Proceedings of Machine Learning Research vol 99:1–31, 2019 32nd Annual Conference on Learning Theory

Disagreement-Based Combinatorial Pure Exploration: Sample Complexity Bounds and an Efficient Algorithm

Tongyi Cao TCAO@CS.UMASS.EDU
University of Massachusetts, Amherst

Akshay Krishnamurthy AKSHAY@CS.UMASS.EDU
Microsoft Research, New York City

Editors: Alina Beygelzimer and Daniel Hsu

Abstract

We design new algorithms for the combinatorial pure exploration problem in the multi-arm bandit framework. In this problem, we are given K distributions and a collection of subsets $\mathcal{V} \subset 2^{[K]}$ of these distributions, and we would like to find the subset $v \in \mathcal{V}$ that has largest mean, while collecting, in a sequential fashion, as few samples from the distributions as possible. In both the fixed budget and fixed confidence settings, our algorithms achieve new sample-complexity bounds that provide polynomial improvements on previous results in some settings. Via an information-theoretic lower bound, we show that no approach based on uniform sampling can improve on ours in any regime, yielding the first interactive algorithms for this problem with this basic property. Computationally, we show how to efficiently implement our fixed confidence algorithm whenever $\mathcal{V}$ supports efficient linear optimization. Our results involve precise concentration-of-measure arguments and a new algorithm for linear programming with exponentially many constraints.

Keywords: Interactive learning; Bandits; Combinatorial optimization

1. Introduction

Driven by applications in engineering and the sciences, much contemporary research in mathematical statistics focuses on recovering structural information from noisy data. Combinatorial structures that have seen intense theoretical investigation include clusterings (Mossel et al., 2014; Abbe et al., 2016; Balakrishnan et al., 2011), submatrices (Butucea and Ingster, 2013; Kolar et al., 2011; Chen and Xu, 2016), and graph theoretic structures like matchings, spanning trees, and paths (Arias-Castro and Candès, 2008; Addario-Berry et al., 2010). In this paper, we design interactive learning algorithms for these structure discovery problems.

Our mathematical formulation is through the *combinatorial pure exploration for multi-armed bandits* framework (Chen et al., 2014), a recent generalization of the best-arm identification problem (Mannor and Tsitsiklis, 2004; Audibert and Bubeck, 2010). In this setting, we are given a combinatorial decision set $\mathcal{V} \subset 2^{[K]}$ and access to K arms, where each arm $a \in [K]$ is associated with a distribution with unknown mean μ_a. We can, in sequential fashion, query an arm and obtain an iid sample from the corresponding distribution, and the goal is to identify the subset $v \in \mathcal{V}$ with maximum mean $\mu(v) = \sum_{a\in v} \mu_a$ while minimizing the number of samples collected.

This model has been studied in recent work both in the general form and with specific decision sets $\mathcal{V}$. For specific structures, a line of work established near-optimal algorithms for any $\mathcal{V}$ that corresponds to the bases of a matroid (Kalyanakrishnan et al., 2012; Kaufmann and Kalyanakrishnan, 2013; Chen et al., 2016) and slight generalizations (Chen et al., 2014). After Chen et al. (2014) introduced the general problem, Gabillon et al. (2016) and Chen et al. (2017) made interesting progress with improved guarantees reflecting precise dependence on the underlying mean vector μ. However, these results fail to capture intricate combinatorial structure of the decision set, and, as we show, they can be polynomially worse than a simple non-interactive algorithm based on maximum likelihood estimation. With this in mind, our goal is to capture this combinatorial structure to design an algorithm that is never worse than the non-interactive baseline, but that can be much better.

Since we are doing combinatorial optimization, we typically consider decision sets $\mathcal{V}$ that are exponentially large but have small description length, so that direct enumeration of the elements in $\mathcal{V}$ is not computationally tractable. Instead, we assume that $\mathcal{V}$ supports efficient linear optimization, and our main algorithm only accesses $\mathcal{V}$ through a linear optimization oracle. To shed further light on purely statistical issues, we also present some results for computationally inefficient algorithms.

Our Contributions. We make the following contributions:

1. First, we derive the minimax optimal sample complexity in the non-interactive setting, where arms are queried uniformly. This precisely characterizes how the structure of $\mathcal{V}$ influences the sample complexity and also provides a baseline for evaluating interactive algorithms.
2. In the fixed confidence setting, we design two algorithms that are never worse than the non-interactive minimax rate, but that can adapt to heterogeneity in the problem to be substantially better. On the computational side, we show how to implement the first algorithm in polynomial time with access to a linear optimization oracle. The second algorithm is computationally inefficient, but has a strictly better sample complexity.
3. In the fixed budget setting, we design an algorithm with similar statistical improvements, improving on the MLE when there is heterogeneity in the problem.
4. We perform a careful comparison to prior work, with several concrete examples. We show that prior results (Chen et al., 2014; Gabillon et al., 2016; Chen et al., 2017) can be polynomially worse than the non-interactive minimax rate, which contrasts with our guarantees. We also describe other settings where our results outperform prior work, and vice versa.

Our Techniques. The core of our statistical analysis is a new deviation bound for combinatorial pure exploration that we call a *normalized regret inequality*. We prove that to recover the optimal subset $v^\star$, it suffices to control, for each $v \in \mathcal{V}$, the sampling error in the mean difference between v and $v^\star$ at a level proportional to the symmetric set difference between the two.[1] In the non-interactive setting, the normalized regret inequality always yields the optimal sample complexity (as we prove in Theorem 2), and is often sharper than more standard uniform convergence arguments (e.g., over the arms or the decision set) that have been used in prior work. Our new guarantees stem from using this new inequality in interactive procedures.

The fixed confidence setting poses a significant challenge, since confidence bounds typically appear algorithmically, but the normalized regret inequality is centered around the optimum, which is of course unknown! We address this difficulty with an elimination-style algorithm that eliminates a hypothesis $v \in \mathcal{V}$ when any other candidate is significantly better and that queries only where the survivors disagree. Using only the normalized regret inequality, we can prove that $v^\star$ is never

1. The name arises because the inequality involves comparison with the optimum and is normalized by the set difference.

eliminated, but also that $v^\star$ will eventually eliminate every other hypothesis. This algorithm resembles approaches for disagreement-based active learning (Hanneke, 2014), but uses a much stronger elimination criteria that is crucial for obtaining our sample complexity guarantees.

