Combinatorial Pure Exploration of Multi-Armed Bandits

Anonymous Author(s)
Affiliation
Address
email

Abstract

We study *combinatorial pure exploration (CPE)* problem in the stochastic multi-armed bandit setting, where a learner explores a set of arms with the objective of identifying the optimal set of arms with certain combinatorial structures, such as size-K subsets, matchings, spanning trees, paths, etc. A collection of subsets of arms with certain combinatorial structure is referred to as a *decision class*. Instead of solving each CPE task in an ad-hoc way, we provide a general framework as well as solutions that work for all decision classes which admit offline maximization oracles. In particular, we present two algorithms for the general CPE task: one for the fixed confidence setting and one for the fixed budget setting. We prove problem-dependent upper bounds of our algorithms. Our analysis exploits the combinatorial structures of the decision classes and introduces a new analytic tool. We also establish a general problem-dependent lower bound for the CPE problem. Our results show that the proposed algorithms achieve optimal sample complexity (within logarithmic factors) for many decision classes. In addition, applying our results back to problems of top-K arms identification and multiple bandit best arms identification, we recover the best known upper bounds and settles two open conjectures on the lower bounds.

1 Introduction

Multi-armed bandit (MAB) is a predominant model for characterizing the tradeoff between exploration and exploitation in decision-making problems. Although this tradeoff is intrinsic in many tasks, some application domains prefer a dedicated exploration procedure in which the goal is to identify an optimal object among a collection of candidates and the reward or loss incurred during exploration is irrelevant. In light of these applications, the related learning problem, called pure exploration in MABs, has received much attention. Recent advances in pure exploration MABs have found potential applications in many domains including crowdsourcing, communication network and online advertising.

In many of these application domains, a recurring problem is to identify an optimal object with certain *combinatorial structures*. For example, a crowdsourcing application may want to find the best assignment from workers to tasks such that overall productivity of workers are maximized. A network routing system during the initialization phase may try to build a spanning tree that minimizes the delay of links, or attempts to identify the shortest path between two sites. An online advertising system may be interested to find the best matching between ads and display slots. The literature of pure exploration MAB problems lacks a framework that encompasses these kinds of problems where the object of interest has a non-trivial combinatorial structure. Our paper contributes such a framework which accounts for general combinatorial structures, and develops a series of results, including algorithms, upper bounds and lower bounds for the framework.

In this paper, we formulate the *combinatorial pure exploration (CPE)* problem for stochastic multi-armed bandits. In the CPE problem, a learner has a fixed set of arms and each arm is associated with an unknown reward distribution. The learner is also given a collection of sets of arms called *decision class*, which satisfy certain combinatorial constraints. During the exploration period, in each round the learner chooses an arm to play and observes a random reward sampled from the associated distribution. The objective is when the exploration period ends, the learner output a member of the decision class that she believes to be optimal, in the sense that the sum of expected rewards of all arms in the output set is maximized among all members in the decision class.

The CPE framework represents a rich class of pure exploration problems. The conventional pure exploration problem in MAB, where the objective is to find the single best arm, clearly fits into this framework, in which the decision class is the collection of all singletons. This framework also naturally encompasses several recent extensions, including the problem of finding the top K arms (henceforth TopK) [16, 17, 8, 23] and the multi-bandit problem of finding the best arms simultaneously from several disjoint sets of arms (henceforth MB) [11, 8]. Furthermore, this framework also covers many more interesting cases where the decision classes correspond to collections of non-trivial combinatorial structures. For example, suppose that the arms represent the edges in a graph. Then a decision class could be the set of all paths between two vertices, all spanning trees or all matchings of the graph. And, in these cases, the objectives of CPE correspond to identifying the optimal paths, spanning trees and matchings, respectively. As we have explained earlier, these types of combinatorial pure exploration problems admit potential applications in diverse domains.

The generality of CPE framework raises several interesting challenges to the design and analysis of pure exploration algorithms. A clear challenge is that, instead of solving each CPE task in an ad-hoc way, one requires a unified algorithm and analysis that supports different decision classes. Another challenge is that the arms with the largest mean rewards may not belong to the optimal set. For example, consider the case where the decision class is the set of all matchings in a bipartite graph. In this case, a matching consisting of edges with relatively small weights may turn out to be optimal. However, in many existing algorithms for pure exploration MABs, arms would no longer be considered once their expected reward are proven suboptimal during the learning process. Therefore, the design and analysis of algorithms for CPE demands different techniques which take both rewards and structures into account.

Our results. In this paper, we propose two general learning algorithms for the CPE problem: one for the fixed confidence setting and one for the fixed budget setting. Both algorithms support a wide range of decision classes in a unified way. In the fixed confidence setting, we present Combinatorial Gap Exploration (CLUCB) algorithm. The CLUCB algorithm does not need to know the definition of the decision class, as long as it has access to the decision class through a maximization oracle. We upper bound the number of samples used by CLUCB. This sample complexity bound depends on both expected reward and the structure of decision class. When specializing our result into TopK and MB, we recover previous sample complexity bounds due to Kalyanakrishnan et al. [17] and Gabillon et al. [12]. While for other decision classes in general, our result establishes the first sample complexity upper bound. Our analysis relies on a novel combinatorial construction called *exchange class* which we believe may be of independent interest for other combinatorial optimization problems. We further show that CLUCB can be easily extended to the fixed budget setting and PAC learning setting and we provide related theoretical guarantees in the supplementary material.

Moreover, we prove a problem-dependent sample complexity lower bound for the CPE problem. Our lower bound shows that the sample complexity of the proposed CLUCB algorithm is optimal (to within logarithmic factors) for a large class of decision classes, including TopK, MB and the decision classes derived from matroids (e.g. spanning tree). Therefore our upper and lower bounds provide a near full characterization of the sample complexity of these CPE problems. For more general decision classes, our results shows that the upper and lower bounds are within a relatively benign factor. To the best of our knowledge, there are few problem-dependent lower bounds known for pure exploration MABs besides the case of identifying the single best arm [19, 2]. We also notice that our result resolves the open conjectures of Kalyanakrishnan et al. [17] and Bubeck et al. [8] on the sample complexity lower bounds of TopK and MB problems.

In the fixed budget setting, we present a parameter-free algorithm called Combinatorial Gap-based Elimination (CSAR) algorithm. We prove a probability of error bound of the CSAR algorithm. This bound can be shown to be equivalent to the sample complexity bound of CLUCB within logarithmic factors, although the two algorithms are based on quite different techniques. Our analysis of CSAR re-uses exchange classes as tools. This suggests that exchange class may be useful for the analysis of similar problems. In addition, when applying the algorithm to back TopK and MB, our bound recovers a recent result due to Bubeck et al. [8].

Useful notations. Let $[n]$ denote the set $\{1, \ldots, n\}$. Suppose that $\boldsymbol{w} \in \mathbb{R}^n$ is a vector and $E \subseteq [n]$ is a set. Let $w(i)$ denote the i-th entry of $\boldsymbol{w}$. We define $w(E) \triangleq \sum_{i \in E} w(i)$ to be the sum of entries indexed by E. Furthermore, we will use the convention that $\max_{M \in \emptyset} f(M) = -\infty$ for any real-valued function f.

2 Problem Formulation

In this section, we formally define the CPE problem. Suppose that there are n arms and the arms are numbered $1, 2, \ldots, n$. Assume that each arm $e \in [n]$ is associated with a reward distribution φ_e. Let $\boldsymbol{w} = \big(w(1), \ldots, w(n)\big)^T$ denote the vector of expected rewards, where each entry $w(e) = \mathbb{E}_{X \sim \varphi_e}[X]$ denote the expected reward of arm e. Following standard assumptions of stochastic MABs, we assume that all reward distributions have R-sub-Gaussian tails for some known constant $R > 0$. Formally, if X is a random variable drawn according to φ_e, then, for all $t \in \mathbb{R}$, one has $\mathbb{E}\big[\exp(tX - t\mathbb{E}[X])\big] \leq \exp(R^2t^2/2)$ and $\mathbb{E}\big[\exp(t\mathbb{E}[X] - tX)\big] \leq \exp(R^2t^2/2)$. It is well-known that the family of R-sub-Gaussian tail distributions encompasses all distributions that are supported on $[0, R]$ as well as Gaussian distributions with variance R^2 (cf. [22]).

We define a *decision class* $\mathcal{M} \subseteq 2^{[n]}$ as a collection of sets of arms. Let $M_* = \arg\max_{M \in \mathcal{M}} w(M)$ denote the optimal set belonging to the decision class $\mathcal{M}$ which maximizes the sum of expected reward[1]. A learner's objective is to identify M_* from $\mathcal{M}$ by playing the following game with the stochastic environment. At the beginning of the game, the decision class $\mathcal{M}$ is revealed to the learner while the reward distributions $\{\varphi_e\}_{e \in [n]}$ are unknown to the learner. Then, the learner plays the game over a sequence of rounds; on each round t, the learner pulls an arm $p_t \in [n]$ and observes a reward sampled from the associated reward distribution φ_{p_t}. The game continues until certain stopping condition is satisfied. After the game finishes, the learner need to output a set $\mathsf{Out} \in \mathcal{M}$.

We consider two different stopping conditions of the game, which are known as *fixed confidence* setting and *fixed budget* setting in the literature. In the fixed confidence setting, the learner can stop the game at any round. The learner need to guarantee that $\Pr[\mathsf{Out} = M_*] \geq 1 - \delta$ for a given confidence parameter δ. The learner's performance is evaluated by her *sample complexity*, i.e. the number of pulls used by the learner. In the fixed budget setting, the game stops after a fixed number T of rounds, where T is given before the game. The learner tries to minimize the *probability of error*, which is formally $\Pr[\mathsf{Out} \neq M_*]$, within T rounds. In this case, the learner's performance is measured by the probability of error.

3 Algorithm, Exchange Class and Sample Complexity

In this section, we present CLUCB, a learning algorithm for the CPE problem in the fixed confidence setting, and analyze its sample complexity. En route to our sample complexity bound, we introduce the notions of exchange class and widths, which characterize the certain exchange properties of combinatorial structures. Furthermore, the CLUCB algorithm can be extended to the fixed budget and PAC learning settings. We will discuss these extensions in the supplementary material (Section B).

Oracle. We allow the CLUCB algorithm to access a *maximization oracle*. A maximization oracle takes a weight vector $\boldsymbol{v} \in \mathbb{R}^n$ as input and finds an optimal set within a given decision class $\mathcal{M}$ with respect to the weight vector $\boldsymbol{v}$. Formally, we call a function Oracle: $\mathbb{R}^n \to \mathcal{M}$ a maximization oracle for $\mathcal{M}$ 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 wide range of decision classes admit such maximization oracles, including decision classes correspond to collections of matchings, paths or bases of matroids (see later for concrete examples). Besides the access to the oracle, CLUCB does not need *any* additional knowledge of the decision class $\mathcal{M}$.

Algorithm. Now we describe the details of CLUCB. During its execution, the CLUCB 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 $|w(e) - \bar{w}_t(e)| \leq \mathrm{rad}_t(e)$ holds with high probability for each arm $e \in [n]$ and each round $t > 0$. CLUCB begins with an initialization phase in which each arm is pulled once. Then, at round $t \geq n$, CLUCB uses the following procedure to choose an arm to play. First, CLUCB calls the oracle which finds the set $M_t = \mathrm{Oracle}(\bar{\boldsymbol{w}}_t)$. The set M_t is the "best" set with respect to the empirical means $\bar{\boldsymbol{w}}_t$. Then, CLUCB explores possible refinements

[1] For convenience, we will assume that M_* is unique. We also assume that $\mathcal{M}$ is non-empty.

Algorithm 1 CLUCB: Combinatorial Lower-Upper Confidence Bound

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

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

- **for** $t = n, n+1, \ldots$ **do**
  - $M_t \leftarrow \text{Oracle}(\bar{\boldsymbol{w}}_t)$
  - Compute confidence radius $\text{rad}_t(e)$ for all $e \in [n]$ ▷ $\text{rad}_t(e)$ is defined later in Theorem 1
  - **for** $e = 1, \ldots, n$ **do**
    - **if** $e \in M_t$ **then**
      - $\tilde{w}_t(e) \leftarrow \bar{w}_t(e) - \text{rad}_t(e)$
    - **else**
      - $\tilde{w}_t(e) \leftarrow \bar{w}_t(e) + \text{rad}_t(e)$
  - $\tilde{M}_t \leftarrow \text{Oracle}(\tilde{\boldsymbol{w}}_t)$
  - **if** $\tilde{w}_t(\tilde{M}_t) = \tilde{w}_t(M_t)$ **then**
    - Out $\leftarrow M_t$
    - **return** Out
  - $p_t \leftarrow \arg\max_{e \in (\tilde{M}_t \setminus M_t) \cup (M_t \setminus \tilde{M}_t)} \text{rad}_t(e)$ ▷ Break ties arbitrarily
  - 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 $e \neq p_t$

of M_t. In particular, CLUCB uses the confidence radius to compute an adjusted expectation vector $\tilde{\boldsymbol{w}}_t$ in the following way: for each arm $e \in M_t$, $\tilde{w}_t(e)$ equals to the lower confidence bound $\tilde{w}_t(e) = \bar{w}_t(e) - \text{rad}_t(e)$; and for each arm $e \notin M_t$, $\tilde{w}_t(e)$ equals to the upper confidence bound $\tilde{w}_t(e) = \bar{w}_t(e) + \text{rad}_t(e)$. Intuitively, the adjusted expectation vector $\tilde{\boldsymbol{w}}_t$ penalizes arms belonging to the current set M_t and encourages exploring arms out of M_t. CLUCB then calls the oracle using the adjusted expectation vector $\tilde{\boldsymbol{w}}_t$ as input to compute a refined set $\tilde{M}_t = \text{Oracle}(\tilde{\boldsymbol{w}}_t)$. If $\tilde{w}_t(\tilde{M}_t) = \tilde{w}_t(M_t)$ then CLUCB stops and returns Out $= M_t$. Otherwise, CLUCB pulls the arm belonging to the symmetric difference $(\tilde{M}_t \setminus M_t) \cup (M_t \setminus \tilde{M}_t)$ between M_t and $\tilde{M}_t$ with the largest confidence radius. This ends the t-th round of CLUCB. The pseudo-code of CLUCB is shown in Algorithm 1. We note that CLUCB generalizes and unifies several different fixed confidence algorithms dedicated to the TOPK and MB problems in the literature [17, 12].

3.1 Sample complexity

Now we prove a problem-dependent sample complexity bound of the CLUCB algorithm. Our sample complexity bound depends on the combinatorial structure of the decision class $\mathcal{M}$. Therefore, to formally state our result, we need to introduce several notions to capture these structural properties.

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

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

where we use the convention that the maximum value of an empty set is $-\infty$. We also define 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 $e \notin M_*$, the gap Δ_e represents sub-optimality of the best set that includes arm e; and, for each arm $e \in M_*$, the gap Δ_e is the sub-optimality of the best set that does not include arm e. This naturally generalizes and unifies previous definitions of gaps [2, 11, 16, 8].

Exchange class and the width of a decision class. A notable challenge of our analysis stems from the generality of CLUCB which, as we have seen, supports a wide range of decision classes $\mathcal{M}$. Indeed, previous algorithms for special cases including TOPK and MB require a separate analysis for each individual type of problem. Such strategy is intractable for our setting and we need a unified analysis for all decision classes. Our solution to this challenge is a novel combinatorial construction called *exchange class*, which is used as a proxy for the structure of the decision class. Intuitively, an exchange class $\mathcal{B}$ for a decision class $\mathcal{M}$ can be seen as a collection of "patches" (borrowing concepts from software engineering) such that, for any two different sets $M, M' \in \mathcal{M}$, one can transform M to M' by applying a series of patches of $\mathcal{B}$; and each application of a patch yields a valid member of $\mathcal{M}$. These patches are later used by our analysis to build gadgets that interpolate between different members of the decision class and serve to bridge different key quantities.

