subseteq 2^{[n]}$ be the family of all feasible solutions to a combinatorial problem. A learner wants to find the optimal solution of $\mathcal{M}$ which maximizes the expected reward $M_* = \arg\max_{M \in \mathcal{M}} w(M)$ by playing the following game. At the beginning of the game, the reward distributions $\{\varphi_e\}_{e \in [n]}$ are unknown to the learner. Then, the game is played for multiple rounds; on each round $t$, the learner pulls an arm $p_t \in [n]$ and observes a reward sampled from the associated reward distribution $\varphi_{p_t}$. The game continues until certain stopping condition is satisfied. After the game finishes, the learner need to output a solution $\mathsf{Out} \in \mathcal{M}$.

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 while uses a small number of pulls. Specifically, given a confidence parameter $\delta$, 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. 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.

**Examples of combinatorial problems.** The formulation of the ExpCMAB problem covers many online learning tasks. We consider the following problems as examples.

- MULTI.
- MATROID.
- MATCH.

- PATH.

We assume that all reward distributions have $R$-sub-Gaussian tails. Formally, for all $t \in \mathbb{R}$, we assume that $\mathbb{E}_{X\sim\varphi_e}\left[\exp(tX - tw(e))\right] \le \exp(R^2t^2/2)$. It is well known that all distributions that are supported on $[0, R]$ satisfy this property [].

# 3 Algorithm, Exchange Class and Main Result

In this section, we present CGapExp, a learning algorithm for the ExpCMAB problem in the fixed confidence setting, and analyze its sample complexity. En route to our main result, we propose the notions of exchange class and width of combinatorial problems, which characterize the exchange properties of combinatorial structures.

The CGapExp algorithm can be extended to the fixed budget and PAC learning settings. We will discuss these extensions in Section 5.

**Oracle.** We allow the CGapExp algorithm to access a *maximization oracle*. A maximization oracle takes a weight vector $\boldsymbol{v} \in \mathbb{R}^n$ as input and computes an optimal solution with respect to the weight vector $\boldsymbol{v}$. Formally, we call a function Oracle: $\mathbb{R}^n \to \mathcal{M}$ a maximization oracle if, for all $\boldsymbol{v} \in \mathbb{R}^n$, we have $\mathrm{Oracle}(\boldsymbol{v}) \in \arg\max_{M\in\mathcal{M}} v(M)$. It is clear that a very broad class of combinatorial problems admit such maximization oracles. Besides the access to the oracle, CGapExp does not need *any* additional knowledge of the combinatorial problem $\mathcal{M}$.

**Algorithm.** The CGapExp algorithm maintains empirical mean $\bar{w}_t(e)$ and confidence radius $\mathrm{rad}_t(e)$ for each arm $e \in [n]$ and each round $t$. The construction of confidence radius ensures that $|w(e) - \bar{w}_t(e)| \le \mathrm{rad}_t(e)$ holds with high probability for each arm $e \in [n]$ and each round $t > 0$. CGapExp begins with an initialization phase in which each arm is pulled once. Then, at round $t \ge n$, CGapExp uses the following procedure to choose an arm to play. First, CGapExp calls the oracle which computes the solution $M_t = \mathrm{Oracle}(\bar{\boldsymbol{w}}_t)$. The solution $M_t$ is the "best" solution with respect 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 $\tilde{w}_t(e) = \bar{w}_t(e) - \mathrm{rad}_t(e)$; and for each arm $e \notin M_t$, $\tilde{w}_t(e)$ equals to the upper confidence bound $\tilde{w}_t(e) = \bar{w}_t(e) + \mathrm{rad}_t(e)$. Intuitively, the adjusted expectation vector $\tilde{\boldsymbol{w}}_t$ penalizes arms belonging to the current solution $M_t$ and encourages exploring arms out of $M_t$. CGapExp then calls the oracle using the adjusted expectation vector $\tilde{\boldsymbol{w}}_t$ as input to compute a refined solution $\tilde{M}_t = \mathrm{Oracle}(\tilde{\boldsymbol{w}}_t)$. If $\tilde{w}_t(\tilde{M}_t) = \tilde{w}_t(M_t)$ then CGapExp stops and returns $\mathsf{Out} = M_t$. 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.

## 3.1 Analysis

Now we prove a problem-dependent sample complexity bound of the CGapExp algorithm. Our sample complexity bound depends on several combinatorial properties of $\mathcal{M}$. Therefore, to formally state our result, we need to introduce several definitions.

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

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

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

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

From Eq. (1), we see that, for each arm $e \notin M_*$, $\Delta_e$ represents the gap between the optimal set $M_*$ and the best set that includes arm $e$; and, for each arm $e \in M_*$, $\Delta_e$ is the sub-optimality of the best

**Algorithm 1** CGapExp: Combinatorial Gap Exploration

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

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

**for** $t = n, n+1, \ldots$ **do**
  $M_t \leftarrow \mathrm{Oracle}(\bar{\boldsymbol{w}}_t)$
  **for** $e = 1, \ldots, n$ **do**
    **if** $e \in M_t$ **then**
      $\tilde{w}_t(e) \leftarrow \bar{w}_t(e) - \mathrm{rad}_t(e)$
    **else**
      $\tilde{w}_t(e) \leftarrow \bar{w}_t(e) + \mathrm{rad}_t(e)$
    **end if**
  **end for**
  $\tilde{M}_t \leftarrow \mathrm{Oracle}(\tilde{\boldsymbol{w}}_t)$
  **if** $\tilde{w}_t(\tilde{M}_t) = \tilde{w}_t(M_t)$ **then**
    $\mathsf{Out} \leftarrow M_t$
    **return** $\mathsf{Out}$
  **end if**
  $p_t \leftarrow \arg\max_{e \in (\tilde{M}_t \setminus M_t) \cup (M_t \setminus \tilde{M}_t)} \mathrm{rad}_t(e)$
  Pull arm $p_t$ and observe the reward
  Update empirical means $\bar{\boldsymbol{w}}_{t+1}$ using the observed reward
  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 $e \neq p_t$
**end for**

set that does not include arm $e$. When specializing to MULTI problem, our definition resembles the previous definition of gaps due to Kalyanakrishnan et al. [3] and Gabillon et al. [2].

**Exchange class and the width of $\mathcal{M}$.** 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 $b = (b_+, b_-)$ where $b_+ \cap b_- = \emptyset$. 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 family 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 $(M' \ominus b) \in \mathcal{M}$. We define the *width* of exchange class $\mathcal{B}$ to be the size of largest exchange set as follows

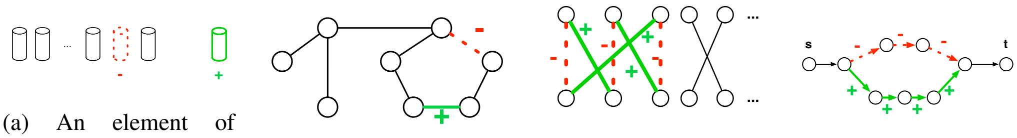
$$\mathrm{width}(\mathcal{B}) = \max_{(b_+, b_-) \in \mathcal{B}} |b_+| + |b_-|. \tag{3}$$

Intuitively, for any feasible sets $M$ and $M'$, there exists an exchange set $(b_+, b_-) \in \mathcal{B}$ belonging to the exchange class $\mathcal{B}$ 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: for any $M \neq M'$, there exists a sequence of exchange sets $b_1, \ldots, b_k$ of $\mathcal{B}$ such that $M' = M \oplus b_1 \oplus \ldots \oplus b_k$.

We notice that an exchange class for $\mathcal{M}$ can be "redundant". It may contains some unnecessary exchange set $b$, such that $M \oplus b \notin \mathcal{M}$ for any $M \in \mathcal{M}$. These redundant exchange sets do not affect our analysis. But allowing them would simplify the construction and description of exchange classes for certain combinatorial problems.

Finally, let $\mathrm{Exchange}(\mathcal{M})$ denote the collection of all possible exchange classes for $\mathcal{M}$. We define the width of a combinatorial problem $\mathcal{M}$ as the width of the thinnest exchange class

$$\mathrm{width}(\mathcal{M}) = \min_{\mathcal{B} \in \mathrm{Exchange}(\mathcal{M})} \mathrm{width}(\mathcal{B}).$$



(a) An element of $\mathcal{B}_{\text{MULTI}}$.

(b) An element of $\mathcal{B}_{\text{MATROID}}$

(c) An element of $\mathcal{B}_{\text{MATCH}}$.

(d) An element of $\mathcal{B}_{\text{PATH}}$.

Figure 1: Examples of exchange sets belonging to the exchange classes $\mathcal{B}_{\text{MULTI}}$, $\mathcal{B}_{\text{MATROID}}$, $\mathcal{B}_{\text{MATCH}}$ and $\mathcal{B}_{\text{PATH}}$: green-solid elements constitute the set $b_+$, red-dotted elements constitute the set $b_-$ and the example exchange set is $b = (b_+, b_-)$. (In Figure 1b, we consider spanning tree as a specific instance for the MATROID problem.)

For many problems, there are exchange classes with small widths that correspond to natural combinatorial structures. To see this, we construct the exchange classes for our running examples. Our constructions are summarized in Fact 1.

**Fact 1.** *There exist exchange classes $\mathcal{B}_{\text{MULTI}}$, $\mathcal{B}_{\text{MATROID}}$, $\mathcal{B}_{\text{MATCH}}$ and $\mathcal{B}_{\text{PATH}}$ for $\mathcal{M}_{\text{MULTI}}$, $\mathcal{M}_{\text{MATROID}}$, $\mathcal{M}_{\text{MATCH}}$ and $\mathcal{M}_{\text{PATH}}$, respectively. These exchange classes can be constructed as follows*

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

*In addition, we have* $\text{width}(\mathcal{B}_{\text{MULTI}}) = 2$, $\text{width}(\mathcal{B}_{\text{MATROID}}) = 2$, $\text{width}(\mathcal{B}_{\text{MATCH}}) = |V|$ *and* $\text{width}(\mathcal{B}_{\text{PATH}}) = |V|$. *This means that* $\text{width}(\mathcal{M}_{\text{MULTI}}) \le 2$, $\text{width}(\mathcal{M}_{\text{MATROID}}) \le 2$, $\text{width}(\mathcal{M}_{\text{MATCH}}) \le |V|$ *and* $\text{width}(\mathcal{M}_{\text{PATH}}) \le |V|$.

We illustrate these exchanges classes in Figure 1. The construction for MULTI problem is straightforward. For MATROID problem, we leverage the basis exchange property of matroids (see Lemma 19 in the supplementary material). And for MATCH and PATH problems, we use standard graph-theoretical properties of matchings and paths. A detailed proof of Fact 1 is deferred to the supplementary material.

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

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

$$T \le O\left(R^2\,\text{width}(\mathcal{M})^2\mathbf{H}\log\left(R^2\,\text{width}(\mathcal{M})^2\mathbf{H}\cdot n/\delta\right)\right), \tag{4}$$

*where $T$ denotes the number of samples used by Algorithm 1 and $\mathbf{H}$ is defined in Eq. (2).*

**Remarks.** For the MULTI problem, we see that Fact 1 shows that $\text{width}(\text{MULTI}) = O(1)$. Therefore, the sample complexity bound of CGapExp is $O\big(\mathbf{H}\log(n\mathbf{H}/\delta)\big)$ for this problem. This matches the previous problem-dependent bounds for the MULTI problem [3, 2]. For the MATROID problem, we know that $\text{width}(\text{MATROID}) = O(1)$ and hence the sample complexity is also $O\big(\mathbf{H}\log(n\mathbf{H}/\delta)\big)$. For MATCH and PATH problem, we see that the sample complexity is bounded by $\tilde{O}(|V|^2\mathbf{H})$.