Computationally, deciding if the surviving candidates disagree on an arm poses further challenges, since the description of the surviving set involves exponentially many constraints, one for each candidate $u \in \mathcal{V}$. This problem can be written as a linear program, which we can solve using the Plotkin-Shmoys-Tardos reduction to online learning (Plotkin et al., 1995). However, since there are exponentially many constraints in the LP, the standard approach of using multiplicative weight updates fails, but, exploiting further structure in the problem, we can run Follow-the-Perturbed-Leader (Kalai and Vempala, 2005), since the online learner's problem is actually linear in the candidates u that parameterize the constraints. Thus with a linear optimization oracle, we obtain an efficient algorithm for the fixed confidence setting.

2. Preliminaries

In the combinatorial pure exploration problem, we are given a finite set of arms $A \triangleq \{1, \ldots, K\}$, where arm a is associated with a sub-Gaussian distribution ν_a with unknown mean $\mu_a \in [-1, 1]$ and variance parameter 1.[2] Further, we are given a decision set $\mathcal{V} \subseteq 2^A$. For $u, v \in \mathcal{V}$, we use $d(u, v) \triangleq |u \ominus v|$ where $\ominus$ denotes the symmetric set difference. The goal is to identify a set $v \in \mathcal{V}$ that has the largest collective mean $\sum_{a \in v} \mu_a$. Throughout the paper, we use the vectorized notation $\mu \triangleq (\mu_1, \ldots, \mu_K)$ and $\mathcal{V} \subset \{0, 1\}^K$. With this notation, we seek to compute

$$v^\star \triangleq \operatorname*{argmax}_{v \in \mathcal{V}} \langle v, \mu \rangle.$$

We are interested in learning algorithms that acquire information about the unknown μ in an interactive, iterative fashion. At the t^{th} iteration, the learning algorithm selects an arm a_t and receives a corresponding observation $y_t \sim \nu_{a_t}$. The algorithm's choice a_t may depend on all previous decisions and observations $\{(a_\tau, y_\tau)\}_{\tau=1}^{t-1}$ and possibly additional randomness.

We consider two related performance goals. In the *fixed budget setting*, the learning algorithm is given a budget of T queries, after which it must produce an estimate $\hat{v}$ of the true optimum $v^\star$, and we seek to minimize the probability of error $\mathbb{P}[\hat{v} \neq v^\star]$. In the *fixed confidence setting*, a failure probability parameter δ is provided as input to the algorithm, which still produces an estimate $\hat{v}$, but must further enjoy the guarantee that $\mathbb{P}[\hat{v} \neq v^\star] \leq \delta$. In this setting, we seek to minimize the number of queries issued by the algorithm.

Since we are performing optimization over $\mathcal{V}$, for computational efficiency, we equip our algorithms with a linear optimization oracle for $\mathcal{V}$. Formally, we assume access to a function

$$\textsc{Oracle}(c) \triangleq \operatorname*{argmax}_{v \in \mathcal{V}} \langle v, c \rangle, \tag{1}$$

that solves the offline combinatorial optimization problem. This oracle is available in many combinatorial problems, including max-weight matchings, spanning trees, and shortest/longest paths in DAGs,[3] and is a basic requirement here since otherwise even if μ were known we would not be able

2. Recall a sub-Gaussian random variable X with mean μ and variance parameter σ^2 satisfies $\mathbb{E}[\exp(s(X - \mu))] \leq \exp(\sigma^2 s^2/2)$. The results easily generalize to arbitrary known variance parameter.
3. Other problems may require slight reformulations of the setup for computational efficiency. For example, shortest paths in undirected graphs requires $\mu_a \in [0, 1]$ and considering a minimization version.

to find $v^\star$. Technically, we allow the oracle to take one additional constraint of the form $v_a = b$ for $a \in A, b \in \{0,1\}$, which preserves computational efficiency in most cases.

To fix ideas, we describe two concrete motivating examples (see also Section 4).

Example 1 (MATCHING) *Consider a complete bipartite graph with $\sqrt{K}$ vertices in each partition cell so that there are K edges, which we identify with A. Let $\mathcal{V}$ denote the perfect bipartite matchings and let μ assign a weight to each edge. Here, the combinatorial pure exploration task amounts to finding the maximum-weight bipartite matching in a graph with edge weights that are initially unknown. Note that the linear optimization oracle (1) is available here.*

Example 2 (BICLIQUE) *In the same graph-theoretic setting, let $\mathcal{V}$ denote the set of bicliques with $\sqrt{s}$ vertices from each partition. Equivalently in a $\sqrt{K} \times \sqrt{K}$ matrix, $\mathcal{V}$ corresponds to all submatrices of $\sqrt{s}$ rows and $\sqrt{s}$ columns. This problem is variously referred to as biclique, biclustering, or submatrix localization, and has applications in genomics (Wang et al., 2007). Unfortunately, (1) is known to be NP-hard for this structure.*

A new complexity measure. We define two complexity measures that govern the performance of our algorithms. We start with the notion of a gap between the decision sets:

$$\Delta_v(\mu) \triangleq \frac{\langle v^\star - v, \mu \rangle}{d(v^\star, v)}.$$

$\Delta_v(\mu)$ captures the difficulty of determining if v is better than $v^\star$, and the normalization $d(v^\star, v)$ accounts for the fact that the numerator is a sum of precisely $d(v, v^\star)$ terms. The gap for arm a is $\Delta_a(\mu) \triangleq \min_{v: a \in v^\star \ominus v} \Delta_v(\mu)$, which captures the difficulty of determining if a is in the optimal set.

We also introduce complexity measures that are independent of μ. For $v \in \mathcal{V}$ and $k \in \mathbb{N}$, let $\mathcal{B}(k, v) \triangleq \{u \in \mathcal{V} \mid d(v, u) = k\}$ be the sphere of radius k centered at v. Then define

$$\Phi \triangleq \Phi(\mathcal{V}) \triangleq \max_{k \in \mathbb{N}, v \in \mathcal{V}} \frac{\log(|\mathcal{B}(k, v)|)}{k}.$$

For some intuition, Φ measures the growth rate of $\mathcal{V}$ as we expand away from some candidate v. Finally, let $\Psi \triangleq \Psi(\mathcal{V}) \triangleq \min_{u,v \in \mathcal{V}} d(u, v)$ denote the smallest distance. In all of these definitions, we omit the dependence on μ and $\mathcal{V}$ when it is clear from context.