Now we formally define the exchange class. We begin with the definition of exchange sets, which formalize the aforementioned "patches". 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 $M \subseteq [n]$ 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 collection 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 $(M' \ominus b) \in \mathcal{M}$. In our analysis, the key quantity associated with the exchange class is called *width*, which is defined as the size of largest exchange set as follows

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

We need one more definition. Let $\operatorname{Exchange}(\mathcal{M})$ denote the family of all possible exchange classes for $\mathcal{M}$. We define the width of a decision class $\mathcal{M}$ as the width of the thinnest exchange class

$$\operatorname{width}(\mathcal{M}) = \min_{\mathcal{B} \in \operatorname{Exchange}(\mathcal{M})} \operatorname{width}(\mathcal{B}), \tag{4}$$

where $\operatorname{width}(\mathcal{B})$ is defined in Eq. (3).

Sample complexity. Our main result of this section is a problem-dependent sample complexity bound of the CLUCB algorithm, which shows that CLUCB returns the optimal set with high probability and uses at most $\tilde{O}\big(\operatorname{width}(\mathcal{M})^2 \mathbf{H}\big)$ samples.

Theorem 1. *Given any $\delta \in (0,1)$, any decision class $\mathcal{M} \subseteq 2^{[n]}$ and any expected rewards $\boldsymbol{w} \in \mathbb{R}^n$. Assume that the reward distribution φ_e for each arm $e \in [n]$ is R-sub-Gaussian with mean $w(e)$. Set $\operatorname{rad}_t(e) = R\sqrt{2\log\left(\frac{4nt^2}{\delta}\right)/T_t(e)}$ for all $t > 0$ and $e \in [n]$. Then, with probability at least $1-\delta$, the CLUCB algorithm (Algorithm 1) returns the optimal set $\mathsf{Out} = \arg\max_{M \in \mathcal{M}} w(M)$ and*

$$T \le O\left(R^2 \operatorname{width}(\mathcal{M})^2 \mathbf{H} \log\left(R^2 \mathbf{H}/\delta\right)\right), \tag{5}$$

where T denotes the number of samples used by Algorithm 1, $\mathbf{H}$ is defined in Eq. (2) and $\operatorname{width}(\mathcal{M})$ is defined in Eq. (4).

3.2 Examples of decision classes

Now we investigate several concrete types of decision classes, which correspond to different CPE tasks. We analyze the width of these decision classes and apply Theorem 1 to obtain the sample complexity bounds. A more detailed analysis and the constructions of exchange classes can be found in the supplementary material. We begin with the problem of top-K arm identification (TopK) and multi-bandit best arms identification (MB).

Example 1 (TopK and MB)**.** *For any $K \in [n]$, the problem of finding the top K arms with the largest expected reward can be modeled by decision class $\mathcal{M}_{\text{TopK}(K)} = \{M \subseteq [n] \mid |M| = K\}$. Let $\mathcal{A} = \{A_1, \ldots, A_m\}$ be a partition of $[n]$. The problem of identifying the best arms from each group of arms $A_1, \ldots, A_m$ can be modeled by decision class $\mathcal{M}_{\text{MB}(\mathcal{A})} = \{M \subseteq [n] \mid \forall i \in [m], |M \cap A_i| = 1\}$.*

Then we have $\operatorname{width}(\mathcal{M}_{\text{TopK}(K)}) \le 2$ and $\operatorname{width}(\mathcal{M}_{\text{MB}(\mathcal{A})}) \le 2$ (see Fact 2 and 3 in the supplementary material) and therefore the sample complexity of CLUCB for solving TopK and MB is $O\big(\mathbf{H}\log(\mathbf{H}/\delta)\big)$, which matches previous results in the fixed confidence setting [17, 12].

Next we consider the problem of identifying the maximum matching or the problem of finding the shortest path in a setting where arms correspond to edges. For these problems, Theorem 1 establishes the first known sample complexity bound.

Example 2 (Matchings and Paths)**.** *Let $G(V,E)$ be a graph with n edges and assume there is a one-to-one mapping between edges E and arms $[n]$. First let us assume that G is a bipartite graph. Let $\mathcal{M}_{\text{MATCH}(G)}$ correspond to the set of all matchings in G. Then we have $\mathrm{width}(\mathcal{M}_{\text{MATCH}(G)}) \le |V|$ (see Fact 4).*

Next suppose that G is a directed acyclic graph and let $s,t \in V$ be two vertices. Let $\mathcal{M}_{\text{PATH}(G,s,t)}$ correspond to the set of all paths from s to t. Then we have $\mathrm{width}(\mathcal{M}_{\text{PATH}(G,s,t)}) \le |V|$ (see Fact 5). Therefore the sample complexity bounds of CLUCB for decision classes $\mathcal{M}_{\text{MATCH}(G)}$ and $\mathcal{M}_{\text{PATH}(G,s,t)}$ are $O\big(|V|^2\mathbf{H}\log(\mathbf{H}/\delta)\big)$.

Last, we investigate the general problem of identifying the maximum-weight basis of a matroid. Again, Theorem 1 is the first sample complexity upper bound for this pure exploration problem.

Example 3 (Matroids)**.** *Let $T = (E, \mathcal{I})$ be a finite matroid, where E is a set of size n (called ground set) and $\mathcal{I}$ is a family of subsets of E (called independent sets) which satisfies the axioms of matroids[2]. Assume that there is a one-to-one mapping between E and $[n]$. And recall that a basis of matroid T is a maximal independent set. Let $\mathcal{M}_{\text{MATROID}(T)}$ correspond to the set of all bases of T. Then we have $\mathrm{width}(\mathcal{M}_{\text{MATROID}(T)}) \le 2$ (see Fact 1) and the sample complexity of CLUCB for $\mathcal{M}_{\text{MATROID}(T)}$ is $O\big(\mathbf{H}\log(\mathbf{H}/\delta)\big)$.*

In our last example, we see that $\mathcal{M}_{\text{MATROID}(T)}$ is a general type of decision class which encompasses many pure exploration problems including TopK and MB as special cases, where TopK corresponds to uniform matroids of rank K and MB corresponds to partition matroids. It is easy to see that $\mathcal{M}_{\text{MATROID}(T)}$ also covers the decision class that contains all spanning trees of a graph. On the other hand, it is well-known that the family of matchings and paths cannot be formulated as matroids since they are matroid intersections [21].

4 Lower Bound

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

We show that, for any decision class $\mathcal{M}$ and any expected rewards $\boldsymbol{w}$, any δ-correct algorithm $\mathbb{A}$ must use at least $\Omega\big(\mathbf{H}\log(1/\delta)\big)$ samples in expectation.

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

$$\mathbb{E}[T] \ge \frac{1}{16}\mathbf{H}\log\left(\frac{1}{4\delta}\right), \tag{6}$$

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 open conjectures of Kalyanakrishnan et al. [17] and Bubeck et al. [8] that the lower bounds of sample complexity of TopK and MB problem are $\Omega\big(\mathbf{H}\log(1/\delta)\big)$. Moreover, in Example 1 and Example 3, we have shown that the sample complexity of CLUCB is $O(\mathbf{H}\log(n\mathbf{H}/\delta))$ for TopK, MB and more generally the decision classes derived from matroids $\mathcal{M}_{\text{MATROID}(T)}$ (including spanning trees). Hence, we see that the CLUCB algorithm achieves the optimal sample complexity within logarithmic factors for these pure exploration problems.

On the other hand, for general decision classes with non-constant widths, we see that there is a gap of $\tilde{\Theta}(\mathrm{width}(\mathcal{M})^2)$ between the upper bound Eq. (5) and the lower bound Eq. (6). Notice that we have $\mathrm{width}(\mathcal{M}) \le n$ for any decision class $\mathcal{M}$ and therefore the gap is relatively benign. Our

[2]The three axioms of matroid are (1) $\emptyset \in \mathcal{I}$ and $\mathcal{I} \ne \{\emptyset\}$; (2) Every subsets of an independent set are independent (heredity property); (3) For all $A, B \in \mathcal{I}$ such that $|B| = |A| + 1$ there exists an element $e \in B \backslash A$ such that $A \cup \{e\} \in \mathcal{I}$ (augmentation property). We refer interested readers to [21] for a general introduction to the matroid theory.

lower bound also suggests that the dependency on $\mathbf{H}$ of the sample complexity of CLUCB cannot be improved up to logarithmic factors. Furthermore, we conjecture that the sample complexity lower bound might inherently depend on the size of exchange sets. In the supplementary material, we provide evidence on this conjecture which is a lower bound on the sample complexity of exploration of exchange sets.

5 Fixed Budget Algorithm

In this section, we present CSAR, a parameter-free learning algorithm for the CPE problem in the fixed budget setting. Then, we upper bound the probability of error CSAR in terms of $\mathbf{H}$ and $\mathrm{width}(\mathcal{M})$.

Constrained oracle. The CSAR algorithm requires access to a *constrained oracle*, which is a function denoted as $\mathrm{COracle} : \mathbb{R}^n \times 2^{[n]} \times 2^{[n]} \to \mathcal{M} \cup \{\perp\}$ and satisfies

$$\mathrm{COracle}(\boldsymbol{v}, A, B) = \begin{cases} \arg\max_{M \in \mathcal{M}_{A,B}} v(M) & \text{if } \mathcal{M}_{A,B} \neq \emptyset \\ \perp & \text{if } \mathcal{M}_{A,B} = \emptyset, \end{cases}$$

where $\mathcal{M}_{A,B} = \{M \in \mathcal{M} \mid A \subseteq M, B \cap M = \emptyset\}$ and $\perp$ is a null symbol. Hence we see that $\mathrm{COracle}(\boldsymbol{v}, A, B)$ returns an optimal set that includes all elements of A while excluding all elements of B; and if there are no feasible sets, the oracle $\mathrm{COracle}(\boldsymbol{v}, A, B)$ returns the null symbol $\perp$. In the supplementary material, we show that constrained oracles are equivalent to maximization oracles up to a transformation on the weight vector. In addition, similar to CLUCB, CSAR does not need any additional knowledge of $\mathcal{M}$ other than accesses to a constrained oracle for $\mathcal{M}$.

Algorithm. The idea of the CSAR algorithm is as follows. The CSAR algorithm divides the budget of T rounds into n phases. In the end of each phase, CSAR 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.

Now we describe the procedure of the CSAR 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 to be active if $e \notin A_t \cup B_t$. Then, in phase t, CSAR 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, CSAR calls the constrained oracle to compute an optimal set 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, CSAR estimate the "empirical gap" of e in the following way. If $e \in M_t$, then CSAR computes an optimal set $\tilde{M}_{t,e}$ 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 CSAR 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 empirical gap of e is calculated as $\bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,e})$. Finally, CSAR chooses the arm p_t with the largest empirical gap. If $p_t \in M_t$ then p_t is accepted otherwise p_t is rejected. The pseudo-code of CSAR is shown in Algorithm 2. We note that CSAR can be considered as a generalization of the two versions of SAR algorithm due to Bubeck et al. [8], which are designed specifically for the TopK and MB problems respectively.

5.1 Probability of error

In the following theorem, we bound that the probability of error of the CSAR algorithm.

Theorem 3. *Given any $T > n$, any decision class $\mathcal{M} \subseteq 2^{[n]}$ and any expected rewards $\boldsymbol{w} \in \mathbb{R}^n$. Assume that the reward distribution φ_e for each arm $e \in [n]$ is R-sub-Gaussian with mean $w(e)$. Let $\Delta_{(1)}, \ldots, \Delta_{(n)}$ be a permutation of $\Delta_1, \ldots, \Delta_n$ (defined in Eq. (1)) such that $\Delta_{(1)} \leq \ldots\ldots \Delta_{(n)}$. Define $\mathbf{H}_2 \triangleq \max_{i \in [n]} i \Delta_{(i)}^{-2}$. Then, the CSAR algorithm uses at most T samples and outputs a solution $\mathsf{Out} \in \mathcal{M} \cup \{\perp\}$ such that*

$$\Pr[\mathsf{Out} \neq M_*] \leq n^2 \exp\left(-\frac{2(T-n)}{9R^2 \tilde{\log}(n)\, \mathrm{width}(\mathcal{M})^2 \mathbf{H}_2}\right), \tag{7}$$

where $\tilde{\log}(n) \triangleq \sum_{i=1}^n i^{-1}$, $M_* = \arg\max_{M \in \mathcal{M}} w(M)$ and $\mathrm{width}(\mathcal{M})$ is defined in Eq. (4).

Algorithm 2 CSAR: Combinatorial Successive Accept Reject

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

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 active arm
  $M_t \leftarrow \mathrm{COracle}(\bar{\boldsymbol{w}}_t, A_t, B_t)$
  if $M_t = \perp$ **then**
    fail: set Out $\leftarrow \perp$ and **return** Out
  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 $\tilde{M}_{t,e} \leftarrow \mathrm{COracle}(\bar{\boldsymbol{w}}_t, A_t \cup \{e\}, B_t)$
  $p_t \leftarrow \arg\max_{i \in [n] \backslash (A_t \cup B_t)} \bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,i})$ ▷ Define $\bar{w}_t(\perp) = -\infty$; Break ties arbitrarily.
  if $p_t \in M_t$ **then**
    $A_{t+1} \leftarrow A_t \cup \{p_t\}, B_{t+1} \leftarrow B_t$
  else
    $A_{t+1} \leftarrow A_t, B_{t+1} \leftarrow B_t \cup \{p_t\}$
Out $\leftarrow A_{n+1}$
return Out

It is well-known that $\mathbf{H}_2$ is equivalent to $\mathbf{H}$ up to a logarithmic factor: $\mathbf{H}_2 \leq \mathbf{H} \leq \log(2n)\mathbf{H}_2$ (see [2]). Therefore, by setting the probability of error (the RHS of Eq. (7)) to a constant, one can see that CSAR requires a budget of $T = \tilde{O}(\mathrm{width}(\mathcal{M})^2\mathbf{H})$ samples. This is equivalent to the sample complexity bound of CLUCB up to logarithmic factors. In addition, applying Theorem 3 back to TopK and MB, our bound matches the previous fixed budget algorithm due to Bubeck et al. [8].

6 Related Work

The multi-armed bandit problem has been extensively studied in both stochastic and adversarial settings [18, 3, 3]. We refer readers to [5] for a survey on recent advances. Many work in MABs focus on minimizing the cumulative regret, which is an objective known to be fundamentally different from the objective of pure exploration MABs [6]. Among these work, a recent line of research considers a generalized setting called combinatorial bandits in which a set of arms (satisfying certain combinatorial constraint) is played on each round [9, 10, 15, 20, 1, 7**?**]. Note that the objective of these work is to minimize the cumulative regret, which differs from ours.

In the literature of pure exploration MABs, the classical problem of identifying the single best arm has been well-studied in both fixed confidence and fixed budget settings [19, 6, 2, 12, 14, 13]. A flurry of recent work extend this classical problem to TopK and MB problems and provide algorithms and upper bounds [16, 17, 23, 8, 11, 12]. Our framework encompasses these two problems as special cases and covers a much larger class of combinatorial pure exploration problems, which are unaddressed in the current literature. Applying our results back to TopK and MB, our upper bounds match the best known problem-dependent bounds due to Gabillon et al. [12], Bubeck et al. [8] and Kalyanakrishnan et al. [17]; and our lower bound provides the first problem-dependent lower bounds for these two problems, which are conjectured earlier by Kalyanakrishnan et al. [17] and Bubeck et al. [8].

7 Conclusion

In this paper, we proposed a general framework called CPE which represents a rich class of structured pure exploration problems and admits potential applications in various domains. We have shown a number of results for the framework, including two novel learning algorithms, their related upper bounds and a novel lower bound. The proposed algorithms support a wide range of decision classes in a unifying way and our analysis introduced a novel tool called exchange class which maybe of independent interest. Our upper and lower bounds characterize the complexity of the CPE problem: the sample complexity of our algorithm is optimal (up to a logarithmic factor) for the decision classes derived from matroids (including TopK and MB), while for general decision classes, our upper and lower bounds are within a relatively benign factor.