## 4 Lower Bound

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

We show that, for any family of feasible solutions $\mathcal{M}$ and any expected rewards $\boldsymbol{w}$, any $\delta$-correct algorithm $\mathbb{A}$ must use at least $\Omega\big(\mathbf{H}\log(1/\delta)\big)$ 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 $\varphi_e$ is given by $\varphi_e = \mathcal{N}(w(e), 1)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with mean $\mu$ and variance $\sigma^2$. Then, for any $\delta \in (0, e^{-16}/4)$ and any $\delta$-correct algorithm $\mathbb{A}$, we have*

$$\mathbb{E}[T] \geq \frac{1}{16}\mathbf{H}\log\left(\frac{1}{4\delta}\right), \tag{5}$$

*where $T$ denote the number of total samples used by algorithm $\mathbb{A}$ and $\mathbf{H}$ is defined in Eq. (2).*

Theorem 2 resolves the conjecture of Kalyanakrishnan et al. [3] that the lower bound of sample complexity of MULTI problem is $\Omega\big(\mathbf{H}\log(1/\delta)\big)$. In addition, our upper bound Theorem 1 shows that, for MULTI and MATROID, the sample complexity of CGapExp is $O(\mathbf{H}\log(n\mathbf{H}/\delta))$. Hence, we see that the CGapExp algorithm achieves the optimal sample complexity within logarithmic factors for these two problems.

On the other hand, for general combinatorial problems with non-constant widths, we see that there is a gap of $\tilde{\Theta}(\mathrm{width}(\mathcal{M})^2)$ between the upper bound Eq. (4) and the lower bound Eq. (5). Notice that we have $\mathrm{width}(\mathcal{M}) \leq n$ for any $\mathcal{M}$ and therefore the gap is relatively benign. Ignoring the logarithmic factors, this gap only depends on the underlying combinatorial structure of $\mathcal{M}$. This suggests that the dependency on $\mathbf{H}$ of the sample complexity of CGapExp cannot be improved up to logarithmic factors for general combinatorial problems. Furthermore, we conjecture that the sample complexity lower bound might intrinsically depend on the size of exchange sets (a quantity which is bounded by $\mathrm{width}(\mathcal{M})$). In the supplementary material, we provide evidence on this conjecture which is a lower bound on the sample complexity of exploration on exchange sets.

## 5 Extensions

CGapExp is a general and flexible learning algorithm for the ExpCMAB problem. In this section, we present two extensions to CGapExp that allow it to work in the fixed budget setting and PAC learning setting.

### 5.1 Fixed Budget Setting

We can extend the CGapExp algorithm to the fixed budget setting using 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 uses at most $T$ samples, which meets the requirement of the fixed budget setting. And the second modification bounds the probability 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 $O\left(Tn\exp\left(\frac{-T}{\mathrm{width}(\mathcal{M})^2\mathbf{H}}\right)\right)$.

**Theorem 3.** *Use the same notations as in Theorem 1. Given $T > n$ and parameter $\alpha > 0$, set the confidence radius $\mathrm{rad}_t(e) = R\sqrt{\frac{\alpha}{T_e(t)}}$ for all arms $e \in [n]$ and all $t > 0$. Run CGapExp algorithm for at most $T$ rounds. Then, for $0 \leq \alpha \leq \frac{1}{9}(T-n)\left(R^2\,\mathrm{width}(\mathcal{M})^2\mathbf{H}\right)^{-1}$, we have*

$$\Pr\left[\mathsf{Out} \neq M_*\right] \leq 2Tn\exp\left(-2\alpha\right). \tag{6}$$

The right-hand side of Eq. (6) equals to $O\left(Tn\exp\left(\frac{-T}{\mathrm{width}(\mathcal{M})^2\mathbf{H}}\right)\right)$ when parameter $\alpha = O(T\mathbf{H}^{-1}\,\mathrm{width}(\mathcal{M})^{-2})$. For MULTI problem, we see that this matches the guarantees of the previous fixed budget algorithm due to Gabillon et al. [2].

### 5.2 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 $(\epsilon, \delta)$-PAC algorithm. Formally, an algorithm $\mathbb{A}$ is called an $(\epsilon, \delta)$-PAC algorithm if its output $\mathsf{Out}$ satisfies $\Pr\left[w(M_*) - w(\mathsf{Out}) > \epsilon\right] \le \delta$.

We show that a simple modification on the CGapExp algorithm gives an $(\epsilon, \delta)$-PAC algorithm, with guarantees similar to Theorem 1. In fact, the only modification needed is to change the stopping condition from $\tilde{w}_t(\tilde{M}_t) \le \tilde{w}_t(M_t)$ to $w(\tilde{M}_t) - w(M_t) \le \epsilon$ on line 15 of Algorithm 1. We let CGapExpPAC denote the modified algorithm. In the following theorem, we show that CGapExpPAC is indeed an $(\epsilon, \delta)$-PAC algorithm and has sample complexity similar to CGapExp.

**Theorem 4.** *Use the same notations as in Theorem 1. Fix $\delta \in (0, 1)$ and $\epsilon \ge 0$. Then, with probability at least $1 - \delta$, the output $\mathsf{Out}$ of CGapExpPAC satisfies $w(M_*) - w(\mathsf{Out}) \le \epsilon$. In addition, the number of samples $T$ used by the algorithm satisfies*

$$T \le O\left(R^2 \sum_{e\in[n]} \min\left\{\frac{\mathrm{width}(\mathcal{M})^2}{\Delta_e^2}, \frac{K^2}{\epsilon^2}\right\} \log\left(\frac{R^2 n}{\delta} \sum_{e\in[n]} \min\left\{\frac{\mathrm{width}(\mathcal{M})^2}{\Delta_e^2}, \frac{K^2}{\epsilon^2}\right\}\right)\right), \quad (7)$$

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

We see that the sample complexity of CGapExpPAC decreases when $\epsilon$ increases. And if $\epsilon = 0$, the sample complexity Eq. (7) of CGapExpPAC equals to that of CGapExp.

There are several PAC algorithms for the MULTI problem in the literature with different guarantees [3, 4, 2]. Zhou et al. [4] proposed an $(\epsilon, \delta)$-PAC algorithm for the MULTI problem with a problem-independent sample complexity bound of $O(\frac{K^2 n}{\epsilon^2} + \frac{Kn\log(1/\delta)}{\epsilon^2})$.[1] If we ignore logarithmic factors, then the sample complexity bound of CGapExpPAC for the MULTI problem is better than theirs since $\tilde{O}(\sum_{e\in[n]} \min\{\Delta_e^{-2}, K^2\epsilon^{-2}\}) \le \tilde{O}(nK^2\epsilon^{-2})$. On the other hand, the algorithms of Kalyanakrishnan et al. [3] and Gabillon et al. [2] guarantee to find $K$ arms such that each of them is better than the $K$-th optimal arm within a factor of $\epsilon$ with probability $1 - \delta$. Unless $\epsilon = 0$, their guarantee is different from ours which concerns the optimality of the sum of $K$ arms.

## 6 Agnostic Fixed Budget Algorithm

In this section, we present CGapKill, a parameter-free learning algorithm for the ExpCMAB problem in the fixed budget setting. We analyze the probability of error CGapKill using the tools of exchange classes. Recall that, in the fixed budget setting, an algorithm is given a budget $T > 0$ such that the algorithm can use at most $T$ pulls. The goal of the algorithm is to minimize the probability of error.

**Constrained oracle.** The CGapKill algorithm requires access to a stronger oracle which is referred as *constrained oracle*. A constrained oracle for $\mathcal{M}$ is a function denoted as $\mathrm{COracle} : \mathbb{R}^n \times 2^{[n]} \times 2^{[n]} \to \mathcal{M}$. The function COracle takes three inputs: (1) a weight vector $\boldsymbol{v}$, (2) a set $A$ of positive constraints and (3) a set $B$ of negative constraints and returns a solution such that

$$\mathrm{COracle}(\boldsymbol{v}, A, B) \in \operatorname*{arg\,max}_{M: M\in\mathcal{M}, A\subseteq M, B\cap M=\emptyset} v(M). \quad (8)$$

Hence we see that $\mathrm{COracle}(\boldsymbol{v}, A, B)$ computes an optimal solution that includes all elements of $A$ while exclude all elements of $B$. In the supplementary, we show that constrained oracles can be easily reduced to maximization oracles. In addition, similar to CGapExp, CGapKill does not need any additional knowledge of $\mathcal{M}$ other than accesses to a constrained oracle for $\mathcal{M}$.

[1] We notice that Zhou et al. [4] allow an $(\epsilon', \delta)$-PAC algorithm to produce an output with *average* sub-optimality of $\epsilon'$ in their exposition. This is equivalent to our definition of $(\epsilon, \delta)$-PAC algorithm with $\epsilon = K\epsilon'$ for the MULTI problem.

**CGapKill algorithm.** The idea of the CGapKill algorithm is as follows. The CGapKill algorithm divides the budget of $T$ pulls into $n$ phases. In the end of each phase, CGapKill either accepts or rejects a single arm. If an arm is accepted, then it is included into the final output. Conversely, if an arm is rejected, then it is excluded from the final output. The arms that are neither accepted nor rejected are sampled for a equal number of times in the next phase. Therefore the major challenge here is to choose a correct arm to accept/reject. Clearly, for problems with non-trivial combinatorial structures, one cannot simply accept/reject arms according to their empirical means, since an arm with small expected reward may belong to the optimal solution. We resolve this challenge by using a novel gap estimation method for decision making and demonstrate that CGapKill can achieve a small probability of error.

Now we describe the procedure of the CGapKill algorithm for choosing an arm to accept/reject. Let $A_t$ denote the set of accepted arms before phase $t$ and let $B_t$ denote the set of rejected arms before phase $t$. We call an arm $e$ active if $e \notin A_t \cup B_t$. Then, in phase $t$, CGapKill samples each active arm for $\tilde{T}_t - \tilde{T}_{t-1}$ times, where the definition of $\tilde{T}_t$ is given in Algorithm 2. Next, CGapKill calls the constrained oracle to compute an optimal solution $M_t$ with respect to the empirical means $\bar{w}_t$, accepted arms $A_t$ and rejected arms $B_t$, i.e. let $M_t = \mathrm{COracle}(\bar{\boldsymbol{w}}_t, A_t, B_t)$. Then, for each arm active arm $e$, CGapKill estimate the gap of $e$ in the following way. If $e \in M_t$, then CGapKill computes an optimal solution $\tilde{M}_{t,e}$ that does not include $e$, i.e. $\tilde{M}_{t,e} = \mathrm{COracle}(\bar{\boldsymbol{w}}_t, A_t, B_t \cup \{e\})$. Conversely, if $e \notin M_t$, then CGapKill computes an optimal $\tilde{M}_{t,e}$ which includes $e$, i.e. $\tilde{M}_{t,e} = \mathrm{COracle}(\bar{\boldsymbol{w}}_t, A_t \cup \{e\}, B_t)$. Then, the gap of $e$ is calculated as $\bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,e})$. Finally, CGapKill chooses the arm $p_t$ with the largest gap. If $p_t \in M_t$ then $p_t$ is accepted otherwise $p_t$ is rejected. The pseudo-code of CGapKill is shown in Algorithm 2.