A deviation bound and the non-interactive setting. As a reference point and to foreshadow our results, we first study the non-interactive setting. With budget T, a non-interactive algorithm queries each arm T/K times and then outputs an estimate $\hat{v}$ of $v^\star$.[4] In this case, we have the following *normalized regret inequality*, which will play a central role in our analysis.

Lemma 1 (Normalized Regret Inequality) *Query each arm T/K times and let $\hat{\mu} \in \mathbb{R}^K$ be the vector of sample averages. Then $\forall \delta \in (0,1)$,*

$$\mathbb{P}\left(\exists v \in \mathcal{V} : \frac{|\langle v^\star - v, \hat{\mu} - \mu \rangle|}{d(v^\star, v)} \geq \sqrt{\frac{2K}{T} \cdot \left(\Phi + \frac{\log(2K/\delta)}{\Psi} \right)} \right) \leq \delta. \tag{2}$$

4. For simplicity, we do not implement a stopping rule, which naïvely incurs a $\max_{v \neq v^\star} \log(1/\Delta_v)$ dependence.

This inequality and a simple argument sharply characterize the performance of the MLE, $\hat{v} = \text{argmax}_{v \in \mathcal{V}} \langle v, \hat{\mu} \rangle$, which we prove is nearly minimax optimal for the non-interactive setting.

Theorem 2 (Non-interactive upper and lower bound) *For any $\mu \in [-1, 1]^K$ and $\delta \in (0, 1)$ the non-interactive MLE guarantees $\mathbb{P}_\mu[\hat{v} \neq v^\star] \leq \delta$ with*

$$T = O\left(\frac{K}{\min_{v \neq v^\star} \Delta_v^2}\left(\Phi + \frac{\log(K/\delta)}{\Psi}\right)\right).$$

Further, with $\mathcal{S}(v^\star, \Delta) \triangleq \{\mu : v^\star = \text{argmax}_{v \in \mathcal{V}} \langle v, \mu \rangle, \forall v \neq v^\star, \Delta_v(\mu) \geq \Delta\}$, any non-interactive algorithm must have $\sup_{v^\star \in \mathcal{V}} \sup_{\mu \in \mathcal{S}(v^\star, \Delta)} \mathbb{P}_\mu[\hat{v} \neq v^\star] \geq 1/2$ as long as $T \leq \frac{K}{\Delta^2}(\Phi - \log\log 3)$.

See Appendix A for the proof of Lemma 1 and Theorem 2. The proofs are not difficult and involve adapting the argument of Krishnamurthy (2016) to our setting. We also extend his result by introducing the combinatorial parameters Φ, Ψ, which are analytically tractable in many cases. Indeed, since $\Phi \leq \log(K)$, a more interpretable, but strictly weaker, upper bound is $O\left(\frac{K \log(K/\delta)}{\min_v \Delta_v^2}\right)$.

To compare the upper and lower bounds, note that $\Phi \geq 1/\Psi$ always but in most examples, including MATCHING and BICLIQUE, we actually have $\Phi \geq \log(K)/\Psi$. As such, in the moderate confidence regime where $\delta = \text{poly}(1/K)$, the upper and lower bounds disagree by at most $\log(K)$ factor, but typically they agree up to constants. Hence, Theorem 2 identifies the minimax non-interactive sample complexity and we may conclude that the MLE is near-optimal here.

Prior results. The departure point for our work is the observation that all prior results for the interactive setting can be polynomially worse than the bound for the MLE. We defer a detailed comparison to Section 4, but as a specific example, on $\sqrt{s}$-BICLIQUE, Theorem 2 can improve on the bound of Chen et al. (2014) by a factor of $s^{3/2}$, and it can improve on the bounds of Gabillon et al. (2016) and Chen et al. (2017) by a factor of $\sqrt{s}$.

For intuition, the analyses of Chen et al. (2014) and Gabillon et al. (2016) involve a uniform convergence argument over individual arms. As noted by Chen et al. (2017), this argument is suboptimal whenever $\Psi \gg 1$ as it does not take advantage of large distances between hypotheses (note that $\Psi = \sqrt{s}$ in BICLIQUE). The analysis of Chen et al. (2017) avoids this argument, but instead involves uniform convergence over the decision set, which can be suboptimal when set differences to $v^\star$ vary in size. (e.g., in BICLIQUE, the minimum and maximum distances are $O(\sqrt{s})$ and $\Omega(s)$ respectively). In comparison, our analysis *always* gives the minimax optimal non-interactive rate (up to a $\log(K/\delta)$ factor), reflecting the advantage of the normalized regret inequality over these other proof techniques. In the next section, we show how this inequality yields an interactive algorithm that is never worse than the MLE but that can also be substantially better.

Related work. Combinatorial pure exploration generalizes the best arm identification problem, which has been extensively studied (c.f., (Even-Dar et al., 2006; Mannor and Tsitsiklis, 2004; Audibert and Bubeck, 2010; Karnin et al., 2013; Russo, 2016; Garivier and Kaufmann, 2016; Carpentier and Locatelli, 2016; Chen et al., 2016; Simchowitz et al., 2017) for some classical and recent results). This problem is much simpler both computationally and statistically than ours, and, accordingly, the results are much more precise. One important difference is that in best arm identification, verifying that the optimal solution is correct is roughly as hard as finding the optimal solution, which motivates many algorithms and lower bounds based on Le Cam's method (Kaufmann et al., 2014; Karnin, 2016; Garivier and Kaufmann, 2016; Chen et al., 2017). However, for combinatorial problems discovering the optimal solution often dominates the sample complexity, and hence these techniques do not immediately produce near-optimal results in the combinatorial setting. Nevertheless our algorithms are inspired by some ideas from this literature, namely elimination and successive-reject techniques (Even-Dar et al., 2006; Audibert and Bubeck, 2010).