References

[1] J.-Y. Audibert and S. Bubeck. Minimax policies for adversarial and stochastic bandits. In *COLT*, 2009.

[2] J.-Y. Audibert, S. Bubeck, and R. Munos. Best arm identification in multi-armed bandits. In *COLT*, 2010.

[3] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

[4] C. Berge. Two theorems in graph theory. *Proceedings of the National Academy of Sciences*, 43(9):842–844, 1957.

[5] S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5:1–122, 2012.

[6] S. Bubeck, R. Munos, and G. Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412:1832–1852, 2010.

[7] S. Bubeck, N. Cesa-bianchi, S. M. Kakade, S. Mannor, N. Srebro, and R. C. Williamson. Towards minimax policies for online linear optimization with bandit feedback. In *COLT*, 2012.

[8] S. Bubeck, T. Wang, and N. Viswanathan. Multiple identifications in multi-armed bandits. In *ICML*, pages 258–265, 2013.

[9] N. Cesa-Bianchi and G. Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

[10] W. Chen, Y. Wang, and Y. Yuan. Combinatorial multi-armed bandit: General framework and applications. In *Proceedings of The 30th International Conference on Machine Learning*, pages 151–159, 2013.

[11] V. Gabillon, M. Ghavamzadeh, A. Lazaric, and S. Bubeck. Multi-bandit best arm identification. In *NIPS*. 2011.

[12] V. Gabillon, M. Ghavamzadeh, and A. Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. In *NIPS*, 2012.

[13] K. Jamieson and R. Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, pages 1–6. IEEE, 2014.

[14] K. Jamieson, M. Malloy, R. Nowak, and S. Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. *COLT*, 2014.

[15] S. Kale, L. Reyzin, and R. E. Schapire. Non-stochastic bandit slate problems. In *NIPS*, pages 1054–1062, 2010.

[16] S. Kalyanakrishnan and P. Stone. Efficient selection of multiple bandit arms: Theory and practice. In *ICML*, pages 511–518, 2010.

[17] S. Kalyanakrishnan, A. Tewari, P. Auer, and P. Stone. Pac subset selection in stochastic multi-armed bandits. In *ICML*, pages 655–662, 2012.

[18] T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[19] S. Mannor and J. N. Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *The Journal of Machine Learning Research*, 5:623–648, 2004.

[20] G. Neu, A. György, and C. Szepesvári. The online loop-free stochastic shortest-path problem. In *COLT*, pages 231–243, 2010.

[21] J. G. Oxley. *Matroid theory*. Oxford university press, 2006.

[22] D. Pollard. Asymptopia. *Manuscript, Yale University, Dept. of Statist., New Haven, Connecticut*, 2000.

[23] Y. Zhou, X. Chen, and J. Li. Optimal pac multiple arm identification with applications to crowdsourcing. In *ICML*, 2014.

A Analysis of CLUCB

In this section, we analyze the sample complexity of CLUCB and prove 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 $\mathbf{rad}_t = \big(\mathrm{rad}_t(1), \ldots, \mathrm{rad}_t(n)\big)^T \in [0, +\infty)^n$ 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

Let us begin with a simple lemma that characterizes the incidence vectors of exchange sets.

Lemma 1. *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 $M_2 = M_1 \backslash b_- \cup b_+$ and $b_+ \cap b_- = \emptyset$. Therefore we see that $M_2 \backslash M_1 = b_+$ and $M_1 \backslash M_2 = b_-$. We can 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

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

□

The next lemma serves as a basic tool derived from exchange classes, which allows us to interpolate between different members of a decision class $\mathcal{M}$. It also characterizes the relationship between gaps and exchange sets.

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

Proof. We decompose our proof into two cases.

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

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

Next, if $M' = M_*$, we see that $e \notin M_*$. Let us consider the set $M_1 = \arg\max_{M' : M' \in \mathcal{M} \wedge e \in M'} w(M')$ [Tian: change to $\arg\max_{M' \in \mathcal{M} : e \in M'} w(M')$ for consistency]. Note that, by definition of gaps, one has that $w(M_*) - w(M_1) = \Delta_e$. Now we define $M_0 = M_* \ominus b$. Note that we already have that $M_0 = M_* \ominus b \in \mathcal{M}$. By combining this 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 1 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 $M_3 = \arg\max_{M': M' \in \mathcal{M} \wedge e \notin M'} w(M')$ [Tian: similarly, change to $\arg\max_{M' \in \mathcal{M}: e \notin M'} w(M')$]. By definition of Δ_e, we see that $w(M_*) - w(M_3) = \Delta_e$. Now we define $M_2 = M_* \ominus b$ and notice that $M_2 \in \mathcal{M}$. By combining with the fact that $e \notin M_2$, we obtain that $w(M_2) \leq w(M_3)$. Hence, we have $w(M_*) - w(M_2) \geq w(M_*) - w(M_3) = \Delta_e$. Similar to Case (1), applying Lemma 1 again, we have

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

$\square$

Next we state two basic lemmas that help us to convert set-theoretical arguments to linear algebraic arguments.

Lemma 3. *Let $M, M' \subseteq [n]$ be two sets. Let $\mathbf{rad}_t$ be an n-dimensional vector. Then, we have*

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

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

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

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

Lemma 4. *Let $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n$ be three vectors. Then, we have $\langle \boldsymbol{a}, \boldsymbol{b} \circ \boldsymbol{c} \rangle = \langle \boldsymbol{a} \circ \boldsymbol{b}, \boldsymbol{c} \rangle$.*

Proof. We have

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

$\square$

The next lemma characterizes the property of $\tilde{\boldsymbol{w}}_t$ which is defined in the CLUCB algorithm.

Lemma 5. *Let M_t, $\tilde{\boldsymbol{w}}_t$ and $\mathbf{rad}_t$ be defined in Algorithm 1 and Theorem 1. Let $M' \in \mathcal{M}$ be an arbitrary member of decision class. We have*

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

[Tian: Change notation from $\tilde{\mathcal{X}}$ to $\hat{\mathcal{X}}$ to make the difference more clear.]

Proof. We begin with proving the first part. [Tian: From Line 5-8 in Algorithm 1, we have $\tilde{w}_t(e) = \begin{cases} \bar{w}_t(e) - \mathrm{rad}_t(e) & e \in M_t \\ \bar{w}_t(e) + \mathrm{rad}_t(e) & e \notin M_t \end{cases}$.] It is easy to verify that $\tilde{\boldsymbol{w}}_{\boldsymbol{t}} = \bar{\boldsymbol{w}}_t + \mathbf{rad}_t \circ (\mathbf{1}_n - 2\boldsymbol{\chi}_{M_t})$. Then, we have

$$
\begin{aligned}
\left\langle \tilde{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle &= \left\langle \bar{\boldsymbol{w}}_t + \mathbf{rad}_t \circ (1 - 2\boldsymbol{\chi}_{M_t}),\ \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle \\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, (\mathbf{1}_n - 2\boldsymbol{\chi}_{M_t}) \circ (\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t}) \right\rangle && (8) \\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} - 2\boldsymbol{\chi}_{M_t} \circ \boldsymbol{\chi}_{M'} + 2\boldsymbol{\chi}_{M_t}^2 \right\rangle \\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, \boldsymbol{\chi}_{M'}^2 - \boldsymbol{\chi}_{M_t}^2 - 2\boldsymbol{\chi}_{M_t} \circ \boldsymbol{\chi}_{M'} + 2\boldsymbol{\chi}_{M_t}^2 \right\rangle && (9) \\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, (\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t})^2 \right\rangle \\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, \left|\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t}\right| \right\rangle, && (10)
\end{aligned}
$$

where Eq. (8) follows from Lemma 4; Eq. (9) holds since $\boldsymbol{\chi}_{M'} \in \{0,1\}^n$ and $\boldsymbol{\chi}_{M_t} \in \{0,1\}^n$ and therefore $\boldsymbol{\chi}_{M'} = \boldsymbol{\chi}_{M'}^2$ and $\boldsymbol{\chi}_{M_t} = \boldsymbol{\chi}_{M_t}^2$; and Eq. (10) follows since $\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t} \in \{-1,0,1\}^n$. ☐

A.2 Confidence Intervals

First, we recall a standard concentration inequality of sub-Gaussian random variables.

Lemma 6 (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*

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

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

$$
\xi_t = \Big\{\forall i \in [n], \quad |w(i) - \bar{w}_t(i)| \leq \mathrm{rad}_t(i)\Big\}. \tag{11}
$$

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

If the confidence bounds are valid, we can generalize Eq. (11) to inner products as follows.

Lemma 7. *Given any $t > 0$, assume that event ξ_t as defined in Eq. (11) occurs. Then, for any vector $\boldsymbol{a} \in \mathbb{R}^n$, we have*

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

Proof. Suppose that ξ_t occurs. Then, we have

$$
\begin{aligned}
\left|\, \langle \boldsymbol{w}, \boldsymbol{a} \rangle - \langle \bar{\boldsymbol{w}}_t, \boldsymbol{a} \rangle \,\right| &= \left|\, \langle \boldsymbol{w} - \bar{\boldsymbol{w}}_t, \boldsymbol{a} \rangle \,\right| \\
&= \left| \sum_{i=1}^n \big(w(i) - \bar{w}_t(i)\big) a(i) \right| \\
&\leq \sum_{i=1}^n \left|w(i) - \bar{w}_t(i)\right| |a(i)| \\
&\leq \sum_{i=1}^n \mathrm{rad}_t(i) \cdot |a(i)| && (12) \\
&= \langle \mathbf{rad}_t, |\boldsymbol{a}| \rangle ,
\end{aligned}
$$

where Eq. (12) follows the definition of event ξ_t in Eq. (11) and the assumption that it occurs. ☐

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

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

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

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

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

Proof. For any $t > 0$ and $e \in [n]$, notice φ_e is a R-sub-Gaussian distribution with mean $w(e)$ and $w_t(e)$ is the empirical mean of φ_e for $T_t(e)$ samples. Using Hoeffding's inequality (see Lemma 6), we obtain

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

[Tian: I find $\mathrm{rad}_t(e) = R\sqrt{\frac{\log\frac{4nt^2}{\delta}}{2T_t(e)}}$ is enough, because substituting t with $\mathrm{rad}_t(e)$ in Hoeffding inequality we have:

$$\begin{aligned}\Pr\left[\left|\bar{w}_t(e) - w(e)\right| \geq R\sqrt{\frac{\log\left(\frac{4nt^2}{\delta}\right)}{2T_t(e)}}\right] &\leq 2\exp\left\{-\frac{2T_t(e)}{R^2}\cdot\left(R\sqrt{\frac{\log\frac{4nt^2}{\delta}}{2T_t(e)}}\right)^2\right\} \\ &= 2\exp\left\{-\log\frac{4nt^2}{\delta}\right\} \\ &\leq \frac{\delta}{2nt^2}.\end{aligned}$$

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

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

□

A.3 Main Lemmas

Now we state our key technical lemmas. In these lemmas, we shall use Lemma 2 to construct gadgets that interpolate between different members of a decision class. The first lemma shows that, if the confidence intervals are valid, then CLUCB always returns the correct answer when it stops.

Lemma 9. *Given any $t > n$, assume that event ξ_t (defined in Eq. (11)) 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 $\langle \boldsymbol{w}, \boldsymbol{\chi}_{M_*}\rangle > \langle \boldsymbol{w}, \boldsymbol{\chi}_{M_t}\rangle$.

Let $\mathcal{B}$ be an exchange class for $\mathcal{M}$. Applying Lemma 2 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

$$\begin{aligned}
\tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) &\geq \tilde{w}_t(M'_t) - \tilde{w}_t(M_t) \\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'_t} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, |\boldsymbol{\chi}_{M'} - \boldsymbol{\chi}_{M_t}| \right\rangle && (13) \\
&\geq \left\langle \boldsymbol{w}, \boldsymbol{\chi}_{M'_t} - \boldsymbol{\chi}_{M_t} \right\rangle && (14) \\
&= w(M'_t) - w(M_t) > 0, && (15)
\end{aligned}$$

where Eq. (13) follows from Lemma 5; and Eq. (14) follows the assumption that event ξ_t occurs and Lemma 7;

Therefore Eq. (15) shows that $\tilde{w}_t(\tilde{M}_t) > \tilde{w}_t(M_t)$. However, this contradicts to the stopping condition of CLUCB: $\tilde{w}_t(\tilde{M}_t) \leq \tilde{w}_t(M_t)$ and the assumption that the algorithm terminates on round t. ☐

The next lemma shows that if the confidence interval of an arm is sufficiently small, then this arm will not be played by the algorithm.

Lemma 10. *Given any $t > 0$ and suppose that event ξ_t (defined in Eq. (11)) 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 $\mathcal{B} \in \arg\min_{\mathcal{B}' \in \mathrm{Exchange}(\mathcal{M})} \mathrm{width}(\mathcal{B}')$. Note that $\mathrm{width}(\mathcal{B}) = \mathrm{width}(\mathcal{M})$. Suppose, on the contrary, that $p_t = e$. By Lemma 2, 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): $(e \in M_* \wedge e \in c_+) \vee (e \notin M_* \wedge e \in c_-)$**.**

First we construct a gadget M'_t as $M'_t = \tilde{M}_t \ominus c$ [Tian: add $= (\tilde{M}_t \setminus c_-) \cup c_+$, then it will be more intuitive.] and recall that $M'_t \in \mathcal{M}$ due to the definition of exchange class.

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 2 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 $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e > 0$. We will also use $M'_t \oplus b$ as a gadget.

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 1, 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 $\|\mathbf{rad}_t \circ (\boldsymbol{d} - \boldsymbol{\chi}_c)\|_\infty < \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}$. Using standard set-algebraic manipulations, we have

$$\begin{aligned}
M_t \backslash M'_t &= M_t \backslash (\tilde{M}_t \ominus c) \\
&= M_t \backslash (\tilde{M}_t \backslash c_+ \cup c_-) \\
&= M_t \backslash (\tilde{M}_t \backslash c_+) \cap (M_t \backslash c_-) \\
&= (M_t \cap c_+) \cup (M_t \backslash \tilde{M}_t) \cap (M_t \backslash c_-) \\
&= (M_t \backslash \tilde{M}_t) \cap (M_t \backslash c_-) && (16) \\
&\subseteq M_t \backslash \tilde{M}_t, && (17)
\end{aligned}$$

where Eq. (16) follows from $c_+ \subseteq \tilde{M}_t \backslash M_t$ and therefore $c_+ \cap M_t = \emptyset$. Similarly, we can derive $M'_t \backslash M_t$ as follows

$$M'_t \backslash M_t = (\tilde{M}_t \ominus c) \backslash M_t = (\tilde{M}_t \backslash c_+ \cup c_-) \backslash M_t$$

$$
\begin{aligned}
&= \big((\tilde{M}_t \backslash c_+) \backslash M_t\big) \cup (c_- \backslash M_t) \\
&= \tilde{M}_t \backslash c_+ \backslash M_t && (18)\\
&\subseteq \tilde{M}_t \backslash M_t, && (19)
\end{aligned}
$$

where Eq. (18) follows from $c_- \subseteq M_t \backslash \tilde{M}_t$ and hence $c_- \backslash M_t = \emptyset$. By combining Eq. (17) and Eq. (19), we see 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 3, we obtain

$$
\begin{aligned}
\|\mathbf{rad}_t \circ (\boldsymbol{d} - \boldsymbol{\chi}_c)\|_\infty &= \left\|\mathbf{rad}_t \circ (\boldsymbol{\chi}_{M'_t} - \boldsymbol{\chi}_{M_t})\right\|_\infty \\
&= \max_{i \in (M_t \backslash M'_t) \cup (M'_t \backslash M_t)} \mathrm{rad}_t(i) \\
&\le \max_{i \in (M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)} \mathrm{rad}_t(i) \\
&= \mathrm{rad}_t(e) < \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}. && (20)
\end{aligned}
$$