---

**Algorithm 2** CGapKill: Combinatorial Gap-based Elimination

---

**Require:** Budget: $T > 0$; Constrained oracle: $\mathrm{COracle} : \mathbb{R}^n \times 2^{[n]} \times 2^{[n]} \to \mathcal{M}$.

Define $\tilde{\log}(n) \triangleq \sum_{i=1}^{n} \frac{1}{i}$
$\tilde{T}_0 \leftarrow 0, A_1 \leftarrow \emptyset, B_1 \leftarrow \emptyset$
**for** $t = 1, \ldots, n$ **do**
  $\tilde{T}_t \leftarrow \left\lceil \frac{T-n}{\tilde{\log}(n)(n-t+1)} \right\rceil$
  Pull each arm $e \in [n] \backslash (A_t \cup B_t)$ for $\tilde{T}_t - \tilde{T}_{t-1}$ times
  Update the empirical means $\bar{\boldsymbol{w}}_t \in \mathbb{R}^n$ of each arm
  $M_t \leftarrow \mathrm{COracle}(\bar{\boldsymbol{w}}_t, A_t, B_t)$
  **for each** $e \in [n] \backslash (A_t \cup B_t)$ **do**
    **if** $e \in M_t$ **then**
      $\tilde{M}_{t,e} \leftarrow \mathrm{COracle}(\bar{\boldsymbol{w}}_t, A_t, B_t \cup \{e\})$
    **else**
      $\tilde{M}_{t,e} \leftarrow \mathrm{COracle}(\bar{\boldsymbol{w}}_t, A_t \cup \{e\}, B_t)$
    **end if**
  **end for**
  $p_t \leftarrow \arg\max_{i \in [n] \backslash (A_t \cup B_t)} \bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,i})$
  **if** $p_t \in M_t$ **then**
    $A_{t+1} \leftarrow A_t \cup \{p_t\}, B_{t+1} \leftarrow B_t$
  **else**
    $A_{t+1} \leftarrow A_t, B_{t+1} \leftarrow B_t \cup \{p_t\}$
  **end if**
**end for**
Out $\leftarrow A_{n+1}$
**return** Out

---

**Analysis.** We show that the CGapKill algorithm guarantees a probability of error upper bounded by $\tilde{O}(\exp(-T\,\mathrm{width}(\mathcal{M})^{-2}\mathbf{H}^{-1}))$ in the following theorem.

**Theorem 5.** *Use the same notations as in Theorem 1. Let $\Delta_{(1)}, \ldots, \Delta_{(n)}$ be a permutation of $\Delta_1, \ldots, \Delta_n$ such that $\Delta_{(1)} \leq \ldots\ldots \Delta_{(n)}$. Define $\mathbf{H}_2 \triangleq \max_{i \in [n]} i\Delta_{(i)}^{-2}$.*

*Given budget $T > n$, there exists an algorithm (in particular, Algorithm 2) which uses at most $T$ samples and outputs a solution $\mathsf{Out} \in \mathcal{M}$ such that*

$$\Pr[\mathsf{Out} \neq M_*] \leq n^2 \exp\left(-\frac{2(T-n)}{9R^2 \tilde{\log}(n)\,\mathrm{width}(\mathcal{M})^2 \mathbf{H}_2}\right), \tag{9}$$

*where $\tilde{\log}(n) \triangleq \sum_{i=1}^{n} \frac{1}{i}$.*

The quantity $\mathbf{H}_2$ defined in Theorem 5 can be considered as a surrogate of $\mathbf{H}$ since these two quantities are equivalent up to a logarithmic factor: $\mathbf{H}_2 \leq \mathbf{H}_1 \leq \log(2n)\mathbf{H}_2$. For the MULTI problem, we see that this matches previous fixed budget algorithm due to Bubeck et al. [1].

We see that $\mathrm{width}(\mathcal{M})^2$ also appears in the probability of error Eq. (9) in Theorem 5, which indicates that $\mathrm{width}(\mathcal{M})$ may reflect the inherent hardness associate with a combinatorial problem $\mathcal{M}$ for the ExpCMAB problem. Moreover, the analysis of CGapKill re-uses many tools and ideas within the framework of exchange classes which are developed for the analysis of our main algorithm CGapExp. We use these tools to design gadgets in order to bound the estimations of gaps. This suggests that the framework of exchange class may be of interest for the analysis of other similar on-line combinatorial optimization problems.

We notice that the CGapKill algorithm uses solely the empirical means of arms to support its decision making; while our main algorithm CGapExp also need to use explicitly constructed confidence intervals of the arms. From this viewpoint, the CGapKill algorithm is very different from our main algorithm CGapExp. This actually leads to an argument substantially different from the proof of Theorem 1. And we have to build different gadgets using the exchange classes.

### 6.1 Analysis of CGapKill

**Notations.** For convenience, we will use the following notations in the rest of this section. Let $\boldsymbol{w} \in \mathbb{R}^n$ be the vector expected rewards of arms. Let $M_* = \arg\max_{M \in \mathcal{M}} w(M)$ be the optimal solution. Let $T$ be the budget of samples. Let $\Delta_{(1)}, \ldots, \Delta_{(n)}$ be a permutation of $\Delta_1, \ldots, \Delta_n$ such that $\Delta_{(1)} \leq \ldots\ldots \Delta_{(n)}$. Let $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ be two sequence of sets which are defined in Algorithm X.

#### 6.1.1 Confidence Intervals

**Lemma 1.** *Given a phase $t \in [n]$, we define random event $\tau_t$ as follows*

$$\tau_t = \left\{ \forall i \in [n] \backslash (A_t \cup B_t) \quad \left|\bar{w}_t(i) - w(i)\right| < \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})} \right\}. \tag{10}$$

*Then, we have*

$$\Pr\left[\bigcap_{t=1}^{T} \tau_t\right] \geq 1 - n^2 \exp\left(-\frac{2(T-n)}{9R^2 \tilde{\log}(n)\,\mathrm{width}(\mathcal{M})^2 \mathbf{H}_2}\right). \tag{11}$$

*Proof.* Let us consider an arbitrary phase $t \in [n]$ and an arbitrary active arm $i \in [n]\backslash(A_t \cup B_t)$ of phase $t$.

Notice that the arm $e$ has been pulled for $\tilde{T}_t$ times during phases $1, \ldots, t$. Therefore, by Hoeffding's inequality, we have

$$\Pr\left[\left|\bar{w}_t(i) - w(i)\right| \geq \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})}\right] \leq 2\exp\left(-\frac{2\tilde{T}_t \Delta_{(n-t+1)}^2}{9R^2\,\mathrm{width}(\mathcal{M})^2}\right). \tag{12}$$

By plugging the definition of $\tilde{T}_t$, the quantity $\tilde{T}_t \Delta_{(n-t+1)}^2$ on the right-hand side of Eq. (12) can be further bounded by

$$\tilde{T}_t \Delta_{(n-t+1)}^2 \geq \frac{T-n}{\tilde{\log}(n)(n-t+1)} \Delta_{(n-t+1)}^2$$

$$\geq \frac{T-n}{\tilde{\log}(n)\mathbf{H}_2},$$

where the last inequality follows from the definition of $\mathbf{H}_2$. By plugging the last inequality into Eq. (12), we have

$$\Pr\left[\left|\bar{w}_t(i) - w(i)\right| \geq \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})}\right] \leq 2\exp\left(-\frac{2(T-n)}{9R^2\tilde{\log}(n)\,\mathrm{width}(\mathcal{M})^2\mathbf{H}_2}\right). \tag{13}$$

Now using Eq. (13) and a union bound for all $t \in [n]$ and all $i \in [n]\backslash(A_t \cup B_t)$, we have

$$\begin{aligned}\Pr\left[\bigcap_{t=1}^{n}\tau_t\right] &\geq 1 - 2\sum_{t=1}^{n}(n-t+1)\exp\left(-\frac{2(T-n)}{9R^2\tilde{\log}(n)\,\mathrm{width}(\mathcal{M})^2\mathbf{H}_2}\right)\\ &\geq 1 - n^2\exp\left(-\frac{2(T-n)}{9R^2\tilde{\log}(n)\,\mathrm{width}(\mathcal{M})^2\mathbf{H}_2}\right).\end{aligned}$$

□

**Lemma 2.** *Fix a phase $t \in [n]$, suppose that random event $\tau_t$ occurs. For any vector $\boldsymbol{a} \in \mathbb{R}^n$, suppose that $\mathrm{supp}(\boldsymbol{a}) \cap (A_t \cup B_t) = \emptyset$, where $\mathrm{supp}(\boldsymbol{a}) \triangleq \{i \mid a(i) \neq 0\}$ is support of $\boldsymbol{a}$. Then, we have*

$$|\langle\bar{\boldsymbol{w}}_t, \boldsymbol{a}\rangle - \langle\boldsymbol{w}, \boldsymbol{a}\rangle| \leq \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})}\left\|\boldsymbol{a}\right\|_1.$$

*Proof.* Suppose that $\tau_t$ occurs. Then, similar to the proof of Lemma 12, we have

$$\begin{aligned}|\langle\bar{\boldsymbol{w}}_t, \boldsymbol{a}\rangle - \langle\boldsymbol{w}, \boldsymbol{a}\rangle| &= |\langle\bar{\boldsymbol{w}}_t - \boldsymbol{w}, \boldsymbol{a}\rangle|\\ &= \left|\sum_{i=1}^{n}\big(\bar{w}_t(i) - w(i)\big)a(i)\right|\\ &\leq \left|\sum_{i\in[n]\backslash(A_t\cup B_t)}\big(\bar{w}_t(i) - w(i)\big)a(i)\right| \qquad (14)\\ &\leq \sum_{i\in[n]\backslash(A_t\cup B_t)}\left|\big(\bar{w}_t(i) - w(i)\big)a(i)\right|\\ &\leq \sum_{i\in[n]\backslash(A_t\cup B_t)}|\bar{w}_t(i) - w(i)|\,|a(i)|\\ &\leq \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})}\sum_{i\in[n]\backslash(A_t\cup B_t)}|a(i)| \qquad (15)\\ &= \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})}\left\|\boldsymbol{a}\right\|_1,\end{aligned}$$

where Eq. (14) follows from the assumption that $\boldsymbol{a}$ supported on $[n]\backslash(A_t \cup B_t)$; Eq. (15) follows from the definition of $\tau_t$ (Eq. (10)). □

### 6.1.2 Main Lemmas

**Lemma 3.** *Fix a phase $t \in [n]$. Suppose that $A_t \subseteq M_*$ and $B_t \cap M_* = \emptyset$. Let $M$ be a set such that $A_t \subseteq M$ and $B_t \cap M = \emptyset$. Let $a$ and $b$ be two sets satisfying that $a \subseteq M\backslash M_*$, $b \subseteq M_*\backslash M$ and $a \cap b = \emptyset$. Then, we have*

$$A_t \subseteq (M\backslash a \cup b) \quad \textit{and} \quad B_t \cap (M\backslash a \cup b) = \emptyset \quad \textit{and} \quad (a \cup b) \cap (A_t \cup B_t) = \emptyset.$$

*Proof.* We prove the first part as follows

$$
\begin{aligned}
A_t \cap (M\backslash a \cup b) &= (A_t \cap (M\backslash a)) \cup (A_t \cap b) \\
&= A_t \cap (M\backslash a) && (16)\\
&= (A_t \cap M)\backslash a \\
&= A_t\backslash a && (17)\\
&= A_t, && (18)
\end{aligned}
$$

where Eq. (16) holds since we have $A_t \cap b \subseteq A_t \cap (M_*\backslash M) \subseteq M \cap (M_*\backslash M) = \emptyset$; Eq. (17) follows from $A_t \subseteq M$; and Eq. (18) follows from $a \subseteq M\backslash M_*$ and $A_t \subseteq M_*$ which imply that $a \cap A_t = \emptyset$. Notice that Eq. (18) is equivalent to $A_t \subseteq (M\backslash a \cup b)$.