The subset selection problem, also called TOP-K, is a special case of combinatorial exploration where $\mathcal{V}$ corresponds to all $\binom{K}{s}$ subsets (Kalyanakrishnan et al., 2012; Bubeck et al., 2013; Kaufmann and Kalyanakrishnan, 2013). This case is minimally structured, and, in particular, there is little to be gained from our approach since $\Phi = \Theta(\log(K))$. A related effect occurs when the decision set corresponds to the basis of a matroid (Chen et al., 2016).

Structure discovery has also been studied in related mathematical disciplines including electrical engineering and statistics. Research on *adaptive sensing* from the signal processing community studies a similar setup but with assumptions on the mean μ, which lead to more specialized algorithms that fail in our general setup (Castro and Tánczos, 2015). Work from information theory and statistics focuses on non-interactive versions of the problem and typically considers specific combinatorial structures (Krishnamurthy, 2016; Balakrishnan et al., 2011; Arias-Castro and Candès, 2008). In particular, the BICLIQUE problem is extensively studied in the non-interactive setting and the minimax rate is well-known (Chen and Xu, 2016; Kolar et al., 2011; Butucea and Ingster, 2013).

Our fixed confidence algorithm is inspired by disagreement-based active learning approaches, which eliminate inconsistent hypotheses and query where the surviving ones disagree (Cohn et al., 1994; Hsu, 2010; Hanneke, 2014). Our algorithm is similar but uses a stronger elimination criteria, leading to sharper results for exact identification. Unfortunately, exact identification is rather different from PAC-learning, and it seems our approach yields no improvement for PAC-active learning.

Lastly, we use an optimization oracle as a computational primitive. This abstraction has been used previously in combinatorial pure exploration (Chen et al., 2014, 2017), but also in other information acquisition problems including active learning (Hsu, 2010; Huang et al., 2015) and contextual bandits (Agarwal et al., 2014; Syrgkanis et al., 2016; Rakhlin and Sridharan, 2016).

3. Results

Pseudocode for our fixed confidence algorithm is given in Algorithm 1. The algorithm proceeds in rounds, and at each round it issues queries to a judiciously chosen subset of the arms. These arms are chosen by implicitly maintaining a version space of plausibly optimal hypotheses and checking for disagreement among the version space.

The key ingredient is the definition of the version space. For a vector $\hat{\mu} \in \mathbb{R}^K$ and a radius parameter Δ, the version space is defined as

$$\mathcal{V}(\hat{\mu}, \Delta) \triangleq \{v \in \text{conv}(\mathcal{V}) \mid \forall u \in \mathcal{V}, \langle \hat{\mu}, u - v \rangle \leq \Delta \|u - v\|_1\}. \tag{3}$$

Here $\text{conv}(\mathcal{V})$ is the convex hull of $\mathcal{V}$ and $\|u - v\|_1$ is the ℓ_1 norm, which is just $d(u, v)$ for binary u, v. The version space is normalized in that the radius is modulated by $\|u - v\|_1$, which is justified by (2). This yields much sharper guarantees than the more standard un-normalized definition $\{v \mid \max_{u \in \mathcal{V}} \langle \hat{\mu}, u - v \rangle \leq \Delta\}$ from the active learning literature (Cohn et al., 1994; Hsu, 2010). At round t, the version space we use is $\mathcal{V}_t \triangleq \mathcal{V}(\hat{\mu}_t, \Delta_t)$ where $\hat{\mu}_t$ is the empirical mean vector and Δ_t is defined in the algorithm based on the right hand side of (2).

Algorithm 1 Fixed Confidence Algorithm

```
1: Input: Class $\mathcal{V}$, failure probability $\delta \in (0,1)$
2: Set $\Delta_t = \min\left\{1, \sqrt{\frac{8}{t}\left(\frac{\Phi\Psi + \log(\frac{K\pi^2 t^2}{\delta})}{\Psi}\right)}\right\}$
3: Sample each arm once $y_0(a) \sim \nu_a$
4: Set $\hat{\mu}_1 = y_0$
5: for $t = 1, 2 \ldots,$ do
6:     Compute $\hat{v}_t = \mathrm{argmax}_{v \in \mathcal{V}} \langle v, \hat{\mu}_t \rangle$
7:     for $a \in [K]$ do
8:         if DIS$(a, 1 - \hat{v}_t(a), \Delta_t, \hat{\mu}_t, \frac{\delta}{t^2\pi^2})$
9:             Query $a$, set $y_t(a) \sim \nu_a$
10:        else
11:            Set $y_t(a) = 2\hat{v}_t(a) - 1$
12:        endif
13:    end for
14:    Update $\hat{\mu}_{t+1} \leftarrow \frac{1}{t+1}\sum_{i=0}^{t} y_i$
15:    If no queries issued this round, output $\hat{v}_t$
16: end for
```

Algorithm 2 Oracle-based Disagreement (DIS)

```
Input: $a, b, \Delta, \hat{\mu}, \delta$
Set $T = \frac{169K^3 \log(4K/\delta)}{\Delta^2}$, $m = T\log(\frac{4KT}{\delta})$
Set $\epsilon = \sqrt{\frac{1}{25KT\log(4K/\delta)}}$, $\ell_0 = 0$
for $t = 1, \ldots, T$ do
    for $i = 1, \ldots, m$ do
        Sample $\sigma_{t,i} \sim$ Unif$([0, 1/\epsilon]^K)$
        $u_{t,i}$ = ORACLE$(\mathcal{V}, \sum_{t=0}^{t-1} \ell_t + \sigma_{t,i})$
    end for
    Let $s, x_t$ be the value and optimum of
        $\max \sum_{i=1}^{m} \Delta\langle v, \mathbf{1} - 2u_{t,i}\rangle + \langle v, \hat{\mu}\rangle$   (5)
        s.t. $v \in$ conv$(\mathcal{V}), v(a) = b$
    val $= \sum_{i=1}^{m} \langle u_{t,i}, \Delta\mathbf{1} - \hat{\mu}\rangle$
    if $s +$ val $< 0$ return FALSE
    Set $\ell_t = \Delta\mathbf{1} - 2\Delta x_t - \hat{\mu}$
end for
return TRUE
```

This version space is used by the disagreement computation (Algorithm 2), which, with parameters $a \in A, b \in \{0,1\}, \Delta, \hat{\mu}, \delta$ approximately solves the feasibility problem

$$?\exists v \in \mathcal{V}(\hat{\mu}, \Delta) \text{ s.t. } v(a) = b. \tag{4}$$

At round t, we use $\hat{\mu}_t, \Delta_t$, and the value for b that we use in line 8 is $1 - \hat{v}_t(a)$, where $\hat{v}_t$ is the empirically best hypothesis on $\hat{\mu}_t$. Since $\hat{v}_t$, being the empirically best hypothesis, is always in $\mathcal{V}_t$, this computation amounts to checking if there exist two surviving hypotheses that *disagree* on arm a. We use this disagreement-based criteria to drive the query strategy.