Next we claim that [Tian: It should be $\|\mathbf{rad}_t \circ |\boldsymbol{\chi}_c|\|_\infty$] $\|\mathbf{rad}_t \circ \boldsymbol{\chi}_c\|_\infty < \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}$. 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

$$
\begin{aligned}
\|\mathbf{rad}_t \circ |\boldsymbol{\chi}_c|\|_\infty &= \max_{i \in c_+ \cup c_-} \mathrm{rad}_t(i) \\
&\le \max_{i \in (\tilde{M}_t \backslash M_t) \cup (M_t \backslash \tilde{M}_t)} \mathrm{rad}_t(i) \\
&= \mathrm{rad}_t(e) < \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}. && (21)
\end{aligned}
$$

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

$$
\begin{aligned}
\langle \mathbf{rad}_t, |\boldsymbol{d}_2| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle &= \langle \mathbf{rad}_t, |\boldsymbol{d}_2| - |\boldsymbol{d}| \rangle = \left\langle \mathbf{rad}_t, \boldsymbol{d}_2^2 - \boldsymbol{d}^2 \right\rangle && (22)\\
&= \left\langle \mathbf{rad}_t, (\boldsymbol{d} - \boldsymbol{\chi}_c + \boldsymbol{\chi}_b)^2 - \boldsymbol{d}^2 \right\rangle \\
&= \left\langle \mathbf{rad}_t, \boldsymbol{\chi}_b^2 + \boldsymbol{\chi}_c^2 - 2\boldsymbol{\chi}_b \circ \boldsymbol{\chi}_c - 2\boldsymbol{d} \circ \boldsymbol{\chi}_c + 2\boldsymbol{d} \circ \boldsymbol{\chi}_b \right\rangle \\
&= \left\langle \mathbf{rad}_t, \boldsymbol{\chi}_b^2 - \boldsymbol{\chi}_c^2 + 2\boldsymbol{\chi}_b \circ (\boldsymbol{d} - \boldsymbol{\chi}_c) \right\rangle && (23)\\
&= \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - 2 \langle \mathbf{rad}_t, \boldsymbol{\chi}_b \circ (\boldsymbol{d} - \boldsymbol{\chi}_c) \rangle \\
&= \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - 2 \langle \mathbf{rad}_t \circ (\boldsymbol{d} - \boldsymbol{\chi}_c), \boldsymbol{\chi}_b \rangle && (24)\\
&\ge \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - 2 \|\mathbf{rad}_t \circ (\boldsymbol{d} - \boldsymbol{\chi}_c)\|_\infty \|\boldsymbol{\chi}_b\|_1 && (25)\\
&> \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - \frac{2\Delta_e}{3\,\mathrm{width}(\mathcal{B})} \|\boldsymbol{\chi}_b\|_1 && (26)\\
&\ge \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - \frac{2\Delta_e}{3}, && (27)
\end{aligned}
$$

where Eq. (22) holds since $\boldsymbol{d} \in \{-1,0,1\}^n$ and $\boldsymbol{d}_2 \in \{-1,0,1\}^n$; Eq. (23) follows from the claim that $\boldsymbol{d} \circ \boldsymbol{\chi}_c = |\boldsymbol{\chi}_c| = \boldsymbol{\chi}_c^2$; Eq. (24) and Eq. (25) follow from Lemma 4 and Hölder's inequality; Eq. (26) follows from Eq. (20); and Eq. (27) holds since $b \in \mathcal{B}$ and $\|\boldsymbol{\chi}_b\|_1 = |b_+| + |b_-| \le \mathrm{width}(\mathcal{B})$.

Applying Lemma 5 by setting $M' = M'_t \oplus b$ [Tian: I think you replace M' in Lemma 5 with $M' = \tilde{M}_t$] and using the fact that $\tilde{w}_t(\tilde{M}_t) \geq \tilde{w}_t(M'_t \oplus b)$ [Tian: from Line 9 in Algorithm 1], we have

$$\begin{aligned}
\langle \bar{\boldsymbol{w}}_t, \boldsymbol{d} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle &= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, |\boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t}| \right\rangle \\
&= \tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) \\
&\geq \tilde{w}_t(M'_t \oplus b) - \tilde{w}_t(M_t) \qquad (28) \\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M'_t \oplus b} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, |\boldsymbol{\chi}_{M'_t \oplus b} - \boldsymbol{\chi}_{M_t}| \right\rangle \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{d}_2 \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_2| \rangle \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{d} \rangle - \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_c \rangle + \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_2| \rangle \,, \qquad (29)
\end{aligned}$$

where Eq. (28) follows from the fact that $\tilde{w}_t(\tilde{M}_t) = \max_{M \in \mathcal{M}} \tilde{w}_t(M)$; and Eq. (29) follows from the fact that $\boldsymbol{d}_2 = \boldsymbol{d} - \boldsymbol{\chi}_c + \boldsymbol{\chi}_b$. Rearranging the above inequality, we obtain

$$\begin{aligned}
\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_c \rangle &\geq \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_2| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle \\
&\geq \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_b| \rangle - \langle \mathbf{rad}_t, |\boldsymbol{\chi}_c| \rangle - \frac{2\Delta_e}{3} \qquad (30) \\
&> \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \langle \mathbf{rad}_t, \boldsymbol{\chi}_c \rangle - \frac{2\Delta_e}{3} \qquad (31) \\
&> \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{\Delta_e}{3} - \frac{2\Delta_e}{3} \qquad (32) \\
&= \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \Delta_e \geq 0, \qquad (33)
\end{aligned}$$

where Eq. (30) uses Eq. (27); Eq. (31) follows from the assumption that event ξ_t occurs and Lemma 7; and Eq. (31) holds since Eq. (21) [Tian: (Index error): Eq. (32) holds since Eq. (21) with $\langle \mathbf{rad}_t, \boldsymbol{\chi}_c \rangle \leq \|\mathbf{rad}_t \circ |\boldsymbol{\chi}_c|\|_\infty \|\chi_c\|_1$].

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

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

However, the definition of M_t ensures that $M_t = \arg\max_{M \in \mathcal{M}} \bar{w}_t(M)$, i.e. $\bar{w}_t(M_t) \geq \bar{w}_t(M'_t)$. This is a contradiction, and therefore we have $p_t \neq e$ for this case.

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 $\tilde{M}_t \neq M_*$. On the other hand, suppose that $e \notin M_*$ 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 2, 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 2 also indicates that $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e > 0$. We will use $\tilde{M}_t \oplus b$ as a gadget for this case.

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 1 gives that $\boldsymbol{d}_1 = \boldsymbol{d} + \boldsymbol{\chi}_b$.

Then, we apply Lemma 3 by setting $M = M_t$ and $M' = \tilde{M}_t$. This shows that

$$\|\mathbf{rad}_t \circ \boldsymbol{d}\|_\infty \leq \max_{i : (\tilde{M}_t \backslash M_t) \cup (M_t \backslash \tilde{M}_t)} \mathrm{rad}_t(i) = \mathrm{rad}_t(e) < \frac{\Delta_e}{3}. \qquad (34)$$

[Tian: I think it is $< \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}$, but I don't know why there are $\mathrm{width}(\mathcal{B})$ and $\mathrm{width}(\mathcal{M})$ for they are exactly the same. Maybe it is better to omit one to reduce the possibility of misunderstanding.]

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

$$\begin{aligned}
\langle \bar{\boldsymbol{w}}_t, \boldsymbol{d}_1 \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_1| \rangle - \langle \bar{\boldsymbol{w}}_t, \boldsymbol{d} \rangle - \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle &= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_1| - |\boldsymbol{d}| \rangle \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \left\langle \mathbf{rad}_t, \boldsymbol{d}_1^2 - \boldsymbol{d}^2 \right\rangle \qquad (35)
\end{aligned}$$

$$
\begin{aligned}
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \left\langle \mathbf{rad}_t, 2\boldsymbol{d} \circ \boldsymbol{\chi}_b + \boldsymbol{\chi}_b^2 \right\rangle && (36)\\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle + \left\langle \mathbf{rad}_t, \boldsymbol{\chi}_b^2 \right\rangle + 2 \left\langle \mathbf{rad}_t \circ \boldsymbol{d}, \boldsymbol{\chi}_b \right\rangle \\
&\geq \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - 2 \left\langle \mathbf{rad}_t \circ \boldsymbol{d}, \boldsymbol{\chi}_b \right\rangle \text{ [Tian:it is } +2 \left\langle \mathbf{rad}_t \circ \boldsymbol{d}, \boldsymbol{\chi}_b \right\rangle\text{]} && (37)\\
&\geq \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - 2 \left\| \mathbf{rad}_t \circ \boldsymbol{d} \right\|_\infty \left\| \boldsymbol{\chi}_b \right\|_1 && (38)\\
&> \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{2\Delta_e}{3} && (39)\\
&\geq 0, && (40)
\end{aligned}
$$

where Eq. (35) follows from the fact that $\boldsymbol{d}_1 \in \{-1, 0, 1\}^n$ and $\boldsymbol{d} \in \{-1, 0, 1\}^n$; Eq. (36) holds since $\boldsymbol{d}_1 = \boldsymbol{d} + \boldsymbol{\chi}_b$; Eq. (37) follows from the assumption that ξ_t occurs and Lemma 7; Eq. (38) follows from Lemma 4 and Hölder's inequality; and Eq. (39) is due to Eq. (34).

Therefore, we have proved that

$$
\langle \bar{\boldsymbol{w}}_t, \boldsymbol{d} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle < \langle \bar{\boldsymbol{w}}_t, \boldsymbol{d}_1 \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_1| \rangle . \tag{41}
$$

However, we have

$$
\begin{aligned}
\langle \bar{\boldsymbol{w}}_t, \boldsymbol{d} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}| \rangle &= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, |\boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t}| \right\rangle && (42)\\
&= \tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) \\
&\geq \tilde{w}_t(\tilde{M}_t \oplus b) - \tilde{w}_t(M_t) && (43)\\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t \oplus b} - \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \mathbf{rad}_t, |\boldsymbol{\chi}_{\tilde{M}_t \oplus b} - \boldsymbol{\chi}_{M_t}| \right\rangle \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{d}_1 \rangle + \langle \mathbf{rad}_t, |\boldsymbol{d}_1| \rangle , && (44)
\end{aligned}
$$

where Eq. (42) follows from Lemma 5; and Eq. (43) follows from the fact that $\tilde{w}_t(\tilde{M}_t) = \max_{M \in \mathcal{M}} \tilde{w}_t(M)$. This contradicts to Eq. (41) and therefore $p_t \neq e$.

□

A.4 Proof of Theorem 1

Theorem 1 is now a straightforward corollary of Lemma 9 and Lemma 10. For the reader's convenience, we first restate Theorem 1 as follows.

Theorem 1. *Given any $\delta \in (0, 1)$, any decision class $\mathcal{M} \subseteq 2^{[n]}$ and any expected rewards $\boldsymbol{w} \in \mathbb{R}^n$. Assume that the reward distribution φ_e for each arm $e \in [n]$ is R-sub-Gaussian with mean $w(e)$. Set $\mathrm{rad}_t(e) = R\sqrt{2 \log\left(\frac{4nt^2}{\delta}\right) / T_t(e)}$ for all $t > 0$ and $e \in [n]$. Then, with probability at least $1 - \delta$, the CLUCB algorithm (Algorithm 1) returns the optimal set $\mathsf{Out} = \arg\max_{M \in \mathcal{M}} w(M)$ and*

$$
T \leq O\left(R^2 \,\mathrm{width}(\mathcal{M})^2 \mathbf{H} \log\left(R^2 \mathbf{H} / \delta\right)\right), \tag{5}
$$

where T denotes the number of samples used by Algorithm 1, $\mathbf{H}$ is defined in Eq. (2) and $\mathrm{width}(\mathcal{M})$ is defined in Eq. (4).

Proof. Lemma 8 indicates that the event $\xi \triangleq \bigcap_{t=1}^{\infty} \xi_t$ occurs with probability at least $1 - \delta$. In the rest of the proof, we shall assume that this event holds.

By Lemma 9 and the assumption on ξ, we see that $\mathsf{Out} = M_*$. Next, we focus on bounding the total number T of samples.

Fix any arm $e \in [n]$. Let $T(e)$ denote the total number of pull of arm $e \in [n]$. Let t_e be the last round which arm e is pulled, i.e. $p_{t_e} = e$. It is easy to see that $T_{t_e}(e) = T(e) - 1$. By Lemma 10, we see that $\mathrm{rad}_{t_e}(e) \geq \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{M})}$. By plugging the definition of rad_{t_e}, we have

$$
\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{M})} \leq R\sqrt{\frac{2\log\left(4nt_e^2/\delta\right)}{T(e) - 1}} \leq R\sqrt{\frac{2\log\left(4nT^2/\delta\right)}{T(e) - 1}}. \tag{45}
$$

Solving Eq. (45) for $T(e)$, we obtain

$$T(e) \leq \frac{18\,\mathrm{width}(\mathcal{M})^2 R^2}{\Delta_e^2} \log(4nT^2/\delta) + 1. \tag{46}$$

Notice that $T = \sum_{i\in[n]} T(i)$. Hence the theorem follows by summing up Eq. (46) for all $e \in [n]$ and solving for T. $\square$

B Extensions of CLUCB

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

B.1 Fixed Budget Setting

We can extend the CLUCB algorithm to the fixed budget setting by using two simple modifications: (1) requiring CLUCB to terminate after T rounds; and (2) using a different construction of confidence intervals. The first modification ensures that CLUCB 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 CLUCB is bounded by $O\left(Tn\exp\left(\frac{-T}{\mathrm{width}(\mathcal{M})^2\mathbf{H}}\right)\right)$.

Theorem 4. *Use the same notations as in Theorem 1. Given $T > n$ and parameter $\alpha > 0$, set the confidence radius $\mathrm{rad}_t(e) = R\sqrt{\frac{\alpha}{T_t(e)}}$ for all arms $e \in [n]$ and all $t > 0$. Run CLUCB 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*

$$\Pr\left[\mathsf{Out} \neq M_*\right] \leq 2Tn\exp\left(-2\alpha\right). \tag{47}$$

In particular, the right-hand side of Eq. (47) equals to $O\left(Tn\exp\left(\frac{-T}{\mathrm{width}(\mathcal{M})^2\mathbf{H}}\right)\right)$ when parameter $\alpha = O(T\mathbf{H}^{-1}\,\mathrm{width}(\mathcal{M})^{-2})$.

Theorem 4 shows that the modified CLUCB algorithm in the fixed budget setting requires the knowledge of quantity $\mathbf{H}$ in order to achieve the optimal performance. However $\mathbf{H}$ is usually unknown. Therefore, although its probability of error guarantee matches the parameter-free CSAR algorithm up to logarithmic factors, this modified algorithm is considered more restricted than CSAR. Nevertheless, Theorem 4 shows that CLUCB can solve CPE in both fixed confidence and fixed budget settings and more importantly this theorem provides additional insights on the behavior CLUCB.

B.2 PAC Learning

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

We show that a simple modification on the CLUCB algorithm gives an (ϵ,δ)-PAC algorithm, with guarantees similar to Theorem 1. In fact, the only modification needed is to change the stopping condition from $\tilde{w}_t(\tilde{M}_t) = \tilde{w}_t(M_t)$ to $w(\tilde{M}_t) - w(M_t) \leq \epsilon$ on line 13 of Algorithm 1. We let CLUCB-PAC denote the modified algorithm. In the following theorem, we show that CLUCB-PAC is indeed an (ϵ,δ)-PAC algorithm and has sample complexity similar to CLUCB.

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

$$T \leq O\left(R^2 \sum_{e\in[n]} \min\left\{\frac{\mathrm{width}(\mathcal{M})^2}{\Delta_e^2}, \frac{K^2}{\epsilon^2}\right\} \log\left(\frac{R^2}{\delta} \sum_{e\in[n]} \min\left\{\frac{\mathrm{width}(\mathcal{M})^2}{\Delta_e^2}, \frac{K^2}{\epsilon^2}\right\}\right)\right), \tag{48}$$

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

We see that the sample complexity of CLUCB-PAC decreases when ϵ increases. And if $\epsilon = 0$, the sample complexity Eq. (48) of CLUCB-PAC equals to that of CLUCB.