Then, we proceed to prove the second part in the following

$$
\begin{aligned}
B_t \cap (M\backslash a \cup b) &= (B_t \cap (M\backslash a)) \cup (B_t \cap b) \\
&= B_t \cap (M\backslash a) && (19)\\
&= (B_t \cap M)\backslash a \\
&= \emptyset\backslash a = \emptyset, && (20)
\end{aligned}
$$

where Eq. (19) follows from the fact that $B_t \cap b \subseteq B_t \cap (M_*\backslash M) \subseteq \neg M_* \cap (M_*\backslash M) = \emptyset$; and Eq. (20) follows from the fact that $B_t \cap M = \emptyset$.

Last, we prove the third part. By combining the assumptions that $A_t \subseteq M_*$ and $A_t \subseteq M$, we see that $A_t \subseteq M \cap M_*$. Therefore, we have

$$
(a \cap A_t) \cup (b \cap A_t) \subseteq ((M\backslash M_*) \cap (M \cap M_*)) \cup ((M_*\backslash M) \cap (M \cap M_*)) = \emptyset. \tag{21}
$$

Similarly, we have $B_t \subseteq \neg M \cap \neg M_*$. Hence, we derive

$$
(a \cap B_t) \cup (b \cap B_t) \subseteq ((M\backslash M_*) \cap (\neg M \cap \neg M_*)) \cup ((M_*\backslash M) \cap (\neg M \cap \neg M_*)) = \emptyset. \tag{22}
$$

By combining Eq. (21) and Eq. (22), we obtain

$$
(a \cup b) \cap (A_t \cup B_t) = (a \cap A_t) \cup (b \cap A_t) \cup (a \cap B_t) \cup (b \cap B_t) = \emptyset.
$$

□

**Lemma 4.** *Fix any round $t > 0$. Suppose that event $\tau_t$ occurs. Also assume that $A_t \subseteq M_*$ and $B_t \cap M_* = \emptyset$. Let $e \in [n]\backslash(A_t \cup B_t)$ be an active arm. Suppose that $\Delta_{(t-n+1)} \leq \Delta_e$. Then, we have $e \in (M_* \cap M_t) \cup (\neg M_* \cap \neg M_t)$.*

*Proof.* Fix an exchange class $\mathcal{B} \in \arg\min_{\mathcal{B}' \in \text{Exchange}(\mathcal{M})} \text{width}(\mathcal{B}')$. Suppose that $e \notin (M_* \cap M_t) \cup (\neg M_* \cap \neg M_t)$. This is equivalent to the following

$$
e \in (M_* \cap \neg M_t) \cup (\neg M_* \cap M_t). \tag{23}
$$

Eq. (23) can be further rewritten as

$$
e \in (M_*\backslash M_t) \cup (M_t\backslash M_*).
$$

From this assumption, it is easy to see that $M_t \neq M_*$. Therefore we can apply Lemma 8. Then we know that there exists $b = (b_+, b_-) \in \mathcal{B}$ such that $e \in b_- \cup b_+$, $b_- \subseteq M_t\backslash M_*$, $b_+ \subseteq M_*\backslash M_t$, $M_t \oplus b \in \mathcal{M}$ and $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e \geq 0$.

Using Lemma 3, we see that $(M_t \oplus b) \cap B_t = \emptyset$, $A_t \subseteq (M_t \oplus b)$ and $(b_+ \cup b_-) \cap (A_t \cup B_t) = \emptyset$. Now recall the definition $M_t \in \arg\max_{M \in \mathcal{M}, A_t \subseteq M, B_t \cap M = \emptyset} \bar{w}_t(M)$ and also recall that $M_t \oplus b \in \mathcal{M}$. Therefore, we obtain that

$$
\bar{w}_t(M_t) \geq \bar{w}_t(M_t \oplus b). \tag{24}
$$

On the other hand, we have

$$
\begin{aligned}
\bar{w}_t(M_t \oplus b) &= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} + \boldsymbol{\chi}_b \rangle && (25)\\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \rangle + \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle
\end{aligned}
$$

$$> \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \rangle + \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})} \|\boldsymbol{\chi}_b\|_1 \tag{26}$$

$$> \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \rangle + \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{M})} \|\boldsymbol{\chi}_b\|_1$$

$$\geq \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \rangle + \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{\Delta_e}{3} \tag{27}$$

$$\geq \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \rangle + \frac{2}{3}\Delta_e \tag{28}$$

$$\geq \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \rangle = \bar{w}_t(M_t), \tag{29}$$

where Eq. (25) follows from Lemma 7; Eq. (26) follows from Lemma 2 and the fact that $(b_+ \cup b_-) \cap (A_t \cup B_t) = \emptyset$; Eq. (27) holds since $b \in \mathcal{B}$ which implies that $\|\boldsymbol{\chi}_b\|_1 = |b_+| + |b_-| \leq \mathrm{width}(\mathcal{B}) = \mathrm{width}(\mathcal{M})$; and Eq. (28) and Eq. (29) hold since we have shown that $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e \geq 0$.

This means that $\bar{w}_t(M_t \oplus b) > \bar{w}_t(M_t)$. This contradicts to Eq. (24). Therefore we have $e \in (M_* \cap M_t) \cup (\neg M_* \cap \neg M_t)$.

□

**Lemma 5.** *Fix any round $t > 0$. Suppose that event $\tau_t$ occurs. Also assume that $A_t \subseteq M_*$ and $B_t \cap M_* = \emptyset$. Let $e \in [n] \backslash (A_t \cup B_t)$ be an active arm such that $\Delta_{(t-n+1)} \leq \Delta_e$. Then, we have*

$$\bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,e}) \geq \frac{2}{3}\Delta_{(t-n+1)}.$$

*Proof.* By Lemma 4, we see that

$$e \in (M_* \cap M_t) \cup (\neg M_* \cap \neg M_t). \tag{30}$$

We claim that $e \in (\tilde{M}_{t,e} \backslash M_*) \cup (M_* \backslash \tilde{M}_{t,e})$ and therefore $M_* \neq \tilde{M}_{t,e}$. By Eq. (30), we see that either $e \in (M_* \cap M_t)$ or $e \in (\neg M_* \cap \neg M_t)$. First let us assume that $e \in M_* \cap M_t$. Then, by definition of $\tilde{M}_{t,e}$, we see that $e \notin \tilde{M}_{t,e}$. Therefore $e \in M_t \backslash \tilde{M}_{t,e}$. On the other hand, suppose that $e \in \neg M_* \cap \neg M_t$. Then, we see that $e \in \tilde{M}_{t,e}$. This means that $e \in \tilde{M}_{t,e} \backslash M_*$.

Hence we can apply Lemma 8. Then we obtain that there exists $b = (b_+, b_-) \in \mathcal{B}$ such that $e \in b_+ \cup b_-$, $b_+ \subseteq M_* \backslash \tilde{M}_{t,e}$, $b_- \subseteq \tilde{M}_{t,e} \backslash M_*$, $\tilde{M}_{t,e} \oplus b \in \mathcal{M}$ and $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e$.

Define $M'_{t,e} \triangleq \tilde{M}_{t,e} \oplus b$. Using Lemma 3, we have $A_t \subseteq M'_{t,e}$, $B_t \cap M'_{t,e} = \emptyset$ and $(b_+ \cup b_-) \cap (A_t \cup B_t) = \emptyset$. Since $M'_{t,e} \in \mathcal{M}$ and by definition $M_t \in \arg\max_{M \in \mathcal{M}, A_t \subseteq M, B_t \cap M = \emptyset} \bar{w}_t(M)$, we have

$$\bar{w}_t(M_t) \geq \bar{w}_t(M'_{t,e}). \tag{31}$$

Hence, we have

$$\begin{aligned}
\bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,e}) &\geq \bar{w}_t(M'_{t,e}) - \bar{w}_t(\tilde{M}_{t,e}) \\
&= \bar{w}_t(\tilde{M}_{t,e} \oplus b) - \bar{w}_t(\tilde{M}_{t,e}) \\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_{t,e}} + \boldsymbol{\chi}_b \right\rangle - \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_{t,e}} \right\rangle && (32) \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle \\
&> \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{B})} \|\boldsymbol{\chi}_b\|_1 && (33) \\
&\geq \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})} \|\boldsymbol{\chi}_b\|_1 && (34) \\
&\geq \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{\Delta_e}{3} && (35) \\
&\geq \frac{2}{3}\Delta_e \geq \frac{2}{3}\Delta_{(n-t+1)}, && (36)
\end{aligned}$$

where Eq. (32) follows from Lemma 7; Eq. (33) follows from Lemma 2, the assumption on event $\tau_t$ and the fact $(b_+ \cup b_-) \cap (A_t \cup B_t) = \emptyset$; Eq. (34) follows from the assumption that $\Delta_e \geq \Delta_{(n-t+1)}$; Eq. (35) holds since $b \in \mathcal{B}$ and therefore $\|\boldsymbol{\chi}_b\|_1 \leq \operatorname{width}(\mathcal{M})$; and Eq. (36) follows from the fact that $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e$.

□

**Lemma 6.** *Fix any phase $t > 0$. Suppose that event $\xi_t$ occurs. Suppose an active arm $e \in [n] \backslash (A_t \cup B_t)$ satisfies that $e \in (M_* \cap \neg M_t) \cup (\neg M_* \cap M_t)$. Then, we have*

$$\bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,e}) \leq \frac{1}{3}\Delta_{(n-t+1)}.$$

*Proof.* Fix an exchange class $\mathcal{B} \in \arg\min_{\mathcal{B}' \in \operatorname{Exchange}(\mathcal{M})} \operatorname{width}(\mathcal{B}')$.

The assumption that $e \in (M_* \cap \neg M_t) \cup (\neg M_* \cap M_t)$ can be rewritten as $e \in (M_* \backslash M_t) \cup (M_t \backslash M_*)$. This shows that $M_t \neq M_*$ Hence Lemma 8 applies here. Therefore we know that there exists $b = (b_+, b_-) \in \mathcal{B}$ such that $e \in (b_+ \cup b_-)$, $b_+ \subseteq M_* \backslash M_t$, $b_- \subseteq M_t \backslash M_*$, $M_t \oplus b \in \mathcal{M}$ and $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e \geq 0$.

Define $M'_{t,e} \triangleq M_t \oplus b$. We claim that

$$\bar{w}_t(\tilde{M}_{t,e}) \geq \bar{w}_t(M'_{t,e}). \tag{37}$$

By definition of $\tilde{M}_{t,e}$, we only need to show that **(a):** $e \in (M'_{t,e} \backslash M_t) \cup (M_t \backslash M'_{t,e})$ and **(b):** $A_t \subseteq M'_{t,e}$ and $B_t \cap M'_{t,e} = \emptyset$. First we prove **(a)**. Notice that $b_+ \cap b_- = \emptyset$ and $b_- \subseteq M_t$. Hence we see that $M'_{t,e} \backslash M_t = (M_t \backslash b_- \cup b_+) \backslash M_t = b_+$ and $M_t \backslash M'_{t,e} = M_t \backslash (M_t \backslash b_- \cup b_+) = b_-$. In addition, we have that $e \in (b_- \cup b_+)$. Therefore we see that **(a)** holds by combining these relations. Next, we notice that **(b)** follows directly from Lemma 3. Hence we have shown that Eq. (37) holds.