Before turning to computational considerations, a few other details warrant some discussion. First, if at any round we detect that there is no disagreement on some arm a, then we use a *hallucinated* observation $y_t(a) = 2\hat{v}_t(a) - 1 \in \{\pm 1\}$. While this leads to bias in our estimates, since all surviving hypotheses $v \in \mathcal{V}_t$ agree with $\hat{v}_t$ on arm a, this bias favors the survivors. As in related work on disagreement-based active learning, this helps enforces monotonicity of the version space (Dasgupta et al., 2007). Finally, we terminate once there are no remaining arms with disagreement, at which point we output the empirically best hypothesis.

Note that to set Δ_t in the algorithm, we must compute Φ. This can be done analytically for many structures including, paths in various graph models, bipartite matching, and the biclique problem. Even when it cannot, Φ is independent of the unknown means, so it can always be computed via enumeration, although this may compromise the efficency of the algorithm. Finally, we can always use the upper bound $\Phi \leq \log(K)$, which may increase the sample complexity, but will not affect the correctness of the algorithm.

Efficient implementation of disagreement computation. Computationally, the bottleneck is the feasibility problem (4) for the disagreement computation. All other computations in Algorithm 1 can trivially be done in polynomial time with access to the optimization oracle (1). Therefore, to derive an oracle-efficient algorithm, we show how to solve (4), with pseudocode in Algorithm 2.

It is not hard to see that (4) is a linear feasibility problem, but it has $|\mathcal{V}|$ constraints, which could be exponentially large. This precludes standard linear programming approaches, and instead we use the Plotkin-Shmoys-Tardos reduction to online learning (Plotkin et al., 1995).[5] The idea is to run an online learner to compute distributions over the constraints and solve simpler feasibility problems to generate the losses. In our case, the constraints are parametrized by candidates $u \in \mathcal{V}$ and we can express each generated loss as a linear function of the constraint parameter u, which enables us to use Follow-The-Perturbed-Leader (FTPL) as the online learning algorithm (Kalai and Vempala, 2005). Importantly, FTPL can be implemented using only the linear optimization oracle. As a technical detail, we must use an empirical distribution based on repeated oracle calls to approximate the true FTPL distribution, since in our reduction the loss function is generated after and based on the random decision of the learner.

First we provide the guarantee for the disagreement routine.

Theorem 3 (Efficient Disagreement Computation) *Algorithm 2 with parameters $a, b, \Delta, \hat{\mu}, \delta$ runs in polynomial time with $\tilde{O}(K^6/\Delta^4)$ calls to* ORACLE*. If it reports* FALSE *then Program (4) is infeasible. If it reports* TRUE *then with probability at least $1-\delta$, $\exists v \in conv(\mathcal{V}), v(a) = b$ such that $\forall u \in \mathcal{V}, \langle \hat{\mu}, u - v\rangle \leq \Delta \|u - v\|_1 + \Delta$.*

This result proves that Algorithm 2 can approximate the feasibility problem in (4) in polynomial time using the optimization oracle. The approximation is one-sided and, since we do not query when the algorithm returns FALSE, only affects the sample complexity of Algorithm 1, but never the correctness. The following theorem, which is the correctness and sample complexity guarantee for Algorithm 1, shows that this approximation has negligible effect. For the theorem, recall the definition of the arm gaps $\Delta_a \triangleq \min_{v: a \in v \ominus v^\star} \Delta_v$.

Theorem 4 (Fixed confidence sample complexity bound) *For any combinatorial exploration instance with mean vector μ, and any $\delta \in (0,1)$, Algorithm 1 guarantees that $\mathbb{P}[\hat{v} \neq v^\star] \leq \delta$. Moreover, it runs in polynomial time with access to the optimization oracle, and the total number of samples is at most*

$$\sum_{a \in K} \frac{144}{\Delta_a^2}\left(\Phi + \frac{2\log(144/(\Delta_a^2\Psi)) + 2\log(K\pi^2/\delta)}{\Psi}\right).$$

The bound replaces the worst case gap, $\frac{K}{\min_v \Delta_v^2}$, in Theorem 2 with a less pessimistic notion, $\sum_a \Delta_a^{-2}$, that accounts for heterogeneity in the problem. Since $\sum_a \Delta_a^{-2} \leq \frac{K}{\min_v \Delta_v^2}$, the bound is never worse than the minimax lower bound for non-interactive algorithms (given in Theorem 2) by more than a logarithmic factor, but it can be much better if many arms have large gaps. To our knowledge, Algorithm 1 is the first combinatorial exploration algorithm that is never worse than non-interactive approaches yet can exploit heterogeneity in the problem.

Theorem 4 is not easily comparable with prior results for combinatorial pure exploration, which use different complexity measures than our gaps Δ_a, Φ, and Ψ. Our observations from Theorem 2

5. Technically, we do have a separation oracle here, so we could use the Ellipsoid algorithm, but a standard application would certify feasibility or approximate infeasibility. Our reduction instead certifies infeasibility or approximate feasibility, which is more convenient.

apply here: since Theorem 4 precisely captures the combinatorial structure of $\mathcal{V}$, it can yield polynomial improvements over prior work. On the other hand, our notion of gap Δ_a is different from, and typically smaller than, prior definitions, so these results can also dominate ours. We defer a detailed comparison with calculations for several concrete examples to Section 4.

We provide the full proof for Theorem 3 and Theorem 4 in Appendix C. As a brief sketch, we prove a martingale version of Lemma 1. This inequality and the choice of Δ_t verifies that $v^\star$ is never eliminated. We then show that once $\Delta_t < \Delta_a$, all hypothesis $v \in \mathcal{V}$ with $a \in v^\star \ominus v$ satisfy $\langle \hat{\mu}, v^\star - v \rangle > \Delta_t \|v^\star - v\|_1$ and so they are eliminated from the version space. Theorem 3 then guarantees that arm a will never be queried again, which yields the sample complexity bound.