There are several PAC learning algorithms dedicated for the TopK problem in the literature with different guarantees [17, 23, 12]. Zhou et al. [23] proposed an (ϵ, δ)-PAC algorithm for the TopK problem with a problem-independent sample complexity bound of $O(\frac{K^2 n}{\epsilon^2} + \frac{Kn\log(1/\delta)}{\epsilon^2})$.[3] If we ignore logarithmic factors, then the sample complexity bound of CLUCB-PAC for the TopK problem is better than theirs since $\tilde{O}(\sum_{e\in[n]} \min\{\Delta_e^{-2}, K^2\epsilon^{-2}\}) \leq \tilde{O}(nK^2\epsilon^{-2})$. On the other hand, the algorithms of Kalyanakrishnan et al. [17] and Gabillon et al. [12] guarantee to find K arms such that each of them is better than the K-th optimal arm within a factor of ϵ with probability $1-\delta$. Unless $\epsilon = 0$, their guarantee is different from ours which concerns the optimality of the sum of K arms.

B.3 Proof of Extension Results

B.3.1 Fixed Budget Setting (Theorem 4)

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

Lemma 11. *Fix parameter $\alpha > 0$ and the number of rounds $T > 0$. Assume that the reward distribution φ_e is a R-sub-Gaussian distribution for all $e \in [n]$. Let the confidence radius $\mathrm{rad}_t(e)$ of arm $e \in [n]$ and round $t > 0$ be $\mathrm{rad}_t(e) = R\sqrt{\frac{\alpha}{T_t(e)}}$. Then, we have*

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

where ξ_t is the random event defined in Eq. (11).

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 4 can be obtained from the key lemmas (Lemma 9 and Lemma 10) and Lemma 11.

Proof of Theorem 4. Define random event $\xi = \bigcap_{t=1}^{T} \xi_t$. By Lemma 11, we see that $\Pr[\xi] \geq 1 - 2nT\exp(-2\alpha)$. In the rest of the proof, we assume that ξ happens.

Let T^* denote the round that the algorithm stops. We claim that the algorithm stops before the budget is exhausted, i.e. $T^* < T$. If the claim is true, then the algorithm stops since it meets the stopping condition on round T^*. Hence $\tilde{w}_t(\tilde{M}_{T^*}) = \tilde{w}_t(M_{T^*})$ and $\mathsf{Out} = M_{T^*}$. By assumption on ξ and Lemma 9, we know that $M_{T^*} = M_*$. Therefore the theorem follows immediately from this claim and the bound of $\Pr[\xi]$.

Next, we show that this claim is true. Let $T(e)$ denote the total number of pulls of arm $e \in [n]$. Let t_e be the last round that arm e is pulled. Hence $T_{t_e}(e) = T_e - 1$. By Lemma 10, we see that $\mathrm{rad}_{t_e}(e) \geq \frac{\Delta}{3\,\mathrm{width}(\mathcal{B})}$. Now plugging in the definition of $\mathrm{rad}_{t_e}(e)$, we have

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

[3] We notice that the definition of Zhou et al. [23] allow an (ϵ', δ)-PAC algorithm to produce an output with *average* sub-optimality of ϵ'. This is equivalent to our definition of (ϵ, δ)-PAC algorithm with $\epsilon = K\epsilon'$ for the TopK problem. In this paper, we translate their guarantees to our definition of PAC algorithm.

Hence we have

$$T_e \le \frac{9R^2 \operatorname{width}(\mathcal{B})^2}{\Delta_e^2} \cdot \alpha + 1. \tag{49}$$

By summing up Eq. (49) for all $e \in [n]$, we have

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

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

□

B.3.2 PAC Learning (Theorem 5)

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

Lemma 12. *If* CLUCB-PAC *stops on round t and suppose that event ξ_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 CLUCB-PAC ensures that $\tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) \le \epsilon$. Therefore, we have

$$\begin{aligned}
\epsilon &\ge \tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) \\
&\ge \tilde{w}_t(M_*) - \tilde{w}_t(M_t) && (50) \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_*} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_{M_*} - \boldsymbol{\chi}_{M_t}| \rangle && (51) \\
&\ge \langle \boldsymbol{w}, \boldsymbol{\chi}_{M_*} - \boldsymbol{\chi}_{M_t} \rangle && (52) \\
&= w(M_*) - w(M_t),
\end{aligned}$$

where Eq. (50) follows from the definition of $\tilde{M}_t \triangleq \arg\max_{M \in \mathcal{M}} \tilde{w}_t(M)$; Eq. (51) follows from Lemma 5; Eq. (52) follows from the assumption that ξ_t occurs and Lemma 7. □

The next lemma generalizes Lemma 10. It shows that on event ξ_t each arm $e \in [n]$ will not be played on round t if $\operatorname{rad}_t(e) \le \max\left\{ \frac{\Delta_e}{3 \operatorname{width}(\mathcal{M})}, \frac{\epsilon}{2K} \right\}$.

Lemma 13. *Let $K = \max_{M \in \mathcal{M}} |M|$. For any arm $e \in [n]$ and any round $t > n$ after initialization, if $\operatorname{rad}_t(e) \le \max\left\{ \frac{\Delta_e}{3 \operatorname{width}(\mathcal{M})}, \frac{\epsilon}{2K} \right\}$ and random event ξ_t occurs, then arm e will not be played on round t, i.e. $p_t \ne e$.*

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

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

$$\begin{aligned}
\epsilon &\ge \sum_{i \in (M_t \backslash \tilde{M}_t) \cup (\tilde{M}_t \backslash M_t)} \operatorname{rad}_t(i) && (53) \\
&= \left\langle \mathbf{rad}_t, \left| \chi_{M_t} - \chi_{\tilde{M}_t} \right| \right\rangle, && (54)
\end{aligned}$$

where Eq. (53) 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. (54) 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

$$\tilde{w}_t(\tilde{M}_t) - \tilde{w}_t(M_t) = \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t} \rangle + \langle \mathbf{rad}_t, |\boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t}| \rangle \tag{55}$$

$$\leq \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_t} - \boldsymbol{\chi}_{M_t} \rangle + \epsilon \tag{56}$$
$$= \bar{w}_t(\tilde{M}_t) - \bar{w}_t(M_t) + \epsilon$$
$$\leq \epsilon, \tag{57}$$

where Eq. (55) follows from Lemma 5; Eq. (56) uses Eq. (54); and Eq. (57) follows from $\bar{w}_t(M_t) \geq \bar{w}_t(\tilde{M}_t)$.

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

Using Lemma 13 and Lemma 12, we are ready to prove Theorem 5.

Proof of Theorem 5. Similar to the proof of Theorem 1, we appeal to Lemma 8, which shows that the event $\xi \triangleq \bigcap_{t=1}^{\infty} \xi_t$ occurs with probability at least $1 - \delta$. And we shall assume that ξ occurs in the rest of the proof.

By the assumption of ξ and Lemma 12, we know that $w(M_*) - w(\mathsf{Out}) \leq \epsilon$. 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 pulls of arm $e \in [n]$. Let t_e be the last round which arm e is pulled, i.e. $p_{t_e} = e$. Hence $T_{t_e}(e) = T(e) - 1$. By Lemma 13, we see that $\mathrm{rad}_{t_e}(e) \geq \min\{\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}, \frac{\epsilon}{2K}\}$. [Tian: it is $\max\{\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}, \frac{\epsilon}{2K}\}$.] Then, by the construction of $\mathrm{rad}_{t_e}(e)$, we have

$$\text{[Tian:it is } \max\{\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}, \frac{\epsilon}{2K}\}] \min\left\{\frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})}, \frac{\epsilon}{2K}\right\} \leq R\sqrt{\frac{2\log\left(4nt_e^2/\delta\right)}{T(e)-1}} \leq R\sqrt{\frac{2\log\left(4nT^2/\delta\right)}{T(e)-1}}. \tag{58}$$

Solving Eq. (58) for $T(e)$, we obtain

$$T(e) \leq R^2 \min\left\{\frac{18\,\mathrm{width}(\mathcal{B})^2}{\Delta_e^2}, \frac{16K^2}{\epsilon^2}\right\} \log(4nT^2/\delta) + 1. \tag{59}$$

Notice that $T = \sum_{i\in[n]} T(e)$. Hence the theorem follows by summing up Eq. (59) for all $e \in [n]$ and solving for T. ☐

C Proof of Lower Bound

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

$$\mathbb{E}[T] \geq \frac{1}{16}\mathbf{H}\log\left(\frac{1}{4\delta}\right), \tag{6}$$

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 δ-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

$$\mathbb{E}[T_e] \geq \frac{1}{16\Delta_e^2}\log(1/4\delta). \tag{60}$$

Notice that the theorem follows immediately by summing up Eq. (60) for all $e \in [n]$.

Fix an arm $e \in [n]$. We now focus on proving Eq. (60) by contradiction. [Tian: Assume $\mathbb{E}[T_e] < \frac{1}{16\Delta_e^2}\log(1/4\delta)$, and consider ...] 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 Δ_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

$$t_e^* = \frac{1}{16\Delta_e^2} \log\left(\frac{1}{\theta}\right). \tag{61}$$

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.

$$\begin{aligned} \Pr_0[T_e > 4t_e^*] &\leq \frac{\mathbb{E}_0[T_e]}{4t_e^*} \\ \text{[Tian:change this = to <]} &= \frac{t_e^*}{4t_e^*} = \frac{1}{4}. \end{aligned}$$

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 φ_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 μ and variance σ^2 is given by $\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. [Tian: Note that H_1 and H_0 only differs in φ_e and the rest distributions are the same.] Hence, we have

$$\begin{aligned} \frac{L_1(W)}{L_0(W)} &= \prod_{i=1}^{T_e} \exp\left(\frac{-(X_i - w(e) - 2\Delta_e)^2 + (X_i - w(e))^2}{2}\right) \\ &= \prod_{i=1}^{T_e} \exp\left(\Delta_e(2X_i - 2w(e)) - 2\Delta_e^2\right) \\ &= \exp\left(\Delta_e(2K - 2w(e)T_e) - 2\Delta_e^2 T_e\right) \\ &= \exp\left(\Delta_e(2K - 2w(e)T_e)\right) \exp(-2\Delta_e^2 T_e). \end{aligned} \tag{62}$$

Next, we bound each individual term on the right-hand side of Eq. (62). We begin with bounding the second term of Eq. (62)

$$\exp(-2\Delta_e^2 T_e) \geq \exp(-8\Delta_e^2 t_e^*) \tag{63}$$

$$= \exp\left(-\frac{8}{16}\log(1/\theta)\right) \tag{64}$$

$$= \theta^{1/2}, \tag{65}$$

where Eq. (63) follows from the assumption that event $\mathcal{S}$ occurred, which implies that event $\mathcal{A}$ occurred and therefore $T_e \leq 4t_e^*$; Eq. (64) follows from the definition of t_e^*.

Then, we bound the first term on the right-hand side of Eq. (62) as follows

$$\exp\left(\Delta_e(2K - 2w(e)T_e)\right) \geq \exp\left(-2\Delta_e\sqrt{t_e^*\log(1/\theta)}\right) \tag{66}$$

$$= \exp\left(-\frac{2}{4}\log(1/\theta)\right) \tag{67}$$

$$= \theta^{1/2}, \tag{68}$$

where Eq. (66) 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. (67) follows from the definition of t_e^*.

Combining Eq. (65) and Eq. (68), we can bound $L_1(W)/L_0(W)$ for this case as follows

$$\frac{L_1(W)}{L_0(W)} \geq \theta. \tag{69}$$

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

$$\begin{aligned} \frac{L_1(W)}{L_0(W)} &= \prod_{i=1}^{T_e} \exp\left(\frac{-(X_i - w(e) + 2\Delta_e)^2 + (X_i - w(e))^2}{2}\right) \\ &= \prod_{i=1}^{T_e} \exp\left(\Delta_e(2w(e) - 2X_i) - 2\Delta_e^2\right) \\ &= \exp\left(\Delta_e(2w(e)T_e - 2K)\right) \exp(-2\Delta_e^2 T_e). \end{aligned} \tag{70}$$

Notice that the right-hand side of Eq. (70) differs from Eq. (62) only in its first term. Now, we bound the first term as follows

$$\exp\left(\Delta_e(2w(e)T_e - 2K)\right) \geq \exp\left(-2\Delta_e\sqrt{t_e^*\log(1/\theta)}\right) \tag{71}$$

$$= \exp\left(-\frac{2}{4}\log(1/\theta)\right) \tag{72}$$

$$= \theta^{1/2}, \tag{73}$$

where the inequalities hold due to reasons similar to Case (1): Eq. (71) 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. (72) follows from the definition of t_e^*.

Combining Eq. (65) and Eq. (68), we can obtain the same bound of $L_1(W)/L_0(W)$ as in Eq. (69), 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. (69) holds, i.e. $\frac{L_1(W)}{L_0(W)} \geq \theta$. Hence, we have

$$\begin{aligned} \frac{L_1(W)}{L_0(W)} &\geq \theta \\ &= 4\delta. \end{aligned} \tag{74}$$

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

C.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 δ-correct algorithm must spend $\tilde{\Omega}\Big(\big(|b_+| + |b_-|\big)^2/\Delta_e^2\Big)$ samples on exploring 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 φ_e is given by $\varphi_e = \mathcal{N}(w(e), 1)$, where $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2. Fix any $\delta \in (0, e^{-16}/4)$ and any δ-correct algorithm $\mathbb{A}$.*

Then, for any $e \in [n]$, 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 $\delta > 0$, $\boldsymbol{w} \in \mathbb{R}^n$ and a δ-correct algorithm $\mathbb{A}$. Assume that $\varphi_e(i) = \mathcal{N}(w(i), 1)$ for all $i \in [n]$. Fix $e \in [n]$, define M_e be the "next-to-optimal" set as follows

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

By definition of Δ_e in Eq. (1), we know that $w(M_*) - w(M_e) = \Delta_e$. Construct the exchange set $b = (b_+, b_-)$ where $b_+ = M_* \backslash M_e$ and $b_- = M_e \backslash M_*$. It is easy to check that $M_e \oplus b = M_*$ and $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle = \Delta_e$.

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

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

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 $w_1(M_*) - w_1(M_e) = -w(b) < 0$ and $w_2(M_*) - w_2(M_e) = -w(b) < 0$. 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 $t^*_{b_-} = \frac{|b_-|^2}{16w(b)^2} \log(1/\theta)$. [Tian: We will prove it by contradiction, thus assume the opposite, that is, $\mathbb{E}_0[T_{b_-}] < t^*_{b_-}$.]

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

$$\mathcal{B} = \{\mathsf{Out} = M_*\}. \tag{75}$$

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

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

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

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

$$= \frac{4t^*_{b_-}}{t^*_{b_-} \log(1/\theta)}$$
$$< \frac{1}{4},$$

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

$$\begin{aligned}
\frac{L_1(W)}{L_0(W)} &= \prod_{i=1}^{T_{b_-}} \exp\left(\frac{-\left(X_i - w(I_i) - \frac{2w(b)}{|b_-|}\right)^2 + (X_i - w(I_i))^2}{2}\right) \\
&= \prod_{i=1}^{T_{b_-}} \exp\left(\frac{w(b)}{|b_-|}(2X_i - 2w(I_i)) - \frac{2w(b)^2}{|b_-|^2}\right) \\
&= \exp\left(\frac{w(b)}{|b_-|}\left(\sum_{i=1}^{T_{b_-}} 2X_i - 2w(I_i)\right) - \frac{2w(b)^2}{|b_-|^2} T_{b_-}\right) \\
&= \exp\left(\frac{w(b)}{|b_-|}\left(\sum_{i=1}^{T_{b_-}} 2X_i - 2w(I_i)\right)\right) \exp\left(-\frac{2w(b)^2}{|b_-|^2} T_{b_-}\right). \qquad (76)
\end{aligned}$$

Then, we bound each term on the right-hand side of Eq. (76). First, we bound the second term of Eq. (76).

$$\begin{aligned}
\exp\left(-\frac{2w(b)^2}{|b_-|^2} T_{b_-}\right) &\geq \exp\left(-\frac{2w(b)^2}{|b_-|^2} 4t^*_b\right) \qquad (77) \\
&= \exp\left(-\frac{8}{16}\log(1/\theta)\right) \qquad (78) \\
&= \theta^{1/2}, \qquad (79)
\end{aligned}$$

where Eq. (77) follows from the assumption that events $\mathcal{S}_1$ and $\mathcal{A}_1$ occurred and therefore $T_{b_-} \leq 4t^*_{b_-}$; Eq. (78) follows from the definition of $t^*_{b_-}$. Next, we bound the first term of Eq. (76) as follows

$$\begin{aligned}
\exp\left(\frac{w(b)}{|b_-|}\left(\sum_{i=1}^{T_{b_-}} 2X_i - 2w(I_i)\right)\right) &\geq \exp\left(-\frac{2w(b)}{|b_-|}\sqrt{t^*_b \log(1/\theta)}\right) \qquad (80) \\
&= \exp\left(-\frac{2}{4}\log(1/\theta)\right) \qquad (81) \\
&= \theta^{1/2}, \qquad (82)
\end{aligned}$$

where Eq. (80) 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. (81) follows from the definition of $t^*_{b_-}$.