Hence, we have

$$\begin{aligned}
\bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,e}) &\leq \bar{w}_t(M_t) - \bar{w}_t(M'_{t,e}) \\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \right\rangle - \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} + \boldsymbol{\chi}_b \right\rangle && (38) \\
&= -\left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \right\rangle \\
&\leq -\left\langle \boldsymbol{w}, \boldsymbol{\chi}_b \right\rangle + \frac{\Delta_{(n-t+1)}}{3 \operatorname{width}(\mathcal{M})} \|\boldsymbol{\chi}_b\|_1 && (39) \\
&\leq \frac{\Delta_{(n-t+1)}}{3 \operatorname{width}(\mathcal{M})} \|\boldsymbol{\chi}_b\|_1 \leq \frac{\Delta_{(n-t+1)}}{3}, && (40)
\end{aligned}$$

where Eq. (38) follows from Lemma 7; Eq. (39) follows from Lemma 2, the assumption on $\tau_t$ and $(b_+ \cup b_-) \cap (A_t \cup B_t) = \emptyset$ (by Lemma 3); and Eq. (40) follows from the fact $\|\boldsymbol{\chi}_b\|_1 \leq \operatorname{width}(\mathcal{M})$ (since $b \in \mathcal{B}$) and that $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e \geq 0$. □

### 6.1.3 Proof of Theorem 5

For reader's convenience, we first restate Theorem 5 in the following.

**Theorem 5.** *Use the same notations as in Theorem 1. Let $\Delta_{(1)}, \ldots, \Delta_{(n)}$ be a permutation of $\Delta_1, \ldots, \Delta_n$ such that $\Delta_{(1)} \leq \ldots\ldots \Delta_{(n)}$. Define $\mathbf{H}_2 \triangleq \max_{i \in [n]} i \Delta_{(i)}^{-2}$.*

*Given budget $T > n$, there exists an algorithm (in particular, Algorithm 2) which uses at most $T$ samples and outputs a solution $\mathsf{Out} \in \mathcal{M}$ such that*

$$\Pr[\mathsf{Out} \neq M_*] \leq n^2 \exp\left(-\frac{2(T-n)}{9R^2 \tilde{\log}(n) \operatorname{width}(\mathcal{M})^2 \mathbf{H}_2}\right), \tag{9}$$

*where $\tilde{\log}(n) \triangleq \sum_{i=1}^{n} \frac{1}{i}$.*

*Proof.* First, we show that the algorithm at most $T$ samples. It is easy to see that exactly one arm is pulled for $\tilde{T}_1$ times, one arm is pulled for $\tilde{T}_2$ times, $\ldots$, and one arm is pulled for $\tilde{T}_n$ times. Therefore, the total number samples used by the algorithm is bounded by

$$\begin{aligned}\sum_{t=1}^{n} \tilde{T}_t &\leq \sum_{t=1}^{n} \left( \frac{T-n}{\tilde{\log}(n)(n-t+1)} + 1 \right) \\ &= \frac{T-n}{\tilde{\log}(n)} \tilde{\log}(n) + n = T.\end{aligned}$$

By Lemma 1, we know that the event $\tau \triangleq \bigcap_{t=1}^{T} \tau_t$ occurs with probability at least $1 - n^2 \exp\left(-\frac{2(T-n)}{9R^2 \tilde{\log}(n)\, \mathrm{width}(\mathcal{M})^2 \mathbf{H}_2}\right)$. Therefore, we only need to prove that, under event $\tau$, the algorithm outputs $M_*$. We will assume that event $\tau$ occurs in the rest of the proof.

We prove by induction. Fix a phase $t \in [t]$. Suppose that the algorithm does not make any error before phase $t$, i.e. $A_t \subseteq M_*$ and $B_t \cap M_* = \emptyset$. We show that the algorithm does not err at phase $t$.

In the beginning phase $t$, there are only $t-1$ inactive arms $|A_t \cup B_t| = t-1$. Therefore there must exists an active arm $e_1 \in [n] \backslash (A_t \cup B_t)$ such that $\Delta_{e_1} \geq \Delta_{(n-t+1)}$. Hence, by Lemma 5, we have

$$\bar{w}_t(M_t) - \bar{w}_t(M_{t,e_1}) \geq \frac{2}{3} \Delta_{(n-t+1)}. \tag{41}$$

Notice that the algorithm makes an error on phase $t$ if and only if it accepts an arm $p_t \notin M_*$ or rejects an arm $p_t \in M_*$. On the other hand, by design, arm $p_t$ is accepted when $p_t \in M_t$ and is rejected when $p_t \notin M_t$. Therefore, we see that the algorithm makes an error on phase $t$ if and only if $p_t \in (M_* \cap \neg M_t) \cup (\neg M_* \cap M_t)$.

Suppose that $p_t \in (M_* \cap \neg M_t) \cup (\neg M_* \cap M_t)$. Now appeal to Lemma 6, we see that

$$\bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,p_t}) \leq \frac{1}{3} \Delta_{(n-t+1)}. \tag{42}$$

By combining Eq. (41) and Eq. (42), we see that

$$\bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,p_t}) \leq \frac{1}{3} \Delta_{(n-t+1)} < \frac{2}{3} \Delta_{(n-t+1)} \leq \bar{w}_t(M_t) - \bar{w}_t(M_{t,e_1}). \tag{43}$$

However Eq. (43) is contradictory to the definition of $p_t \triangleq \arg\max_{i \in [n] \backslash (A_t \cup B_t)} \bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,i})$. Therefore we have proved that $p_t \notin (M_* \cap \neg M_t) \cup (\neg M_* \cap M_t)$. This means that the algorithm does not err at phase $t$, or equivalently $A_{t+1} \subseteq M_*$ and $B_{t+1} \cap M_* = \emptyset$. By induction, we have proved that the algorithm does not err at any phase $t \in [n]$.

Hence we have $A_{n+1} \subseteq M_*$ and $B_{n+1} \subseteq \neg M_*$. Notice that $|A_{n+1}| + |B_{n+1}| = n$ and $A_{n+1} \cap B_{n+1} = \emptyset$. This means that $A_{n+1} = M_*$ and $B_{n+1} = \neg M_*$. Therefore the algorithm outputs $\mathsf{Out} = A_{n+1} = M_*$ after phase $n$. □

## References


[1] Séebastian Bubeck, Tengyao Wang, and Nitin Viswanathan. Multiple identifications in multi-armed bandits. In *ICML*, pages 258–265, 2013.

[2] Victor Gabillon, Mohammad Ghavamzadeh, Alessandro Lazaric, and Team SequeL. Best arm identification: A unified approach to fixed budget and fixed confidence. In *NIPS*, pages 3221–3229, 2012.

[3] Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In *ICML*, pages 655–662, 2012.

[4] Yuan Zhou, Xi Chen, and Jian Li. Optimal pac multiple arm identification with applications to crowdsourcing. In *ICML*, 2014.

# A 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 $\mathbf{rad}_t = \big(\mathrm{rad}_t(1), \ldots, \mathrm{rad}_t(n)\big)^T$ 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 $\boldsymbol{u} \circ \boldsymbol{v} \triangleq \big(u(1) \cdot v(1), \ldots, u(n) \cdot v(n)\big)^T$ as the element-wise product of $\boldsymbol{u}$ and $\boldsymbol{v}$. For any $s \in \mathbb{R}$, we also define $\boldsymbol{u}^s \triangleq \big(u(1)^s, \ldots, u(n)^s\big)^T$ as the element-wise exponentiation of $\boldsymbol{u}$. Let $|\boldsymbol{u}| = \big(|u(1)|, \ldots, |u(n)|\big)^T$ denote the element-wise absolute value of $\boldsymbol{u}$.

## A.1 Preparatory Lemmas

**Lemma 7.** *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

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

□

**Lemma 8.** *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 $\Delta_e$ 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 7 as follows

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

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 $w(M_2) \le w(M_3)$. Hence, we have $w(M_*) - w(M_2) \ge w(M_*) - w(M_3) = \Delta_e$. Similar to Case (1), applying Lemma 7 again, we have

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

□

**Lemma 9.** *Let $M$ and $M'$ be two sets. Then, we have*

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

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

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

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

**Lemma 10.** *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

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

□

**Lemma 11.** *Let $M_t$ and $\tilde{\boldsymbol{w}}_t$ be defined in Algorithm 1. Let $M' \in \mathcal{M}$ be a feasible set. We have*

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

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

$$\begin{aligned} \left\langle \tilde{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle &= \left\langle \bar{\boldsymbol{w}}_t + \mathbf{rad}_t \circ (1 - 2\boldsymbol{\chi}_{M_t}),\ \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle \\ &= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, (\mathbf{1}_n - 2\boldsymbol{\chi}_{M_t}) \circ (\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t}) \right\rangle \qquad (44) \\ &= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} - 2\boldsymbol{\chi}_{M_t} \circ \boldsymbol{\chi}_{M'} + 2\boldsymbol{\chi}^2_{M_t} \right\rangle \\ &= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, \boldsymbol{\chi}^2_{M'} - \boldsymbol{\chi}^2_{M_t} - 2\boldsymbol{\chi}_{M_t} \circ \boldsymbol{\chi}_{M'} + 2\boldsymbol{\chi}^2_{M_t} \right\rangle \qquad (45) \\ &= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, (\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t})^2 \right\rangle \\ &= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, |\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t}| \right\rangle, \qquad (46) \end{aligned}$$

where Eq. (44) follows from Lemma 10; Eq. (45) holds since $\boldsymbol{\chi}_{M'} \in \{0, 1\}^n$ and $\boldsymbol{\chi}_{M_t} \in \{0, 1\}^n$ and therefore $\boldsymbol{\chi}_{M'} = \boldsymbol{\chi}^2_{M'}$ and $\boldsymbol{\chi}_{M_t} = \boldsymbol{\chi}^2_{M_t}$; and Eq. (46) follows since $\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \in \{-1, 0, 1\}^n$. □

### A.2 Confidence Intervals

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

$$\xi_t = \Big\{\forall i \in [n], \quad |w(i) - \bar{w}_t(i)| \le \mathrm{rad}_t(i)\Big\}. \tag{47}$$

We notice that random event $\xi_t$ 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. (47) to inner products as follows.

**Lemma 12.** *Given any $t > 0$, assume that event $\xi_t$ as defined in Eq. (47) occurs. Then, for any vector $\boldsymbol{a} \in \mathbb{R}^n$, we have*

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

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

$$\begin{aligned}
\big| \langle \boldsymbol{w}, \boldsymbol{a} \rangle - \langle \bar{\boldsymbol{w}}_t, \boldsymbol{a} \rangle \big| &= \big| \langle \boldsymbol{w} - \bar{\boldsymbol{w}}_t, \boldsymbol{a} \rangle \big| \\
&= \left| \sum_{i=1}^{n} \big(w(i) - \bar{w}_t(i)\big) a(i) \right| \\
&\le \sum_{i=1}^{n} \big|w(i) - \bar{w}_t(i)\big| |a(i)| \\
&\le \sum_{i=1}^{n} \mathrm{rad}_t(i) \cdot |a(i)| \qquad (48) \\
&= \langle \mathbf{rad}_t, |\boldsymbol{a}| \rangle ,
\end{aligned}$$

where Eq. (48) follows the definition of event $\xi_t$ in Eq. (47) and the assumption that it occurs. □

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

**Lemma 13.** *Suppose that the reward distribution $\varphi_e$ 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*

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

*Proof.* For any $t > 0$ and $e \in [n]$, notice $\varphi_e$ is a $R$-sub-Gaussian distribution with mean $w(e)$ and $w_t(e)$ is the empirical mean of $\varphi_e$ for $T_e(t)$ samples. Using Hoeffding's inequality (see Lemma 20 in Section D), we obtain

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

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

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

□

### A.3 Main Lemmas

**Lemma 14.** *Given any $t > 0$, assume that event $\xi_t$ (defined in Eq. (47)) 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 8 by setting $M = M_t$ and $M' = M_*$, we see that there exists $b = (b_+, b_-) \in \mathcal{B}$ such that $(M_t \oplus b) \in \mathcal{M}$.