3.1. Deferred Results

In this section we state two related results: a guarantee for a disagreement-based algorithm in the fixed budget setting, and a more refined sample complexity bound for a computationally inefficient fixed confidence algorithm. Both algorithms and all proof details are deferred to the appendices.

3.1.1. A FIXED BUDGET ALGORITHM

Recall that in the fixed budget setting, the algorithm is given a budget of T queries and after issuing these queries, it must output an estimate $\hat{v}$. The goal is to minimize $\mathbb{P}[\hat{v} \neq v^\star]$. As is common in the literature, this setting requires a modified definition of the instance complexity, which for our fixed confidence result is $H \triangleq \sum_{a \in [K]} \Delta_a^{-2}$. For the definition, let $\Delta^{(j)}$ denote the j^{th} largest Δ_a value, breaking ties arbitrarily. The complexity measure for the fixed budget setting is

$$\tilde{H} \triangleq \max_j (K + 1 - j)(\Delta^{(j)})^{-2}.$$

It is not hard to see that $\tilde{H} \leq H \leq \widetilde{\log}(K)\tilde{H}$, where $\widetilde{\log}(t) = \sum_{i=1}^t 1/i$ is the partial harmonic sum. With these new definitions, we can state our fixed budget guarantee.

Theorem 5 (Fixed Budget Guarantee) *Given budget $T \geq K$, there exists a fixed budget algorithm (Algorithm 3 in Appendix E) that guarantees*

$$\mathbb{P}[\hat{v} \neq v] \leq K^2 \exp\left\{ \Psi \left(\Phi - \frac{(T - K)}{9\widetilde{\log}(K)\tilde{H}} \right) \right\}.$$

See Appendix E for the proof. At a high level, the savings over a naïve analysis are similar to Theorem 4. By using the normalized regret inequality, we obtain a refined dependence on the hypothesis complexity, replacing $\log|\mathcal{V}|$ with the potentially much smaller $\log|\mathcal{B}(k, v^\star)|$ (implicitly through the Φ parameter). To compare, non-interactive methods scale with $K(\Delta^{(1)})^{-2}$ instead of $\tilde{H}$, so the bound is never worse, but it can yield an improvement when the arm gaps are not all equal, which results in $\tilde{H} < K(\Delta^{(1)})^{-2}$. Unfortunately, the algorithm is not oracle-efficient.

3.1.2. A REFINED FIXED CONFIDENCE GUARANTEE

We also derive a sharper sample complexity bound for the fixed confidence setting. First, define

$$D(v, v') \triangleq \max\{\log|\mathcal{B}(d(v, v'), v)|, \log|\mathcal{B}(d(v, v'), v')|\},$$

to be the symmetric log-volume. We use two new instance-dependent complexity measures:

$$H_a^{(1)} \triangleq \max_{v: a \in v \ominus v^\star} \frac{d(v, v^\star)}{\langle \mu, v^\star - v \rangle^2}, \qquad H_a^{(2)} \triangleq \max_{v: a \in v \ominus v^\star} \frac{d(v, v^\star) D(v, v^\star)}{\langle \mu, v^\star - v \rangle^2}.$$

The two definitions provide more refined control on the two terms in Theorem 4. Specifically $H^{(1)}$ replaces the minimum distance Ψ with the distance to the hypothesis maximizing the complexity measure. Similarly $H^{(2)}$ replaces the volume measure Φ with a notion particular to the maximizing hypothesis. Using these definitions, we have the following fixed-confidence guarantee.

Theorem 6 (Refined fixed confidence guarantee) *There exists a fixed confidence algorithm (Algorithm 4 in Appendix F) that guarantees $\mathbb{P}[\hat{v} \neq v^\star] \leq \delta$ with sample complexity*

$$T \leq 64 \sum_{a \in [K]} H_a^{(1)} \left(2 \log(64 H_a^{(1)}) + \log \frac{\pi^2 K}{\delta} \right) + 64 H_a^{(2)}.$$

To understand this bound and compare with Theorem 4, note that we always have $H_a^{(1)} \leq \left(\Delta_a^2 \Psi\right)^{-1}$ and $H_a^{(2)} \leq \frac{\Phi}{\Delta_a^2}$. As such, Theorem 6 improves on Theorem 4 by replacing worst case quantities Φ, Ψ with instance-specific variants. However, the algorithm is not oracle-efficient.

4. Examples and Comparisons

As mentioned, the bound in Theorem 4 is somewhat incomparable to previous results (Chen et al., 2014; Gabillon et al., 2016; Chen et al., 2017). To provide general insights, we perform an instance-independent, structure-specific analysis, fixing $\mathcal{V}$ and tracking combinatorial quantities but considering the least favorable choice of μ. Such an analysis reveals when one method dominates another for hypothesis class $\mathcal{V}$ *for all mean vectors*, but is less informative about specific instances. For a complementary view, we study the *homogeneous setting*, where $\mu = \Delta(2v^\star - \mathbf{1})$ for some $v^\star \in \mathcal{V}$.

We also consider four specific examples. The first is the TOP-K problem, where $\mathcal{V}$ corresponds to all $\binom{K}{s}$ subsets. The second is DISJSET, where there are K arms and K/s hypotheses each corresponding to a disjoint set of s arms, generalizing an example of Chen et al. (2017).[6] The third and fourth examples are MATCHING from Example 1 and $\sqrt{s}$-BICLIQUE from Example 2.

Throughout, we ignore constant and logarithmic factors, and we use $\lesssim, \asymp$ to denote such asymptotic comparisons. The main combinatorial parameters are $\Psi \triangleq \min_{u,v \in \mathcal{V}} d(u, v)$, $D \triangleq \max_{u,v \in \mathcal{V}} d(u, v)$, $\log |\mathcal{V}|$, and our combinatorial term $\Lambda \triangleq (\Phi + 1/\Psi)$. T denotes our sample complexity bound from Theorem 4, which we instantiate in the the top row of Table 4 for the four examples in the homogeneous setting. All calculations are deferred to Appendix B.