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

$$\frac{L_1(W)}{L_0(W)} \geq \theta = 4\delta. \qquad (83)$$

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

$$\begin{aligned}
\Pr_1[\mathcal{B}] &\geq \Pr_1[\mathcal{S}_1] = \mathbb{E}_1[1_{S_1}] \\
&= \mathbb{E}_0\left[\frac{L_1(W)}{L_0(W)} 1_{S_1}\right] \\
&\geq 4\delta \mathbb{E}_0[1_{S_1}] \\
&= 4\delta \Pr_0[\mathcal{S}_1] > \delta. \qquad (84)
\end{aligned}$$

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

$$\mathbb{E}_0[T_{b_-}] \geq t^*_{b_-} = \frac{|b_-|^2}{16w(b)^2} \log(1/4\delta). \qquad (85)$$

(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 $t^*_{b_+} = \frac{|b_+|^2}{16w(b)^2} \log(1/\theta)$. [Tian: We will prove it by contradiction, thus assume the opposite, i.e., $\mathbb{E}_0[T_{b_+}] < t^*_{b_+}$.] 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. (75). 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

$$\begin{aligned}
\frac{L_2(W)}{L_0(W)} &= \prod_{i=1}^{T_{b_+}} \exp\left(\frac{-\left(Y_i - w(J_i)) + \frac{2w(b)}{|b_-|}\right)^2 + (Y_i - w(J_i))^2}{2}\right) \\
&= \prod_{i=1}^{T_{b_+}} \exp\left(\frac{w(b)}{|b_+|}(2w(J_i) - 2Y_i) - \frac{2w(b)^2}{|b_+|^2}\right) \\
&= \exp\left(\frac{w(b)}{|b_+|}\left(\sum_{i=1}^{T_{b_+}} 2w(J_i) - 2Y_i\right) - \frac{2w(b)^2}{|b_+|^2} T_{b_+}\right) \\
&= \exp\left(\frac{w(b)}{|b_+|}\left(\sum_{i=1}^{T_{b_+}} 2w(J_i) - 2Y_i\right)\right) \exp\left(-\frac{2w(b)^2}{|b_+|^2} T_{b_+}\right) \\
&\geq \theta \qquad (86) \\
&= 4\delta,
\end{aligned}$$

where Eq. (86) 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. (84), 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 $\mathbb{E}_0[T_{b_+}] < t^*_{b_+}$, then, under H_2, the probability of algorithm $\mathbb{A}$ returning M_* as output is at least δ, which is not the optimal set under H_2. This contradicts to the assumption that algorithm $\mathbb{A}$ is a δ-correct algorithm. In sum, we have proved that

$$\mathbb{E}_0[T_{b_+}] \geq t^*_{b_+} = \frac{|b_+|^2}{16w(b)^2}\log(1/4\delta). \tag{87}$$

(End of Part (2))

Finally, we combine the results from both parts, i.e. Eq. (85) and Eq. (87). We obtain

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

□

D Analysis of CSAR

Notations. For convenience, we will use the following additional 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 2. Let $M \subseteq [n]$ be a set, we denote $\neg M$ to be the complement of M.

D.1 Confidence Intervals

First we establish the confidence bounds used for the analysis of CSAR.

Lemma 14. *Given a phase $t \in [n]$, we define random event τ_t as follows*

$$\tau_t = \left\{\forall i \in [n]\backslash(A_t \cup B_t) \quad \left|\bar{w}_t(i) - w(i)\right| < \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})}\right\}. \tag{88}$$

Then, we have

$$\Pr\left[\bigcap_{t=1}^{T}\tau_t\right] \geq 1 - n^2\exp\left(-\frac{2(T-n)}{9R^2\tilde{\log}(n)\,\mathrm{width}(\mathcal{M})^2\mathbf{H}_2}\right). \tag{89}$$

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

$$\Pr\left[\left|\bar{w}_t(i) - w(i)\right| \geq \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})}\right] \leq 2\exp\left(-\frac{2\tilde{T}_t\Delta^2_{(n-t+1)}}{9R^2\,\mathrm{width}(\mathcal{M})^2}\right). \tag{90}$$

By plugging the definition of $\tilde{T}_t$, the quantity $\tilde{T}_t\Delta^2_{(n-t+1)}$ on the right-hand side of Eq. (90) can be further bounded by

$$
\begin{aligned}
\tilde{T}_t\Delta^2_{(n-t+1)} &\geq \frac{T-n}{\tilde{\log}(n)(n-t+1)}\Delta^2_{(n-t+1)} \\
&\geq \frac{T-n}{\tilde{\log}(n)\mathbf{H}_2},
\end{aligned}
$$

where the last inequality follows from the definition of $\mathbf{H}_2 = \max_{i\in n} i\Delta^{-2}_{(i)}$. By plugging the last inequality into Eq. (90), we have

$$
\Pr\left[\left|\bar{w}_t(i) - w(i)\right| \geq \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})}\right] \leq 2\exp\left(-\frac{2(T-n)}{9R^2\tilde{\log}(n)\,\mathrm{width}(\mathcal{M})^2\mathbf{H}_2}\right). \tag{91}
$$

Now using Eq. (91) and a union bound for all $t \in [n]$ and all $i \in [n]\backslash(A_t \cup B_t)$, we have

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

□

The following lemma builds the confidence bound of inner products.

Lemma 15. *Fix a phase $t \in [n]$, suppose that random event τ_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*

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

Proof. Suppose that τ_t occurs. Then, similar to the proof of Lemma 7, we have

$$
\begin{aligned}
|\langle \bar{\boldsymbol{w}}_t, \boldsymbol{a}\rangle - \langle \boldsymbol{w}, \boldsymbol{a}\rangle| &= |\langle \bar{\boldsymbol{w}}_t - \boldsymbol{w}, \boldsymbol{a}\rangle| \\
&= \left|\sum_{i=1}^{n}\big(\bar{w}_t(i) - w(i)\big)a(i)\right| \\
&\leq \left|\sum_{i\in[n]\backslash(A_t\cup B_t)}\big(\bar{w}_t(i) - w(i)\big)a(i)\right| \qquad (92) \\
&\leq \sum_{i\in[n]\backslash(A_t\cup B_t)}\left|\big(\bar{w}_t(i) - w(i)\big)a(i)\right| \\
&\leq \sum_{i\in[n]\backslash(A_t\cup B_t)}|\bar{w}_t(i) - w(i)|\,|a(i)| \\
&< \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})}\sum_{i\in[n]\backslash(A_t\cup B_t)}|a(i)| \qquad (93) \\
&= \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})}\left\|\boldsymbol{a}\right\|_1,
\end{aligned}
$$

where Eq. (92) follows from the assumption that $\boldsymbol{a}$ supported on $[n]\backslash(A_t \cup B_t)$; Eq. (93) follows from the definition of τ_t (Eq. (88)). □

D.1.1 Main Lemmas

Lemma 16. *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

$$\begin{aligned} A_t \cap (M \backslash a \cup b) &= (A_t \cap (M \backslash a)) \cup (A_t \cap b) \\ &= A_t \cap (M \backslash a) && (94) \\ &= (A_t \cap M) \backslash a \\ &= A_t \backslash a && (95) \\ &= A_t, && (96) \end{aligned}$$

where Eq. (94) holds since we have $A_t \cap b \subseteq A_t \cap (M_* \backslash M) \subseteq M \cap (M_* \backslash M) = \emptyset$; Eq. (95) follows from $A_t \subseteq M$; and Eq. (96) follows from $a \subseteq M \backslash M_*$ and $A_t \subseteq M_*$ which imply that $a \cap A_t = \emptyset$. Notice that Eq. (96) is equivalent to $A_t \subseteq (M \backslash a \cup b)$.

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

$$\begin{aligned} B_t \cap (M \backslash a \cup b) &= (B_t \cap (M \backslash a)) \cup (B_t \cap b) \\ &= B_t \cap (M \backslash a) && (97) \\ &= (B_t \cap M) \backslash a \\ &= \emptyset \backslash a = \emptyset, && (98) \end{aligned}$$

where Eq. (97) 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. (98) 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 $A_t \subseteq M \cap M_*$. Also note that $a \subseteq M \backslash M_*$ and $b \subseteq M_* \backslash M$, we have

$$(a \cap A_t) \cup (b \cap A_t) \subseteq \big((M \backslash M_*) \cap (M \cap M_*)\big) \cup \big((M_* \backslash M) \cap (M \cap M_*)\big) = \emptyset. \qquad (99)$$

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

$$(a \cap B_t) \cup (b \cap B_t) \subseteq ((M \backslash M_*) \cap (\neg M \cap \neg M_*)) \cup ((M_* \backslash M) \cap (\neg M \cap \neg M_*)) = \emptyset. \qquad (100)$$

By combining Eq. (99) and Eq. (100), 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 17. *Fix any round $t > 0$. Suppose that event τ_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

$$e \in (M_* \cap \neg M_t) \cup (\neg M_* \cap M_t). \qquad (101)$$

Eq. (101) 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 2. 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 16, 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

$$\bar{w}_t(M_t) \geq \bar{w}_t(M_t \oplus b). \tag{102}$$

On the other hand, we have

$$\begin{aligned}
\bar{w}_t(M_t \oplus b) &= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} + \boldsymbol{\chi}_b \right\rangle && (103)\\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \right\rangle \\
&> \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \boldsymbol{w}, \boldsymbol{\chi}_b \right\rangle - \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})} \|\boldsymbol{\chi}_b\|_1 && (104)\\
&\geq \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \boldsymbol{w}, \boldsymbol{\chi}_b \right\rangle - \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{M})} \|\boldsymbol{\chi}_b\|_1 \\
&\geq \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \right\rangle + \left\langle \boldsymbol{w}, \boldsymbol{\chi}_b \right\rangle - \frac{\Delta_e}{3} && (105)\\
&\geq \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \right\rangle + \frac{2}{3}\Delta_e && (106)\\
&\geq \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \right\rangle = \bar{w}_t(M_t), && (107)
\end{aligned}$$

where Eq. (103) follows from Lemma 1; Eq. (104) follows from Lemma 15 and the fact that $(b_+ \cup b_-) \cap (A_t \cup B_t) = \emptyset$; Eq. (105) 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. (106) and Eq. (107) 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. (102). Therefore we have $e \in (M_* \cap M_t) \cup (\neg M_* \cap \neg M_t)$.

□

Lemma 18. *Fix any round $t > 0$. Suppose that event τ_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*

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

Proof. By Lemma 17, we see that

$$e \in (M_* \cap M_t) \cup (\neg M_* \cap \neg M_t). \tag{108}$$

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. (108), 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 2. 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 16, 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

$$\bar{w}_t(M_t) \geq \bar{w}_t(M'_{t,e}). \tag{109}$$

Hence, we have

$$\begin{aligned}
\bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,e}) &\geq \bar{w}_t(M'_{t,e}) - \bar{w}_t(\tilde{M}_{t,e}) \\
&= \bar{w}_t(\tilde{M}_{t,e} \oplus b) - \bar{w}_t(\tilde{M}_{t,e}) \\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_{t,e}} + \boldsymbol{\chi}_b \right\rangle - \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{\tilde{M}_{t,e}} \right\rangle && (110)\\
&= \left\langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \right\rangle
\end{aligned}$$

$$> \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{B})} \|\boldsymbol{\chi}_b\|_1 \tag{111}$$
$$\geq \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{\Delta_e}{3\,\mathrm{width}(\mathcal{B})} \|\boldsymbol{\chi}_b\|_1 \tag{112}$$
$$\geq \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle - \frac{\Delta_e}{3} \tag{113}$$
$$\geq \frac{2}{3}\Delta_e \geq \frac{2}{3}\Delta_{(n-t+1)}, \tag{114}$$

where Eq. (110) follows from Lemma 1; Eq. (111) follows from Lemma 15, the assumption on event τ_t and the fact $(b_+ \cup b_-) \cap (A_t \cup B_t) = \emptyset$; Eq. (112) follows from the assumption that $\Delta_e \geq \Delta_{(n-t+1)}$; Eq. (113) holds since $b \in \mathcal{B}$ and therefore $\|\boldsymbol{\chi}_b\|_1 \leq \mathrm{width}(\mathcal{M})$; and Eq. (114) follows from the fact that $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e$.

□

Lemma 19. *Fix any phase $t > 0$. Suppose that event τ_t occurs. Also assume that $A_t \subseteq M_*$ and $B_t \cap M_* = \emptyset$. 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 $\mathcal{B} \in \arg\min_{\mathcal{B}' \in \mathrm{Exchange}(\mathcal{M})} \mathrm{width}(\mathcal{B}')$.

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 2 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

$$\bar{w}_t(\tilde{M}_{t,e}) \geq \bar{w}_t(M'_{t,e}). \tag{115}$$

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 16. Hence we have shown that Eq. (115) holds.

Hence, we have

$$\begin{aligned}
\bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,e}) &\leq \bar{w}_t(M_t) - \bar{w}_t(M'_{t,e}) \\
&= \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} \rangle - \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_{M_t} + \boldsymbol{\chi}_b \rangle && (116) \\
&= - \langle \bar{\boldsymbol{w}}_t, \boldsymbol{\chi}_b \rangle \\
&\leq - \langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle + \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})} \|\boldsymbol{\chi}_b\|_1 && (117) \\
&\leq \frac{\Delta_{(n-t+1)}}{3\,\mathrm{width}(\mathcal{M})} \|\boldsymbol{\chi}_b\|_1 \leq \frac{\Delta_{(n-t+1)}}{3}, && (118)
\end{aligned}$$

where Eq. (116) follows from Lemma 1; Eq. (117) follows from Lemma 15, the assumption on τ_t and $(b_+ \cup b_-) \cap (A_t \cup B_t) = \emptyset$ (by Lemma 16); and Eq. (118) follows from the fact $\|\boldsymbol{\chi}_b\|_1 \leq \mathrm{width}(\mathcal{M})$ (since $b \in \mathcal{B}$) and that $\langle \boldsymbol{w}, \boldsymbol{\chi}_b \rangle \geq \Delta_e \geq 0$. □

D.2 Proof of Theorem 3

For reader's convenience, we first restate Theorem 3 in the following.