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

$$
\begin{aligned}
\tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) &\geq \tilde{w}_t(M'_t) - \tilde{w}_t(M_t) \\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'_t} - \boldsymbol{\chi}_{M_t} \right\rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t}| \rangle && (49) \\
&\geq \left\langle \boldsymbol{w}, \boldsymbol{\chi}_{M'_t} - \boldsymbol{\chi}_{M_t} \right\rangle && (50) \\
&= w(M'_t) - w(M_t) > 0, && (51)
\end{aligned}
$$

where Eq. (49) follows from Lemma 11; and Eq. (50) follows the assumption that event $\xi_t$ occurs and Lemma 12;

Therefore Eq. (51) 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 $t$. □

**Lemma 15.** *Given any $t > 0$ and suppose that event $\xi_t$ (defined in Eq. (47)) occurs. For any $e \in [n]$, if $\mathrm{rad}_t(e) < \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{M})}$, then, arm $e$ will not be pulled on round $t$, i.e. $p_t \neq e$.*

*Proof.* Fix an exchange class $\mathcal{B} \in \arg\min_{\mathcal{B}' \in \mathrm{Exchange}(\mathcal{M})} \mathrm{width}(\mathcal{B}')$. Suppose, in the contrary, that $p_t = e$. By Lemma 8, 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 8 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 8, 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 9, we obtain

$$
\begin{aligned}
\|\mathbf{rad}_t \circ (\boldsymbol{d} - \boldsymbol{\chi}_c)\|_\infty &= \left\| \mathbf{rad}_t \circ (\boldsymbol{\chi}_{M'_t} - \boldsymbol{\chi}_{M_t}) \right\|_\infty \\
&= \max_{i \in (M_t \backslash M'_t) \cup (M'_t \backslash M_t)} \mathrm{rad}_t(i) \\
&\leq \max_{i \in (M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)} \mathrm{rad}_t(i)
\end{aligned}
$$

$$= \mathrm{rad}_t(e) < \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}. \tag{52}$$

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

$$\begin{aligned}
\|\mathbf{rad}_t \circ |\boldsymbol{\chi}_c|\|_\infty &= \max_{i \in c_+ \cup c_-} \mathrm{rad}_t(i) \\
&\le \max_{i \in (\tilde{M}_t \backslash M_t) \cup (M_t \backslash \tilde{M}_t)} \mathrm{rad}_t(i) \\
&= \mathrm{rad}_t(e) < \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}. 
\end{aligned} \tag{53}$$

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

$$\begin{aligned}
\langle \mathbf{rad}_t, |\boldsymbol{d}_2| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle &= \langle \mathbf{rad}_t, |\boldsymbol{d}_2| - |\boldsymbol{d}| \rangle = \left\langle \mathbf{rad}_t, \boldsymbol{d}_2^2 - \boldsymbol{d}^2 \right\rangle && (54) \\
&= \left\langle \mathbf{rad}_t, (\boldsymbol{d} - \boldsymbol{\chi}_c + \boldsymbol{\chi}_b)^2 - \boldsymbol{d}^2 \right\rangle \\
&= \left\langle \mathbf{rad}_t, \boldsymbol{\chi}_b^2 + \boldsymbol{\chi}_c^2 - 2\boldsymbol{\chi}_b \circ \boldsymbol{\chi}_c - 2\boldsymbol{d} \circ \boldsymbol{\chi}_c + 2\boldsymbol{d} \circ \boldsymbol{\chi}_b \right\rangle \\
&= \left\langle \mathbf{rad}_t, \boldsymbol{\chi}_b^2 - \boldsymbol{\chi}_c^2 + 2\boldsymbol{\chi}_b \circ (\boldsymbol{d} - \boldsymbol{\chi}_c) \right\rangle && (55) \\
&= \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - 2 \langle \mathbf{rad}_t, \boldsymbol{\chi}_b \circ (\boldsymbol{d} - \boldsymbol{\chi}_c) \rangle \\
&= \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - 2 \langle \mathbf{rad}_t \circ (\boldsymbol{d} - \boldsymbol{\chi}_c), \boldsymbol{\chi}_b \rangle && (56) \\
&\ge \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - 2 \|\mathbf{rad}_t \circ (\boldsymbol{d} - \boldsymbol{\chi}_c)\|_\infty \|\boldsymbol{\chi}_b\|_1 && (57) \\
&> \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - \frac{2\Delta_e}{3\,\mathrm{width}(\mathcal{B})} \|\boldsymbol{\chi}_b\|_1 && (58) \\
&\ge \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - \frac{2\Delta_e}{3}, && (59)
\end{aligned}$$

where Eq. (54) holds since $\boldsymbol{d} \in \{-1,0,1\}^n$ and $\boldsymbol{d}_2 \in \{-1,0,1\}^n$; Eq. (55) follows from the claim that $\boldsymbol{d} \circ \boldsymbol{\chi}_c = |\boldsymbol{\chi}_c| = \boldsymbol{\chi}_c^2$; Eq. (56) and Eq. (57) follow from Lemma 10 and Hölder's inequality; Eq. (58) follows from Eq. (52); and Eq. (59) holds since $b \in \mathcal{B}$ and $\|\boldsymbol{\chi}_b\|_1 = |b_+| + |b_-| \le \mathrm{width}(\mathcal{B})$.

Applying Lemma 11 by setting $M' = M_t' \oplus b$ and using the fact that $\tilde{w}_t(\tilde{M}_t) \ge \tilde{w}_t(M_t' \oplus b)$, we have

$$\begin{aligned}
\langle \bar{\boldsymbol{w}}_t, \boldsymbol{d} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle &= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, |\boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t}| \right\rangle \\
&= \tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) \\
&\ge \tilde{w}_t(M_t' \oplus b) - \tilde{w}_t(M_t) \\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t' \oplus b} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, |\boldsymbol{\chi}_{M_t' \oplus b} - \boldsymbol{\chi}_{M_t}| \right\rangle \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{d}_2 \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_2| \rangle \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{d} \rangle - \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_c \rangle + \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_2| \rangle,
\end{aligned}$$

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

$$
\begin{aligned}
\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_c \rangle &\geq \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_2| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle \\
&\geq \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - \frac{2\Delta_e}{3} && (60) \\
&> \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \langle \mathbf{rad}_t, \boldsymbol{\chi}_c \rangle - \frac{2\Delta_e}{3} && (61) \\
&> \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{\Delta_e}{3} - \frac{2\Delta_e}{3} && (62) \\
&= \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \Delta_e \geq 0, && (63)
\end{aligned}
$$

where Eq. (60) uses Eq. (59); Eq. (61) follows from the assumption that event $\xi_t$ occurs and Lemma 12; and Eq. (61) holds since Eq. (53).

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

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

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 8, 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 8 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 8 gives that $\boldsymbol{d}_1 = \boldsymbol{d} + \boldsymbol{b}$.

Then, we apply Lemma 9 by setting $M = M_t$ and $M' = \tilde{M}_t$. This shows that

$$
\|\mathbf{rad}_t \circ \boldsymbol{d}\|_\infty \leq \max_{i:(\tilde{M}_t \backslash M_t) \cup (M_t \backslash \tilde{M}_t)} \mathrm{rad}_t(i) = \mathrm{rad}_t(e) < \frac{\Delta_e}{3}. \tag{64}
$$

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

$$
\begin{aligned}
\langle \bar{\boldsymbol{w}}_t, \boldsymbol{d}_1 \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_1| \rangle - \langle \bar{\boldsymbol{w}}_t, \boldsymbol{d} \rangle - \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle &= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_1| - |\boldsymbol{d}| \rangle \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \left\langle \mathbf{rad}_t, \boldsymbol{d}_1^2 - \boldsymbol{d}^2 \right\rangle && (65) \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \left\langle \mathbf{rad}_t, 2\boldsymbol{d} \circ \boldsymbol{\chi}_b + \boldsymbol{\chi}_b^2 \right\rangle && (66) \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \left\langle \mathbf{rad}_t, \boldsymbol{\chi}_b^2 \right\rangle + 2 \langle \mathbf{rad}_t \circ \boldsymbol{d}, \boldsymbol{\chi}_b \rangle \\
&\geq \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - 2 \langle \mathbf{rad}_t \circ \boldsymbol{d}, \boldsymbol{\chi}_b \rangle && (67) \\
&\geq \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - 2 \|\mathbf{rad}_t \circ \boldsymbol{d}\|_\infty \|\boldsymbol{\chi}_b\|_1 && (68) \\
&> \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{2\Delta_e}{3} && (69) \\
&\geq 0, && (70)
\end{aligned}
$$

where Eq. (65) follows from the fact that $\boldsymbol{d}_1 \in \{-1, 0, 1\}^n$ and $\boldsymbol{d} \in \{-1, 0, 1\}^n$; Eq. (66) holds since $\boldsymbol{d}_1 = \boldsymbol{d} + \boldsymbol{\chi}_b$; Eq. (67) follows from the assumption that $\xi_t$ occurs and Lemma 12; Eq. (68) follows from Lemma 10 and Hölder's inequality; and Eq. (69) is due to Eq. (64).

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 11 shows that

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

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

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

□

### A.4 Proof of Theorem 1

Theorem 1 is now a straightforward corollary of Lemma 14 and Lemma 15. For the readers' convenience, we first restate Theorem 1 in the following.

**Theorem 1.** *Given any $\delta \in (0,1)$, any $\mathcal{M} \subseteq 2^{[n]}$ and any $\boldsymbol{w} \in \mathbb{R}^n$. Assume that the reward distribution $\varphi_e$ for each arm $e \in [n]$ is $R$-sub-Gaussian with mean $w(e)$. Set $\mathrm{rad}_t(e) = R\sqrt{\frac{2\log\left(\frac{4nt^2}{\delta}\right)}{T_e(t)}}$ 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*

$$T \leq O\left(R^2 \,\mathrm{width}(\mathcal{M})^2 \mathbf{H} \log\left(R^2 \,\mathrm{width}(\mathcal{M})^2 \mathbf{H} \cdot n/\delta\right)\right), \tag{4}$$

*where $T$ denotes the number of samples used by Algorithm 1 and $\mathbf{H}$ is defined in Eq. (2).*

*Proof.* Lemma 13 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 14 and the assumption on $\xi$, 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 15, we see that $\mathrm{rad}_{t_e}(e) \geq \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{M})}$. By plugging the definition of $\mathrm{rad}_{t_e}$, we have

$$\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{M})} \leq R\sqrt{\frac{2\log\left(4nt_e^2/\delta\right)}{T_e - 1}} \leq R\sqrt{\frac{2\log\left(4nT^2/\delta\right)}{T_e - 1}}. \tag{71}$$

Solving Eq. (71) for $T_e$, we obtain

$$T_e \leq \frac{18\,\mathrm{width}(\mathcal{M})^2 R^2}{\Delta_e^2} \log(4nT^2/\delta) + 1. \tag{72}$$

Notice that $T = \sum_{i \in [n]} T_i$. Hence the theorem follows by summing up Eq. (72) for all $e \in [n]$ and solving for $T$. □

## B Proof of Lower Bound

**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 $\varphi_e$ is given by $\varphi_e = \mathcal{N}(w(e), 1)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with mean $\mu$ and variance $\sigma^2$. Then, for any $\delta \in (0, e^{-16}/4)$ and any $\delta$-correct algorithm $\mathbb{A}$, we have*

$$\mathbb{E}[T] \geq \frac{1}{16}\mathbf{H}\log\left(\frac{1}{4\delta}\right), \tag{5}$$

*where $T$ denote the number of total samples used by algorithm $\mathbb{A}$ and $\mathbf{H}$ is defined in Eq. (2).*

*Proof.* Fix $\delta > 0$, $\boldsymbol{w} = \big(w(1), \ldots, w(n)\big)^T$ and a $\delta$-correct algorithm $\mathbb{A}$. 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

$$\mathbb{E}[T_e] \geq \frac{1}{16\Delta_e^2} \log(1/4\delta). \tag{73}$$

Notice that the theorem follows immediately by summing up Eq. (73) for all $e \in [n]$.