Comparison with Chen et al. (2014). The bound of Chen et al. (2014) is $\tilde{O}\left(\sum_a \frac{\text{width}^2}{(\Delta_a^{(C)})^2}\right)$, where $\Delta_a^{(C)} \triangleq \min_{v: a \in v \ominus v^\star} \langle \mu, v^\star - v \rangle$ is an unnormalized gap, and width is the size of the largest augmenting set in the best collection of augmenting sets for $\mathcal{V}$. In contrast, in our bound of $\tilde{O}(\Lambda \sum_a \Delta_a^{-2})$, the normalized gaps that we use incorporate the distance between sets and our combinatorial term $\Lambda \lesssim 1$ is small. The structure-specific relationship with our bound is

$$T_{\text{Chen14}} \cdot \frac{\Psi^2}{\text{width}^2} \Lambda \lesssim T \lesssim T_{\text{Chen14}} \cdot \frac{D^2}{\text{width}^2} \Lambda.$$

6. As they argue, DISJSET is important because it can be embedded in other combinatorial structures, like disjoint paths.

Sample complexity	TOP-K	DISJSET	MATCHING	BICLIQUE
Theorem 4	$\Theta(K)$	$\Theta(K/s)$	$O(K)$	$O(K/\sqrt{s})$
Chen et al. (2014)	$\Theta(K)$	$\Theta(K)$	$\Theta(K^2)$	$\Theta(Ks)$
Chen et al. (2017)	$\Theta(K)$	$\Theta(K/s)$	$\Omega(K^{3/2})$	$\Omega(\sqrt{Ks}+K/\sqrt{s})$
Gabillon et al. (2016)	$\Theta(K)$	$\Theta(K)$	$\Theta(K)$	$\Theta(K)$

Table 1: Guarantees for four algorithms on specific examples, with $\mu = \Delta(2v^\star - 1)$, ignoring logarithmic factors. All bounds scale with $1/\Delta^2$, which is suppressed. For homogeneous problems, Theorem 4 is never worse than prior results and can be polynomially better.

Therefore, whenever width is comparable with the diameter D, our bound is never worse and, further, if $\Lambda < 1$ our bound provides a strict improvement. DISJSET is precisely such an example, where our bound is a factor of s better than that of Chen et al. (2014) for all instances. On the other hand, in TOP-K, we have width $= \Psi = 1$ but $D = 2s$ and $\Lambda \asymp 1$ so our bound is never better, but can be worse by a factor of up to s^2, depending on μ. As a final observation, in the homogeneous case, we have $\Delta_a \geq \frac{\Delta_a^{(C)}}{\text{width}}$ and so our bound is never worse and is strictly better whenever $\Lambda < 1$.

Turning to the examples in the homogeneous case, we instantiate the bound of Chen et al. (2014) in the second row of Table 4. Our bound matches theirs for TOP-K and is polynomially better for DISJSET, MATCHING, and BICLIQUE.

Comparison to Gabillon et al. (2016). Gabillon et al. (2016) use a normalized definition of hypothesis complexity similar to our $\Delta_v(\mu)$, but they compare each hypothesis v to its complement $C_v \in \mathcal{V}$ where $C_v \triangleq \text{argmax}_{v' \neq v} \frac{\langle \mu, v'-v\rangle}{d(v',v)}$. They then define an arm complexity as $\Delta_a^{(G)} \triangleq \min_{v:a\in v\ominus C_v} \frac{\langle \mu, C_v - v\rangle}{d(C_v,v)}$, and obtain the final bound $\tilde{O}\left(\sum_a (\Delta_a^{(G)})^{-2}\right)$. In contrast, we always compare v with $v^\star$ and so $\Delta_a \leq \Delta_a^{(G)}$, but our bound exploits favorable structural properties of the hypothesis class by scaling with Λ, which is small. The structure-specific relationship is

$$T_{\text{Gabillon16}}\Lambda \lesssim T \lesssim T_{\text{Gabillon16}} \cdot \frac{D^2}{\Psi^2}\Lambda.$$

Three observations from above apply here as well: (1) For DISJSET, our bound yields a factor of s improvement on all instances, (2) for TOP-K, our bound is never better but can be a factor of s^2 worse on some instances, and (3) our bound is never worse in the homogeneous case and is an improvement whenever $\Lambda < 1$. On the specific examples in the homogeneous case, we obtain a polynomial improvement on BICLIQUE, where $\Psi = 1/\sqrt{s}$ (See Table 4).

Comparison to Chen et al. (2017). Finally, Chen et al. (2017) introduce a third arm complexity parameter based on the solution to an optimization problem, which they call Low. They prove a fixed-confidence lower bound of $\Omega(\text{Low}\log(1/\delta))$ and an upper bound of $\tilde{O}\left(\text{Low}\log(|\mathcal{V}|/\delta)\right)$. In general, the sharpest structure-specific relationship we can obtain is

$$T_{\text{Chen17}} \cdot \frac{\Lambda}{\log|\mathcal{V}|} \lesssim T \lesssim T_{\text{Chen17}} \cdot \frac{D^2}{\log|\mathcal{V}|}\Lambda.$$

The second inequality results from a rather crude lower bound on Low and is therefore quite pessimistic. Indeed, the factor on the right hand side is always at least 1, so this bound does not reveal any structure where we can improve on Chen et al. (2017) for all choices of μ. Unfortunately it is difficult to relate Low to natural problem parameters in general.

On the other hand, in the homogeneous case, we can bound Low precisely in examples, yielding the third row of Table 4. Roughly speaking, our bound replaces $\log |\mathcal{V}|/\Psi$ with Λ, which is always smaller, leading to polynomial improvements in BICLIQUE, when $s > \sqrt{K}$, and MATCHING.

Final Remarks. We close this section with some final remarks. First, our result shows that $\log(|\mathcal{V}|)$ dependence is not necessary for many structured classes. This does not contradict the lower bound in Theorem 1.9 of Chen et al. (2017), which constructs certain pathological classes.