Theorem 3. *Given any $T > n$, any decision class $\mathcal{M} \subseteq 2^{[n]}$ and any expected rewards $\boldsymbol{w} \in \mathbb{R}^n$. Assume that the reward distribution φ_e for each arm $e \in [n]$ is R-sub-Gaussian with mean $w(e)$. Let $\Delta_{(1)}, \ldots, \Delta_{(n)}$ be a permutation of $\Delta_1, \ldots, \Delta_n$ (defined in Eq. (1)) such that $\Delta_{(1)} \leq \ldots\ldots \Delta_{(n)}$.*

Define $\mathbf{H}_2 \triangleq \max_{i\in[n]} i\Delta_{(i)}^{-2}$. *Then, the* CSAR *algorithm uses at most* T *samples and outputs a solution* $\mathsf{Out} \in \mathcal{M} \cup \{\perp\}$ *such that*

$$\Pr[\mathsf{Out} \neq M_*] \leq n^2 \exp\left(-\frac{2(T-n)}{9R^2 \tilde{\log}(n)\, \mathrm{width}(\mathcal{M})^2 \mathbf{H}_2}\right), \tag{7}$$

where $\tilde{\log}(n) \triangleq \sum_{i=1}^{n} i^{-1}$, $M_* = \arg\max_{M\in\mathcal{M}} w(M)$ *and* $\mathrm{width}(\mathcal{M})$ *is defined in Eq.* (4).

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, ..., 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 14, 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 τ, the algorithm outputs M_*. We will assume that event τ 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 18, we have

$$\bar{w}_t(M_t) - \bar{w}_t(M_{t,e_1}) \geq \frac{2}{3}\Delta_{(n-t+1)}. \tag{119}$$

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 19, we see that

$$\bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,p_t}) \leq \frac{1}{3}\Delta_{(n-t+1)}. \tag{120}$$

By combining Eq. (119) and Eq. (120), we see that

$$\bar{w}_t(M_t) - \bar{w}_t(\tilde{M}_{t,p_t}) \leq \frac{1}{3}\Delta_{(n-t+1)} < \frac{2}{3}\Delta_{(n-t+1)} \leq \bar{w}_t(M_t) - \bar{w}_t(M_{t,e_1}). \tag{121}$$

However Eq. (121) 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. $\square$

E Analysis of the uniform allocation algorithm

In this section, we analyze the performance of a simple benchmark strategy Uniform which plays each arm for a equal number of times and then calls a maximization oracle using the empirical means of arms as input. The Uniform algorithm is a natural generalization of previous uniform allocation algorithms in the literature of pure exploration MABs [6]. The pseudo-code of the Uniform algorithm is listed in Algorithm 3.

The next theorem upper bounds the probability of error of Uniform.

Algorithm 3 Uniform: Uniform Allocation

Require: Budget: $T > 0$; Maximization oracle: Oracle : $\mathbb{R}^n \to \mathcal{M}$.
1: Pull each arm $e \in [n]$ for $\lfloor T/n \rfloor$ times.
2: Compute the empirical means $\bar{\boldsymbol{w}} \in \mathbb{R}^n$ of each arm.
3: $\mathsf{Out} \leftarrow \mathrm{Oracle}(\bar{\boldsymbol{w}})$
4: **return:** Out

Theorem 7. *Given any $T > n$, any decision class $\mathcal{M} \subseteq 2^{[n]}$ and any expected rewards $\boldsymbol{w} \in \mathbb{R}^n$. Assume that the reward distribution φ_e for each arm $e \in [n]$ is R-sub-Gaussian with mean $w(e)$. Define $\Delta_{(1)} = \min_{i\in[n]} \Delta_i$ and $\mathbf{H}_3 = n\Delta_{(1)}^{-2}$. Then, the output* Out *of the* Uniform *algorithm satisfies*

$$\Pr[\mathsf{Out} \neq M_*] \le 2n \exp\left(-\frac{2T}{9R^2 \operatorname{width}(\mathcal{M})^2 \mathbf{H}_3}\right), \tag{122}$$

where $M_* = \arg\max_{M\in\mathcal{M}} w(M)$.

From Theorem 7, we see that the Uniform algorithm could be significantly worse than CLUCB and CSAR, since it is clear that $\mathbf{H}_3 \ge \mathbf{H} \ge \mathbf{H}_2$ and potentially one has $\mathbf{H}_3 \gg \mathbf{H} \ge \mathbf{H}_2$ for a large number of arms with heterogeneous gaps.

Now we prove Theorem 7. The proof is straightforward using tools of exchange classes.

Proof. Define $\Delta_{(1)} = \min_{i\in[n]} \Delta_i$. Define random event ξ as follows

$$\xi = \left\{\forall i \in [n], \quad |\bar{w}(i) - w(i)| \le \frac{\Delta_{(1)}}{3 \operatorname{width}(\mathcal{M})}\right\}.$$

Notice that each arm is sampled for $\lfloor \frac{T}{n} \rfloor$ times. Therefore, using Hoeffding's inequality and union bound, we can bound $\Pr[\xi]$ as follows. Fix any $i \in [n]$, by Hoeffding's inequality, we have

$$\Pr\left[|\bar{w}(i) - w(i)| > \frac{\Delta_{(1)}}{3 \operatorname{width}(\mathcal{M})}\right] \le 2\exp\left(-\frac{2T\Delta_{(1)}^2}{9R^2 n \operatorname{width}(\mathcal{M})^2}\right).$$

Then, using a union bound, we obtain

$$\Pr\left[\xi\right] \ge 1 - 2n\exp\left(-\frac{2T\Delta_{(1)}^2}{9nR^2 \operatorname{width}(\mathcal{M})^2}\right).$$

In addition, using an argument very similar to Lemma 15, one can show that, on event ξ, for any vector $\boldsymbol{a} \in \mathbb{R}^n$, it holds that

$$|\langle \bar{\boldsymbol{w}}, \boldsymbol{a}\rangle - \langle \boldsymbol{w}, \boldsymbol{a}\rangle| \le \frac{\Delta_{(1)}}{3 \operatorname{width}(\mathcal{M})} \|\boldsymbol{a}\|_1 . \tag{123}$$

Now we claim that, on the event ξ, we have $\mathsf{Out} = M_*$, where $M_* = \arg\max_{M\in\mathcal{M}} w(M)$. Note that theorem follows immediately from the claim. Next, we prove this claim.

Suppose that, on the contrary, $\mathsf{Out} \neq M_*$. First, we write $M = \mathsf{Out}$. We also fix $\mathcal{B} \in \arg\min_{\mathcal{B}'\in \mathrm{Exchange}(\mathcal{M})} \operatorname{width}(\mathcal{B})$. Notice that by definition $\operatorname{width}(\mathcal{B}) = \operatorname{width}(\mathcal{M})$.

Since $M \neq M_*$, we see that there exists $e \in (M\backslash M_*) \cup (M_*\backslash M)$. Now, by Lemma 2, we obtain that there exists $b = (b_+, b_-) \in \mathcal{B}$ such that $e \in b_+ \cup b_-$, $b_- \subseteq M\backslash M_*$, $b_+ \subseteq M_*\backslash M$, $M \oplus b \in \mathcal{M}$ and $\langle \boldsymbol{w}, \boldsymbol{\chi}_b\rangle \ge \Delta_e$. Also notice that $\Delta_e \ge \Delta_{(1)}$. Therefore $\langle \boldsymbol{w}, \boldsymbol{\chi}_b\rangle \ge \Delta_{(1)}$.

Consider $M' \triangleq M \oplus b$. We have

$$\begin{aligned}
\bar{w}(M') - \bar{w}(M) &= \langle \bar{\boldsymbol{w}}, \boldsymbol{\chi}_{M'}\rangle - \langle \bar{\boldsymbol{w}}, \boldsymbol{\chi}_M\rangle \\
&= \langle \bar{\boldsymbol{w}}, \boldsymbol{\chi}_b\rangle \qquad (124)\\
&\ge \langle \boldsymbol{w}, \boldsymbol{\chi}_b\rangle - \frac{\Delta_{(1)}}{3 \operatorname{width}(\mathcal{M})} \|\boldsymbol{\chi}_b\|_1 \qquad (125)
\end{aligned}$$

$$\geq \Delta_{(1)} - \frac{\Delta_{(1)}}{3} \tag{126}$$

$$= \frac{2}{3}\Delta_{(1)} > 0, \tag{127}$$

where Eq. (124) follows from Lemma 1; Eq. (125) follows from Eq. (123); and Eq. (126) follows from the fact that $b \in \mathcal{B}$ and hence $\|\boldsymbol{\chi}_b\|_1 = |b_+| + |b_-| \leq \mathrm{width}(\mathcal{B}) = \mathrm{width}(\mathcal{M})$.

Hence, we have shown that $\bar{w}(M') > \bar{w}(M)$. However this contradicts to the fact that $w(M) = \max_{M_1 \in \mathcal{M}} \bar{w}(M_1)$ (by the definition of maximization oracle). Hence, by contradiction, we have proven that $\mathsf{Out} = M_*$. □

F Exchange classes for example decision classes

Notation. We need one extra notation. Let $\sigma : E \to [n]$ be a bijection from some set E with n elements to $[n]$. Let $A \subseteq E$ be an arbitrary set, we define $\sigma(A) \triangleq \{\sigma(a) \mid a \in A\}$. Conversely, for all $M \subseteq [n]$, we define $\sigma^{-1}(M) \triangleq \{\sigma^{-1}(e) \mid e \in M\}$.

Fact 1 (Matroid). *Let $T = (E, \mathcal{I})$ be an arbitrary matroid, where E is the ground set of n elements and $\mathcal{I}$ is the family of independent sets. Let $\sigma\colon E \to [n]$ be a bijection from E to $[n]$. Let $\mathcal{M}_{\textsc{Matroid}(T)}$ correspond to the collection of all bases of matroid T and formally we define*

$$\mathcal{M}_{\textsc{Matroid}(T)} = \left\{ M \subseteq [n] \mid \sigma^{-1}(M) \textit{ is a basis of } T \right\}. \tag{128}$$

Define the exchange class

$$\mathcal{B}_{\textsc{Matroid}(n)} = \left\{ (\{i\}, \{j\}) \mid \forall i \in [n], j \in [n] \right\}. \tag{129}$$

Then we have $\mathcal{B}_{\textsc{Matroid}(n)} \in \mathrm{Exchange}(\mathcal{M}_{\textsc{Matroid}(T)})$. In addition, we have $\mathrm{width}(\mathcal{B}_{\textsc{Matroid}(n)}) = 2$, which implies that $\mathrm{width}(\mathcal{M}_{\textsc{Matroid}(T)}) \leq 2$.

To prove Fact 1, we first recall a well-known result from matroid theory which is referred as the strong basis exchange property.

Lemma 20 (Strong basis exchange [21])**.** *Let $\mathcal{A}$ the set of all bases of a matroid $T = (E, \mathcal{I})$. Let $A_1, A_2 \in \mathcal{A}$ be two members of $\mathcal{A}$. Then for all $x \in A_1 \backslash A_2$, there exists $y \in A_2 \backslash A_1$ such that $A_1 \backslash \{x\} \cup \{y\} \in \mathcal{A}$ and $A_2 \backslash \{y\} \cup \{x\} \in \mathcal{A}$.*

We refer readers to [21] for a proof of Lemma 20.

Proof of Fact 1. Fix a matroid $T = (E, \mathcal{I})$ where $|E| = n$ and fix the bijection $\sigma\colon E \to [n]$. Let $\mathcal{M}_{\textsc{Matroid}(T)}$ be defined as in Eq. (128) and let $\mathcal{B}_{\textsc{Matroid}(n)}$ be defined as in Eq. (129). Let $\mathcal{A}$ denote the set of all bases of T. By definition, we have $\mathcal{M}_{\textsc{Matroid}(T)} = \{\sigma(A) \mid A \in \mathcal{A}\}$.

Now we show that $\mathcal{B}_{\textsc{Matroid}(n)}$ is an exchange class for $\mathcal{M}_{\textsc{Matroid}(T)}$. Let M, M' be two different elements of $\mathcal{M}_{\textsc{Matroid}(T)}$. By definition, we see that $\sigma^{-1}(M)$ and $\sigma^{-1}(M')$ are two bases of T. Consider any $e \in M \backslash M'$. Let $x = \sigma^{-1}(e)$. We see that $x \in \sigma^{-1}(M) \backslash \sigma^{-1}(M')$.

By Lemma 20, we see that there exists $y \in \sigma^{-1}(M') \backslash \sigma^{-1}(M)$ such that

$$\sigma^{-1}(M) \backslash \{x\} \cup \{y\} \in \mathcal{A} \quad \text{and} \quad \sigma^{-1}(M') \backslash \{y\} \cup \{x\} \in \mathcal{B}. \tag{130}$$

Now we define exchange set $b = (b_+, b_-)$ where $b_+ = \{\sigma(y)\}$ and $b_- = \{\sigma(x)\}$. By Eq. (130) and the fact that σ is a bijection, we see that $M \oplus b \in \mathcal{M}_{\textsc{Matroid}(T)}$ and $M' \ominus b \in \mathcal{M}_{\textsc{Matroid}(T)}$. We also have $b \in \mathcal{B}_{\textsc{Matroid}(n)}$. Due to M, M' and e are chosen arbitrarily, we have verified that $\mathcal{B}_{\textsc{Matroid}(n)}$ is an exchange class for $\mathcal{M}_{\textsc{Matroid}(T)}$.

To conclude, we observe that $\mathrm{width}(\mathcal{B}_{\textsc{Matroid}(n)}) = 2$. □

Now we show that TopK and MB can be reduced to the decision classes of matroids. And therefore we can apply Fact 1 to construct exchange classes and bound the widths of these decision classes.

Fact 2 (TopK)**.** *For all $K \in [n]$, let $\mathcal{M}_{\textsc{TopK}(K)} = \{M \subseteq [n] \mid |M| = K\}$ be the collection of all subsets of size K. Then we have $\mathrm{width}(\mathcal{M}_{\textsc{TopK}(K)}) \leq 2$.*

Proof. Recall the definition of a uniform matroid U_n^K [21]. We know that a subset M of $[n]$ is basis of U_n^K if and only if $|M| = K$. Therefore, we have $\mathcal{M}_{\text{TopK}(K)} = \mathcal{M}_{\text{Matroid}(U_n^K)}$. Then the conclusion follows immediately from Fact 1. □

Fact 3 (MB)**.** *For any partition $\mathcal{A} = \{A_1, \ldots, A_m\}$ of $[n]$, we define*

$$\mathcal{M}_{\text{MB}(\mathcal{A})} = \Big\{M \subseteq [n] \mid \forall i \in [m] \quad |M \cap A_i| = 1\Big\}.$$

Then we have $\text{width}(\mathcal{M}_{\text{MB}(\mathcal{A})}) \leq 2$.

Proof. Let $P_{\mathcal{A}} = ([n], \mathcal{I}_{\mathcal{A}})$ where $\mathcal{I}_{\mathcal{A}}$ is given by

$$\mathcal{I}_{\mathcal{A}} = \big\{M \subseteq [n] \mid \forall i \in [m] \quad |M \cap A_i| \leq 1\big\}.$$

It can be shown that $P_{\mathcal{A}}$ is a matroid (known as partition matroid [21]) and each basis M of $P_{\mathcal{A}}$ satisfies $|M \cap A_i| = 1$ for all $i \in [m]$. Therefore we have $\mathcal{M}_{\text{MB}(\mathcal{A})} = \mathcal{M}_{\text{Matroid}(P_{\mathcal{A}})}$. Then the conclusion follows immediately from Fact 1. □

Fact 4 (Matching)**.** *Let $G(V, E)$ be a bipartite graph with n edges. Let $\sigma\colon E \to [n]$ be a bijection. Let $\mathcal{A}$ be the set of all valid matchings in G. We define $\mathcal{M}_{\text{Match}(G)}$ as follows*

$$\mathcal{M}_{\text{Match}(G)} = \big\{\sigma(A) \mid A \in \mathcal{A}\big\}.$$

Then we have $\text{width}(\mathcal{M}_{\text{Match}(G)}) \leq |V|$.

To prove Fact 4, we recall a classical result on graph matching which characterizes the properties of augmenting cycles and augmenting paths [4].