Fix an arm $e \in [n]$. We now focus on proving Eq. (73). 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

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

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 $M_e$ be the "next-to-optimal" set as follows

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

By definition of $\Delta_e$ in Eq. (1), 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, we have

$$\begin{aligned} w_1(M_*) - w_1(M_e) &= w(M_*) - w(M_e) - 2\Delta_e \\ &= -\Delta_e < 0. \end{aligned}$$

This means that under $H_1$, the set $M_*$ is not the optimal set.

Define $\theta = 4\delta$. Define

$$t_e^* = \frac{1}{16\Delta_e^2} \log\left(\frac{1}{\theta}\right). \tag{74}$$

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

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

$$\begin{aligned} \Pr_0[T_e > 4t_e^*] &\leq \frac{\mathbb{E}_0[T_e]}{4t_e^*} \\ &= \frac{t_e^*}{4t_e^*} = \frac{1}{4}. \end{aligned}$$

Let $X_1, \ldots, X_{T_e}$ denote the sequence of reward outcomes of arm $e$. For all $t > 0$, we define $K_t = \sum_{i \in [t]} X_i$ as the sum of outcomes of arm $e$ up to round $t$. Next, we define the event

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

We now show that $\Pr_0[\mathcal{C}] \geq 3/4$. First, notice that $\{K_t - t \cdot w(e)\}_{t=1,\ldots}$ is a martingale under $H_0$. Then, by Kolmogorov's inequality, we have

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

where the second inequality follows from the fact that the variance of $\varphi_e$ equals to 1 and therefore $\mathbb{E}_0[(K_{4t_e^*} - 4w(e)t_e^*)^2] = 4t_e^*$; 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.

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

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

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

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

where $p_l$ is the probability density function under hypothesis $H_l$. Let $K$ be the shorthand of $K_{T_e}$.

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 $w(e) + 2\Delta_e$ and variance 1. Recall that the probability density function of a Gaussian distribution with mean $\mu$ and variance $\sigma^2$ is given by $\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. Hence, we have

$$\begin{aligned}
\frac{L_1(W)}{L_0(W)} &= \prod_{i=1}^{T_e} \exp\left(\frac{-(X_i - w(e) - 2\Delta_e)^2 + (X_i - w(e))^2}{2}\right) \\
&= \prod_{i=1}^{T_e} \exp\left(\Delta_e(2X_i - 2w(e)) - 2\Delta_e^2\right) \\
&= \exp\left(\Delta_e(2K - 2w(e)T_e) - 2\Delta_e^2 T_e\right) \\
&= \exp\left(\Delta_e(2K - 2w(e)T_e)\right) \exp(-2\Delta_e^2 T_e). \qquad (75)
\end{aligned}$$

Next, we bound each individual term on the right-hand side of Eq. (75). We begin with bounding the second term of Eq. (75)

$$\begin{aligned}
\exp(-2\Delta_e^2 T_e) &\geq \exp(-8\Delta_e^2 t_e^*) \qquad (76) \\
&= \exp\left(-\frac{8}{16}\log(1/\theta)\right) \qquad (77) \\
&= \theta^{1/2}, \qquad (78)
\end{aligned}$$

where Eq. (76) follows from the assumption that event $\mathcal{S}$ occurred, which implies that event $\mathcal{A}$ occurred and therefore $T_e \leq 4t_e^*$; Eq. (77) follows from the definition of $t_e^*$.

Then, we bound the first term on the right-hand side of Eq. (75) as follows

$$\begin{aligned}
\exp\left(\Delta_e(2K - 2w(e)T_e)\right) &\geq \exp\left(-2\Delta_e\sqrt{t_e^* \log(1/\theta)}\right) \qquad (79) \\
&= \exp\left(-\frac{2}{4}\log(1/\theta)\right) \qquad (80) \\
&= \theta^{1/2}, \qquad (81)
\end{aligned}$$

where Eq. (79) 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. (80) follows from the definition of $t_e^*$.

Combining Eq. (78) and Eq. (81), we can bound $L_1(W)/L_0(W)$ for this case as follows

$$\frac{L_1(W)}{L_0(W)} \geq \theta. \qquad (82)$$

*(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 $w(e) - 2\Delta$. Therefore, the likelihood ratio $L_1(W)/L_0(W)$ is given by

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

$$
\begin{aligned}
&= \prod_{i=1}^{T_e} \exp\left(\Delta_e(2w(e) - 2X_i) - 2\Delta_e^2\right) \\
&= \exp\left(\Delta_e(2w(e)T_e - 2K)\right)\exp(-2\Delta_e^2 T_e). \qquad (83)
\end{aligned}
$$

Notice that the right-hand side of Eq. (83) differs from Eq. (75) only in its first term. Now, we bound the first term as follows

$$
\begin{aligned}
\exp\left(\Delta_e(2w(e)T_e - 2K)\right) &\geq \exp\left(-2\Delta_e\sqrt{t_e^* \log(1/\theta)}\right) \qquad (84) \\
&= \exp\left(-\frac{2}{4}\log(1/\theta)\right) \qquad (85) \\
&= \theta^{1/2}, \qquad (86)
\end{aligned}
$$

where the inequalities hold due to reasons similar to Case (1): Eq. (84) 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. (85) follows from the definition of $t_e^*$.

Combining Eq. (78) and Eq. (81), we can obtain the same bound of $L_1(W)/L_0(W)$ as in Eq. (82), 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. (82) holds, i.e. $\frac{L_1(W)}{L_0(W)} \geq \theta$. Hence, we have

$$
\begin{aligned}
\frac{L_1(W)}{L_0(W)} &\geq \theta \\
&= 4\delta. \qquad (87)
\end{aligned}
$$

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

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

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

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

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

### B.1 Exchange set size dependent lower bound

We show that, for any arm $e \in [n]$, there exists an exchange set $b$ which contains $e$ such that a $\delta$-correct algorithm must spend $\tilde{\Omega}\left(\left(|b_+| + |b_-|\right)^2/\Delta_e^2\right)$ samples on the arms belonging to $b$. This result is formalized in the following theorem.

**Theorem 6.** *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 $\varphi_e$ is given by $\varphi_e = \mathcal{N}(w(e), 1)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with mean $\mu$ and variance $\sigma^2$. Fix any $\delta \in (0, e^{-16}/4)$ and any $\delta$-correct algorithm $\mathbb{A}$.*

*Then, for any $e \in [n]$, there exists an exchange set $b = (b_+, b_-)$, such that $e \in b_+ \cup b_-$ and*

$$\mathbb{E}\left[\sum_{i \in b_+ \cup b_-} T_i\right] \geq \frac{(|b_+| + |b_-|)^2}{32\Delta_e^2} \log(1/4\delta),$$

*where $T_i$ is the number of samples of arm $i$.*

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

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

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

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

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

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

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

Since $b \in \mathcal{B}_{\mathsf{opt}}$, 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 $t^*_{b_-} = \frac{|b_-|^2}{16w(b)^2} \log(1/\theta)$.

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

$$\mathcal{B} = \{\mathsf{Out} = M_*\}. \tag{88}$$

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

$$\Pr_0[T_{b_-} > 4t^*_{b_-}] \leq \frac{\mathbb{E}_0[T_{b_-}]}{4t^*_{b_-}}$$

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

Next, we show that $\Pr_0[\mathcal{C}_1] \geq 3/4$. Notice that the sequence $\left\{ \sum_{i=1}^{t} X_i - \sum_{i=1}^{t} p_{I_i} \right\}_{t \in [4t^*_{b_-}]}$ is a martingale. Hence, by Kolmogorov's inequality, we have

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

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

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

Now, we show that if $\mathbb{E}_0[T_{b_-}] \leq t^*_{b_-}$, then $\Pr_1[\mathcal{B}] \geq \delta$. Assume that the event $\mathcal{S}_1$ occurred. We bound the likelihood ratio $L_1(W)/L_0(W)$ under this assumption as follows

$$\begin{aligned}
\frac{L_1(W)}{L_0(W)} &= \prod_{i=1}^{T_{b_-}} \exp\left(\frac{-\left(X_i - w(I_i) - \frac{2w(b)}{|b_-|}\right)^2 + (X_i - w(I_i)^2}{2}\right) \\
&= \prod_{i=1}^{T_{b_-}} \exp\left(\frac{w(b)}{|b_-|}(2X_i - 2w(I_i)) - \frac{2w(b)^2}{|b_-|^2}\right) \\
&= \exp\left(\frac{w(b)}{|b_-|}\left(\sum_{i=1}^{T_{b_-}} 2X_i - 2w(I_i)\right) - \frac{2w(b)^2}{|b_-|^2} T_{b_-}\right) \\
&= \exp\left(\frac{w(b)}{|b_-|}\left(\sum_{i=1}^{T_{b_-}} 2X_i - 2w(I_i)\right)\right) \exp\left(-\frac{2w(b)^2}{|b_-|^2} T_{b_-}\right). \qquad (89)
\end{aligned}$$

Then, we bound each term on the right-hand side of Eq. (89). First, we bound the second term of Eq. (89).

$$\begin{aligned}
\exp\left(-\frac{2w(b)^2}{|b_-|^2} T_{b_-}\right) &\geq \exp\left(-\frac{2w(b)^2}{|b_-|^2} 4t^*_b\right) \qquad (90) \\
&= \exp\left(-\frac{8}{16} \log(1/\theta)\right) \qquad (91) \\
&= \theta^{1/2}, \qquad (92)
\end{aligned}$$

where Eq. (90) follows from the assumption that events $\mathcal{S}_1$ and $\mathcal{A}_1$ occurred and therefore $T_{b_-} \leq 4t^*_{b_-}$; Eq. (91) follows from the definition of $t^*_{b_-}$. Next, we bound the first term of Eq. (89) as follows

$$\exp\left(\frac{w(b)}{|b_-|}\left(\sum_{i=1}^{T_{b_-}} 2X_i - 2w(I_i)\right)\right) \geq \exp\left(-\frac{2w(b)}{|b_-|}\sqrt{t^*_b \log(1/\theta)}\right) \qquad (93)$$

$$
\begin{aligned}
&= \exp\left(-\frac{2}{4}\log(1/\theta)\right) &(94)\\
&= \theta^{1/2}, &(95)
\end{aligned}
$$

where Eq. (93) follows since event $\mathcal{S}_1$ and $\mathcal{C}_1$ occurred and therefore $|2K - 2p_e T_e| \leq \sqrt{t_e^* \log(1/\theta)}$; Eq. (94) follows from the definition of $t_{b_-}^*$.

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

$$
\frac{L_1(W)}{L_0(W)} \geq \theta = 4\delta. \tag{96}
$$

Let $1_{S_1}$ denote the indicator variable of event $\mathcal{S}_1$. Then, we have $\frac{L_1(W)}{L_0(W)} 1_{S_1} \geq 4\delta 1_{S_1}$. Therefore, we can bound $\Pr_1[\mathcal{B}]$ as follows

$$
\begin{aligned}
\Pr_1[\mathcal{B}] \geq \Pr_1[\mathcal{S}_1] &= \mathbb{E}_1[1_{S_1}] \\
&= \mathbb{E}_0\left[\frac{L_1(W)}{L_0(W)} 1_{S_1}\right] \\
&\geq 4\delta \mathbb{E}_0[1_{S_1}] \\
&= 4\delta \Pr_0[\mathcal{S}_1] > \delta. &(97)
\end{aligned}
$$

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

$$
\mathbb{E}_0[T_{b_-}] \geq t_{b_-}^* = \frac{|b_-|^2}{16 w(b)^2} \log(1/4\delta). \tag{98}
$$

*(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. (88). 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

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

$$
\begin{aligned}
&= \exp\left(\frac{w(b)}{|b_+|}\left(\sum_{i=1}^{T_{b_+}} 2w(J_i) - 2Y_i\right) - \frac{2w(b)^2}{|b_+|^2}T_{b_+}\right) \\
&= \exp\left(\frac{w(b)}{|b_+|}\left(\sum_{i=1}^{T_{b_+}} 2w(J_i) - 2Y_i\right)\right)\exp\left(-\frac{2w(b)^2}{|b_+|^2}T_{b_+}\right) \\
&\geq \theta \qquad (99)\\
&= 4\delta,
\end{aligned}
$$

where Eq. (99) 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. (97), 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 $\delta$, which is not the optimal set under $H_2$. This contradicts to the assumption that algorithm $\mathbb{A}$ is a $\delta$-correct algorithm. In sum, we have proved that

$$
\mathbb{E}_0[T_{b_+}] \geq t^*_{b_+} = \frac{|b_+|^2}{16w(b)^2}\log(1/4\delta). \qquad (100)
$$

*(End of Part (2))*

Finally, we combine the results from both parts, i.e. Eq. (98) and Eq. (100). We obtain

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

□

# C Proof of Extension Results

## C.1 Fixed Budget Setting (Theorem 3)

In this part, we analyze the probability of error of the modified CGapExp algorithm in the fixed budget setting and prove Theorem 3. First, we prove a lemma which characterizes the confidence intervals constructed in Theorem 3.

**Lemma 16.** *Fix parameter $\alpha > 0$ and the number of rounds $T > 0$. Assume that the reward distribution $\varphi_e$ 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 $t \in [T]$. □

Then, Theorem 3 can be obtained from the key lemmas (Lemma 14 and Lemma 15) and Lemma 16.

**Theorem 3.** *Use the same notations as in Theorem 1. Given $T > n$ and parameter $\alpha > 0$, set the confidence radius $\mathrm{rad}_t(e) = R\sqrt{\frac{\alpha}{T_e(t)}}$ for all arms $e \in [n]$ and all $t > 0$. Run* CGapExp *algorithm for at most $T$ rounds. Then, for $0 \le \alpha \le \frac{1}{9}(T-n)\left(R^2\,\mathrm{width}(\mathcal{M})^2\mathbf{H}\right)^{-1}$, we have*

$$\Pr\left[\mathsf{Out} \neq M_*\right] \le 2Tn\exp\left(-2\alpha\right). \tag{6}$$

*Proof.* Define random event $\xi = \bigcap_{t=1}^{T} \xi_t$. By Lemma 16, we see that $\Pr[\xi] \ge 1 - 2nT\exp(-2\alpha)$. In the rest of the proof, we assume that $\xi$ happens.

Let $T^*$ denote the round that the algorithm stops. We claim that the algorithm $T^* < T$. If the claim is true, then the algorithm stops since it meets the stopping condition on round $T^*$. Hence $\tilde{M}_{T^*} = M_{T^*}$ and $\mathsf{Out} = M_{T^*}$. By assumption on $\xi$ and Lemma 14, 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. Let $t_e$ be the last round that arm $e$ is pulled. Hence $T_e(t_e) = T_e - 1$. By Lemma 15, we see that $\mathrm{rad}_{t_e}(e) \ge \frac{\Delta}{3\,\mathrm{width}(\mathcal{B})}$. Now plugging in the definition of $\mathrm{rad}_{t_e}(e)$, we have

$$\begin{aligned}\frac{\Delta}{3\,\mathrm{width}(\mathcal{B})} &\le \mathrm{rad}_{t_e}(e)\\ &= R\sqrt{\frac{\alpha}{T_e(t_e)}} = R\sqrt{\frac{\alpha}{T_e - 1}}.\end{aligned}$$

Hence we have

$$T_e \le \frac{9R^2\,\mathrm{width}(\mathcal{B})^2}{\Delta_e^2} \cdot \alpha + 1. \tag{101}$$

By summing up Eq. (101) for all $e \in [n]$, we have

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

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

### C.2 PAC Learning (Theorem 4)

First, we prove a $(\epsilon,\delta)$-PAC counterpart of Lemma 14.

**Lemma 17.** *If* CGapExpPAC *stops on round $t$ and suppose that event $\xi_t$ occurs. Then, we have $w(M_*) - w(\mathsf{Out}) \le \epsilon$.*

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

$$\begin{aligned}\epsilon &\ge \tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t)\\ &\ge \tilde{w}_t(M_*) - \tilde{w}_t(M_t) \qquad (102)\\ &= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_*} - \boldsymbol{\chi}_{M_t}\right\rangle + \left\langle \mathbf{rad}_t, |\boldsymbol{\chi}_{M_*} - \boldsymbol{\chi}_{M_t}|\right\rangle \qquad (103)\\ &\ge \left\langle \boldsymbol{w}, \boldsymbol{\chi}_{M_*} - \boldsymbol{\chi}_{M_t}\right\rangle \qquad (104)\\ &= w(M_*) - w(M_t),\end{aligned}$$

where Eq. (102) follows from the definition of $\tilde{M}_t \triangleq \arg\max_{M\in\mathcal{M}} \tilde{w}_t(M)$; Eq. (103) follows from Lemma 11; Eq. (104) follows from the assumption that $\xi_t$ occurs and Lemma 12. □

The next lemma generalizes Lemma 15. It shows that, with high probability, each arm $e \in [n]$ will not be played on round $t$ if $\mathrm{rad}_t(e) \le \max\left\{\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{M})}, \frac{\epsilon}{2K}\right\}$.

**Lemma 18.** *Let $K = \max_{M \in \mathcal{M}} |M|$. For any arm $e \in [n]$ and any round $t > n$ after initialization, if $\mathrm{rad}_t(e) \le \max\left\{\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{M})}, \frac{\epsilon}{2K}\right\}$, then arm $e$ will not be played on round $t$, i.e. $p_t \neq e$.*

*Proof.* If $\mathrm{rad}_t(e) \le \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{M})}$, then we can apply Lemma 15 which immediately gives that $p_t \neq e$. Hence, we only need to prove the case that $\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{M})} \le \mathrm{rad}_t(e) \le \frac{\epsilon}{2K}$.

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 $\mathrm{rad}_t(i) \le \mathrm{rad}_t(e) \le \frac{\epsilon}{2K}$. By summing up this inequality for all $i \in (M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)$, we have

$$\begin{aligned} \epsilon &\ge \sum_{i \in (M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)} \mathrm{rad}_t(i) && (105) \\ &= \left\langle \mathbf{rad}_t, \left|\chi_{M_t} - \chi_{\tilde{M}_t}\right| \right\rangle, && (106) \end{aligned}$$

where Eq. (105) follows from the fact that $|(M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)| \le |M_t| + |\tilde{M}_t| \le 2K$; and Eq. (106) 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, we have

$$\begin{aligned} \tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) &= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, |\boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t}| \right\rangle && (107) \\ &\le \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t} \right\rangle + \epsilon && (108) \\ &= \bar{w}_t(\tilde{M}_t) - \bar{w}_t(M_t) + \epsilon \\ &\le \epsilon, && (109) \end{aligned}$$

where Eq. (107) follows from Lemma 11; Eq. (108) uses Eq. (106); and Eq. (109) follows from $\bar{w}_t(M_t) \ge \bar{w}_t(\tilde{M}_t)$.

Therefore, we see that $\tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) \le \epsilon$. By the stopping condition of CGapExpPAC, the algorithm must terminate on round $t$. This contradicts to the assumption that $p_t = e$. □

Using Lemma 18 and Lemma 17, we are ready to prove Theorem 4.

**Theorem 4.** *Use the same notations as in Theorem 1. Fix $\delta \in (0, 1)$ and $\epsilon \ge 0$. Then, with probability at least $1 - \delta$, the output $\mathsf{Out}$ of CGapExpPAC satisfies $w(M_*) - w(\mathsf{Out}) \le \epsilon$. In addition, the number of samples $T$ used by the algorithm satisfies*

$$T \le O\left( R^2 \sum_{e \in [n]} \min\left\{ \frac{\mathrm{width}(\mathcal{M})^2}{\Delta_e^2}, \frac{K^2}{\epsilon^2} \right\} \log\left( \frac{R^2 n}{\delta} \sum_{e \in [n]} \min\left\{ \frac{\mathrm{width}(\mathcal{M})^2}{\Delta_e^2}, \frac{K^2}{\epsilon^2} \right\} \right) \right), \quad (7)$$

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

*Proof.* Similar to the proof of Theorem 1, we appeal to Lemma 13, 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 $\xi$ occurs in the rest of the proof.

By the assumption of $\xi$ and Lemma 17, 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 18, we see that $\mathrm{rad}_{t_e}(e) \ge \min\{\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}, \frac{\epsilon}{2K}\}$. Then, by the construction of $\mathrm{rad}_{t_e}(e)$, we have

$$\min\left\{ \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}, \frac{\epsilon}{2K} \right\} \le R\sqrt{\frac{2\log\left(4nt_e^2/\delta\right)}{T_e - 1}} \le R\sqrt{\frac{2\log\left(4nT^2/\delta\right)}{T_e - 1}}. \quad (110)$$

Solving Eq. (110) for $T_e$, we obtain

$$T_e \leq R^2 \min\left\{\frac{18\,\mathrm{width}(\mathcal{B})^2}{\Delta_e^2}, \frac{16K^2}{\epsilon^2}\right\}\log(4nT^2/\delta) + 1. \tag{111}$$

Notice that $T = \sum_{i\in[n]} T_i$. Hence the theorem follows by summing up Eq. (111) for all $e \in [n]$ and solving for $T$. □

# D Technical Lemmas

**Lemma 19** (Basis exchange property)**.** *AA*

**Lemma 20** (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*

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

# References


[1] Séebastian Bubeck, Tengyao Wang, and Nitin Viswanathan. Multiple identifications in multi-armed bandits. In *ICML*, pages 258–265, 2013.

[2] Victor Gabillon, Mohammad Ghavamzadeh, Alessandro Lazaric, and Team SequeL. Best arm identification: A unified approach to fixed budget and fixed confidence. In *NIPS*, pages 3221–3229, 2012.

[3] Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In *ICML*, pages 655–662, 2012.

[4] Yuan Zhou, Xi Chen, and Jian Li. Optimal pac multiple arm identification with applications to crowdsourcing. In *ICML*, 2014.