Lemma 21. *Let $G(V, E)$ be a bipartite graph. Let M and M' be two different matching. Then the induced graph G' from the symmetric difference $(M \backslash M') \cup (M' \backslash M)$ consists of connected components that are one of the following*

- *An even cycle whose edges alternate between M and M'.*
- *A simple path whose edges alternate between M and M'.*

Proof of Fact 4. Fix a bipartite graph $G(V, E)$ and a bijection $\sigma\colon E \to [n]$. Let $M, M' \in \mathcal{M}_{\text{Match}(G)}$ be two different elements of $\mathcal{M}_{\text{Match}(G)}$ and consider an arbitrary $e \in M \backslash M'$. On a high level perspective, we construct an exchange class which contains all augmenting cycles and paths of G. We know that the symmetric difference between M and M' can be decomposed into a collection of disjoint augmenting cycles and paths. And e must be on one of the augmenting cycle or path. Then, since "applying" the augmenting cycle/path on M will yield another valid matching which does not contains e. We see that this meets the requirements of an exchange class. In the rest of the proof, we carry out the technical details of this argument.

Define $A = \sigma^{-1}(M)$ and $A' = \sigma^{-1}(M')$. Let $a = \sigma^{-1}(e)$. Then A, A' are two matchings of G. Let G' be the induced graph from the symmetric difference $(A \backslash A') \cup (A' \backslash A)$. Let C be the connected component of G' which contains the edge a. Therefore, by Lemma 21, we see that C is either an even cycle or a simple path with edges alternating between A and A'. Let C_+ contains the edges of C that belongs to $A' \backslash A$. Similarly, let C_- contains the edges of C that belongs to $A \backslash A'$. Define $b_+ = \sigma(C_+)$ and $b_- = \sigma(C_-)$. Let $b = (b_+, b_-)$ be an exchange set.

Now we construct the exchange class. Let $\mathcal{C}$ be the set of all cycles in G. Let $\mathcal{P}$ be the set of all paths in G. We define $\mathcal{B}_{\text{Match}(G)}$ to correspond to the set of all cycles and all paths of G with edges alternating between b_+ and b_-. Formally, we have

$$\mathcal{B}_{\text{Match}(G)} = \Big\{(\sigma(c_+), \sigma(c_-)) \mid \exists c \in \mathcal{C} \cup P,\ \text{the edges of } c \text{ alternate between } c_+, c_-\Big\}.$$

We see that $b \in \mathcal{B}_{\text{Match}(G)}$. Since $a \in C_-$, we obtain that $e \in b_-$. In addition, note that $C_+ \subseteq A' \backslash A$ and $C_- \subseteq A \backslash A'$. Therefore we have $b_+ \subseteq M' \backslash M$ and $b_- \subseteq M \backslash M'$.

Since C is an A-augmenting path/cycle, therefore it immediately holds that $A \backslash C_- \cup C_+$ is a matching for G. Therefore, we have $M \backslash b_- \cup b_+ \in \mathcal{M}_{\text{Match}(G)}$. Similarly, one can show that

$M'\backslash b_+ \cup b_- \in \mathcal{M}_{\textsc{Match}(G)}$. Hence we have shown that $\mathcal{B}_{\textsc{Match}(G)}$ is an exchange class for $\mathcal{M}_{\textsc{Match}(G)}$.

Observe that $\mathrm{width}(\mathcal{B}_{\textsc{Match}(G)}) \leq |V|$. We conclude that $\mathrm{width}(\mathcal{M}_{\textsc{Match}(G)}) \leq |V|$. $\square$

Fact 5 (Path)**.** *Let $G(V, E)$ be a directed acyclic graph with n edges. Let $s, t \in V$ be two different vertices. Let $\sigma : E \to [n]$ be a bijection. Let $\mathcal{A}(s, t)$ be the set of all valid paths from s to t in G. We define $\mathcal{M}_{\textsc{Path}(G,s,t)}$ as follows*

$$\mathcal{M}_{\textsc{Path}(G,s,t)} = \big\{\sigma(A) \mid A \in \mathcal{A}(s, t)\big\}.$$

Then we have $\mathrm{width}(\mathcal{M}_{\textsc{Path}(G,s,t)}) \leq |V|$.

Proof. Fix a directed acyclic graph $G(V, E)$ and a bijection $\sigma : E \to [n]$. Fix two vertices $s, t \in V$. We define $\mathcal{B}_{\textsc{Path}(G)}$ as follows

$$\mathcal{B}_{\textsc{Path}(G)} = \{(\sigma^{-1}(p), \sigma^{-1}(q)) \mid p, q \text{ are the arcs of two disjoint paths of } G \text{ with same endpoints}\}.$$

Next, we prove that $\mathcal{B}_{\textsc{Path}(G)}$ is an exchange class for $\mathcal{M}_{\textsc{Path}(G,s,t)}$. Let $M, M' \in \mathcal{M}_{\textsc{Path}(G,s,t)}$ be two different sets. Then $\sigma^{-1}(M), \sigma^{-1}(M')$ are the sets of arcs of two different paths from s to t. Let $P = (v_1, \ldots, v_{n_1}), P' = (v'_1, \ldots, v'_{n_2})$ denote the two paths, respectively. Also denote $E(P) = \sigma^{-1}(M)$ and $E(P') = \sigma^{-1}(M')$.

Fix $e \in M\backslash M'$ and define $a = \sigma^{-1}(e)$. Suppose that a is an arc from u to v. Suppose $v_i = u$ and $v_{i+1} = v$. Define $j_1 = \arg\max_{j \leq i, v_j \in P'} j$ and $j_2 = \arg\min_{j \geq i+1, v_j \in P'} j$. Let $v'_{k_1} = v_{j_1}$ and $v'_{k_2} = v_{j_2}$ be the corresponding indices in P'. Then, we see that $Q_1 = (v_{j_1}, v_{j_1+1}, \ldots, v_{j_2})$ and $Q_2 = (v'_{k_1}, v'_{k_1+1}, \ldots, v'_{k_2})$ are two different paths from v_{j_1} to v_{j_2}. Denote the sets of arcs of Q_1 and Q_2 as $E(Q_1)$ and $E(Q_2)$.

Let $b = (b_+, b_-)$, where $b_+ = \sigma(E(Q_2)), b_- = \sigma(E(Q_1))$. We see that $b \in \mathcal{B}_{\textsc{Path}(G)}$. It is clear that $a \in E(Q_1)$, $E(Q_1) \subseteq E(P)\backslash E(P')$ and $E(Q_2) \subseteq E(P')\backslash E(P)$. Therefore $e \in b_-$, $b_- \subseteq M\backslash M'$ and $b_+ \subseteq M'\backslash M$.

Now it is easy to check that $E(P_1)\backslash E(Q_1) \cup E(Q_2)$ equals the set of arcs of path $(v_1, \ldots, v_{j_1}, v'_{k_1+1}, \ldots, v'_{k_2-1}, v_{j_2}, \ldots, v_{n_1})$ (recall that $v_{j_1} = v'_{k_1}$ and $v_{j_2} = v'_{k_2}$). This means that $E(P_1)\backslash E(Q_1) \cup E(Q_2) \in \mathcal{A}(s, t)$ and therefore $M\backslash b_- \cup b_+ \in \mathcal{M}_{\textsc{Path}(G,s,t)}$. Using a similar argument, one can show that $M'\backslash b_+ \cup b_- \in \mathcal{M}_{\textsc{Path}(G,s,t)}$ and hence we have verified that $\mathcal{B}_{\textsc{Path}(G)} \in \mathrm{Exchange}(\mathcal{M}_{\textsc{Path}(G,s,t)})$.

Hence we can conclude by observing that $\mathrm{width}(\mathcal{B}_{\textsc{Path}}(G)) \leq |V|$. $\square$

G Reducing constrained oracles to maximization oracles

In this section, we present a general method to implement constrained oracles using maximization oracles. The idea of the reduction is simple: one can impose the negative constrains B by setting the corresponding weights to be sufficiently small; and one can impose the positive constrains A by setting the corresponding weights to very large. The reduction method is shown in Algorithm 4. The correctness of the reduction is proved in Fact 6. Furthermore, it is trivial to reduce from maximization oracles to constrained oracles. Therefore, Fact 6 shows that maximization oracles are equivalent to constrained oracles up to a transformation on the weight vector.

Fact 6. *Given $\mathcal{M} \subseteq 2^{[n]}$, $\boldsymbol{w} \in \mathbb{R}^n$, $A \subseteq [n]$ and $B \subseteq [n]$, suppose that $A \cap B = \emptyset$. Then the output* Out *of Algorithm 4 satisfies* $\mathsf{Out} \in \arg\max_{M \in \mathcal{M}, A \subseteq M, B \cap M = \emptyset} w(M)$ *where we use the convention that the* $\arg\max$ *of an empty set is $\perp$. Therefore Algorithm 4 is a valid constrained oracle.*

Proof. Let $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ be defined as in Algoritm 4. Let $M = \mathrm{Oracle}(\boldsymbol{w}_2)$. Let $\mathcal{M}_{A,B} = \{M \in \mathcal{M} \mid A \subseteq M, B \cap M = \emptyset\}$ be the subset of $\mathcal{M}$ which satisfies the constraints. If $\mathcal{M}_{A,B} = \emptyset$. Then it is clear M cannot satisfy both of the constraints $A \subseteq M$ and $B \cap M = \emptyset$. Therefore Algorithm 4 returns $\perp$ in this case.

Algorithm 4 COracle($\boldsymbol{w}, A, B$)

Require: $\boldsymbol{w} \in \mathbb{R}^n$, $A \subseteq [n]$, $B \subseteq [n]$; Maximization oracle Oracle : $\mathbb{R}^n \to \mathcal{M}$

```
L_1 ← ‖w‖_1
for i = 1, ..., n do
    if i ∈ A then
        w_1(i) ← 3L_1
    else
        w_1(i) ← w(i)
L_2 ← ‖w_1‖_1
for i = 1, ..., n do
    if i ∈ B then
        w_2(i) ← −3L_2
    else
        w_2(i) ← w_1(i)
M ← Oracle(w_2)
if B ∩ M = ∅ and A ⊆ M then
    Out = M
else
    Out = ⊥
return: Out
```

In the rest of the proof, we assume that $\mathcal{M}_{A,B} \neq \emptyset$. Since $\mathcal{M}_{A,B}$ is non-empty, we can fix an arbitrary $M_0 \in \mathcal{M}_{A,B}$, which will be used later in the proof. We will also frequently use the fact that for all $\boldsymbol{v} \in \mathbb{R}^n$ and all $S \subseteq [n]$, we have $-\|\boldsymbol{v}\|_1 \leq v(S) \leq \|\boldsymbol{v}\|_1$.

First we claim that $B \cap M = \emptyset$. Suppose that $B \cap M \neq \emptyset$. Then there exists $i \in B \cap M$. Notice that $w_2(i) = -3L_2$ and recall that $L_2 = \|\boldsymbol{w}_1\|_1$. Therefore we have

$$
\begin{aligned}
w_2(M) &= w_2(M \backslash \{i\}) + w_2(i) \leq w_2(M \backslash B) + w_2(i) \\
&= w_1(M \backslash B) + w_2(i) \leq L_2 - 3L_2 \leq -2L_2.
\end{aligned}
$$

On the other hand, observing that $B \cap M_0 = \emptyset$, we can bound $w_2(M_0)$ as follows

$$
w_2(M_0) = w_1(M_0) \geq -L_2.
$$

Therefore we see that $w_2(M_0) \geq w_2(M)$. However, this contradicts to the definition of M since $M \in \arg\max_{M' \in \mathcal{M}} w_2(M')$. Therefore our claim $B \cap M = \emptyset$ is true. Since $\boldsymbol{w}_2$ and $\boldsymbol{w}_1$ coincide on entries of $[n] \backslash B$, we have

$$
w_2(M) = w_1(M). \tag{131}
$$

Next we claim that $A \subseteq M$. Suppose that $A \not\subseteq M$. We see that $|M \cap A| \leq |A| - 1$. By combining Eq. (131), we have

$$
\begin{aligned}
w_2(M) &= w_1(M) = w_1(M \cap A) + w_1(M \backslash A) \\
&= 3|M \cap A| L_1 + w(M \backslash A) \leq 3(|A| - 1)L_1 + L_1 = (3|A| - 2)L_1.
\end{aligned}
$$

On the other hand, using the fact that $A \subseteq M_0$, we see that

$$
w_2(M_0) = w_1(M_0) = w_1(A) + w_1(M \backslash A) = 3|A|L_1 + w(M_0 \backslash A) \geq 3|A|L_1 - L_1 = (3|A| - 1)L_1.
$$

Therefore, we see that $w_2(M_0) > w_2(M)$. Again this contradicts to the definition of M, which proves the claim.

Now we see that $M \in \mathcal{M}_{A,B}$. Therefore, we remain to verify that $w(M) = \max_{M' \in \mathcal{M}_{A,B}} w(M')$. Suppose that there exists $M_1 \in \mathcal{M}_{A,B}$ such that $w(M_1) > w(M)$. Notice that $B \cap M_1 = \emptyset$ and $A \subseteq M_1$, we have

$$
w_2(M_1) = w_1(M_1) = w_1(M_1 \backslash A) + w_1(B) = w(M_1 \backslash A) + 3|A|L_1 = w(M_1) + 3|A|L_1 - w(A).
$$

Similarly, one can show that $w_2(M) = w(M) + 3|A|L_1 - w(A)$. By combining with the assumption that $w(M_1) > w(M)$ we see that $w_2(M_1) > w_2(M)$, which contradicts to the definition of M. Hence we have verified that $w(M) = \max_{M' \in \mathcal{M}_{A,B}} w(M')$. $\square$

References

[1] J.-Y. Audibert and S. Bubeck. Minimax policies for adversarial and stochastic bandits. In *COLT*, 2009.

[2] J.-Y. Audibert, S. Bubeck, and R. Munos. Best arm identification in multi-armed bandits. In *COLT*, 2010.

[3] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

[4] C. Berge. Two theorems in graph theory. *Proceedings of the National Academy of Sciences*, 43(9):842–844, 1957.

[5] S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5:1–122, 2012.

[6] S. Bubeck, R. Munos, and G. Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412:1832–1852, 2010.

[7] S. Bubeck, N. Cesa-bianchi, S. M. Kakade, S. Mannor, N. Srebro, and R. C. Williamson. Towards minimax policies for online linear optimization with bandit feedback. In *COLT*, 2012.

[8] S. Bubeck, T. Wang, and N. Viswanathan. Multiple identifications in multi-armed bandits. In *ICML*, pages 258–265, 2013.

[9] N. Cesa-Bianchi and G. Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

[10] W. Chen, Y. Wang, and Y. Yuan. Combinatorial multi-armed bandit: General framework and applications. In *Proceedings of The 30th International Conference on Machine Learning*, pages 151–159, 2013.

[11] V. Gabillon, M. Ghavamzadeh, A. Lazaric, and S. Bubeck. Multi-bandit best arm identification. In *NIPS*. 2011.

[12] V. Gabillon, M. Ghavamzadeh, and A. Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. In *NIPS*, 2012.

[13] K. Jamieson and R. Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, pages 1–6. IEEE, 2014.

[14] K. Jamieson, M. Malloy, R. Nowak, and S. Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. *COLT*, 2014.

[15] S. Kale, L. Reyzin, and R. E. Schapire. Non-stochastic bandit slate problems. In *NIPS*, pages 1054–1062, 2010.

[16] S. Kalyanakrishnan and P. Stone. Efficient selection of multiple bandit arms: Theory and practice. In *ICML*, pages 511–518, 2010.

[17] S. Kalyanakrishnan, A. Tewari, P. Auer, and P. Stone. Pac subset selection in stochastic multi-armed bandits. In *ICML*, pages 655–662, 2012.

[18] T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[19] S. Mannor and J. N. Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *The Journal of Machine Learning Research*, 5:623–648, 2004.

[20] G. Neu, A. György, and C. Szepesvári. The online loop-free stochastic shortest-path problem. In *COLT*, pages 231–243, 2010.

[21] J. G. Oxley. *Matroid theory*. Oxford university press, 2006.

[22] D. Pollard. Asymptopia. *Manuscript, Yale University, Dept. of Statist., New Haven, Connecticut*, 2000.

[23] Y. Zhou, X. Chen, and J. Li. Optimal pac multiple arm identification with applications to crowdsourcing. In *ICML*, 2014.