Second, we believe that the optimization-based measure Low, corresponds to the sample complexity for verifying that a proposed v is optimal. Indeed the bound of Chen et al. (2017) is optimal in the extremely high-confidence setting ($\delta \leq 1/|\mathcal{V}|$), where high-probability verification dominates the sample complexity, yet it is more natural to consider polynomially- rather than exponentially-small δ. In the moderate-confidence case, exploration to find a suitable hypothesis v is the dominant cost, but the upper bound of Chen et al. (2017) can be suboptimal here. We believe the $\Omega(\text{Low} \log(1/\delta))$ lower bound is loose in this regime, but are not aware of sharper lower bounds.

Finally, we note that for the fixed-confidence setting it is easy to achieve the best of all of these guarantees, simply by running the algorithms in parallel. For example, by interleaving queries issued by Algorithm 1 and the algorithm of Chen et al. (2017), we obtain a sample complexity of $2\min\{T, T_{\text{chen17}}\}$ with probability $1 - 2\delta$. This yields an algorithm that is never worse than the non-interactive minimax optimal rate *and* is instance-optimal in the high-confidence regime.

5. Discussion and Open Problems

This paper derives new algorithms for combinatorial pure exploration. The algorithms represent a new sample complexity trade-off and importantly are never worse than any non-interactive algorithm, contrasting with prior results. Moreover, our fixed confidence algorithm can be efficiently implemented whenever the combinatorial family supports efficient linear optimization.

We close with some open problems.

- In the homogeneous BICLIQUE, our bound is $O(K/(\sqrt{s}\Delta^2))$, yet one can actually achieve $O(\frac{1}{\Delta^2}(\sqrt{K} + K/s))$ with a specialized algorithm (Castro and Tánczos, 2015). Whether the faster rate is achievable beyond the homogeneous case is open, and seems related to the fact that active learning at best provides distribution-dependent savings in general but can provide exponential savings with random classification noise (analogous to our homogeneous setting).

- Relatedly, settling the optimal sample complexity for combinatorial pure exploration is open. For lower bounds, the technical barrier is to capture the multiple testing phenomena, which typically requires Fano's Lemma. For upper bounds, some interesting algorithms to study are median-elimination (Even-Dar et al., 2006), explore-then-verify (Karnin, 2016), and sample-and-prune (Chen et al., 2016), all of which yield optimal algorithms in special cases.

We hope to study these questions in future work.

Acknowledgments

We thank Sivaraman Balakrishnan for formative and insightful discussion.

References

Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 2016.

Louigi Addario-Berry, Nicolas Broutin, Luc Devroye, and Gábor Lugosi. On combinatorial testing problems. *The Annals of Statistics*, 2010.

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

András Antos, Varun Grover, and Csaba Szepesvári. Active learning in heteroscedastic noise. *Theoretical Computer Science*, 2010.

Ery Arias-Castro and Emmanuel J. Candès. Searching for a trail of evidence in a maze. *The Annals of Statistics*, 2008.

Jean-Yves Audibert and Sébastien Bubeck. Best arm identification in multi-armed bandits. In *Conference on Learning Theory*, 2010.

Sivaraman Balakrishnan, Min Xu, Akshay Krishnamurthy, and Aarti Singh. Noise Thresholds for Spectral Clustering. In *Advances in Neural Information Processing Systems*, 2011.

Sébastian Bubeck, Tengyao Wang, and Nitin Viswanathan. Multiple identifications in multi-armed bandits. In *International Conference on Machine Learning*, 2013.

Cristina Butucea and Yuri I. Ingster. Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli*, 2013.

Alexandra Carpentier and Andrea Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Conference on Learning Theory*, 2016.

Rui Castro and Ervin Tánczos. Adaptive sensing for estimation of structured sparse signals. *IEEE Transactions on Information Theory*, 2015.

Lijie Chen, Anupam Gupta, and Jian Li. Pure exploration of multi-armed bandit under matroid constraints. In *Conference on Learning Theory*, 2016.

Lijie Chen, Anupam Gupta, Jian Li, Mingda Qiao, and Ruosong Wang. Nearly optimal sampling algorithms for combinatorial pure exploration. In *Conference on Learning Theory*, 2017.

Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In *Advances in Neural Information Processing Systems*, 2014.

Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *Journal of Machine Learning Research*, 2016.

David Cohn, Les Atlas, and Richard Ladner. Improving generalization with active learning. *Machine Learning*, 1994.

Sanjoy Dasgupta, Daniel Hsu, and Claire Monteleoni. A general agnostic active learning algorithm. In *Advances in Neural Information Processing Systems*, 2007.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 2006.

Victor Gabillon, Alessandro Lazaric, Mohammad Ghavamzadeh, Ronald Ortner, and Peter Bartlett. Improved learning complexity in combinatorial pure exploration bandits. In *Artificial Intelligence and Statistics*, 2016.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, 2016.

Steve Hanneke. Theory of disagreement-based active learning. *Foundations and Trends in Machine Learning*, 2014.

Daniel Hsu. *Algorithms for Active Learning*. PhD thesis, University of California at San Diego, 2010.

Tzu-Kuo Huang, Alekh Agarwal, Daniel Hsu, John Langford, and Robert E. Schapire. Efficient and parsimonious agnostic active learning. In *Advances in Neural Information Processing Systems*, 2015.

Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 2005.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In *International Conference on Machine Learning*, 2012.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *International Conference on Machine Learning*, 2013.

Zohar S Karnin. Verification based solution for structured mab problems. In *Advances in Neural Information Processing Systems*, 2016.

Emilie Kaufmann and Shivaram Kalyanakrishnan. Information complexity in bandit subset selection. In *Conference on Learning Theory*, 2013.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of a/b testing. In *Conference on Learning Theory*, 2014.

Mladen Kolar, Sivaraman Balakrishnan, Alessandro Rinaldo, and Aarti Singh. Minimax localization of structural information in large noisy matrices. *Advances in Neural Information Processing Systems*, 2011.

Akshay Krishnamurthy. Minimax structured normal means inference. *IEEE International Symposium on Information Theory*, 2016.

Shie Mannor and John N Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 2004.

Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. In *Conference on Learning Theory*, 2014.

Serge A. Plotkin, David B. Shmoys, and Éva Tardos. Fast approximation algorithms for fractional packing and covering problems. *Mathematics of Operations Research*, 1995.

Alexander Rakhlin and Karthik Sridharan. Bistro: An efficient relaxation-based method for contextual bandits. In *International Conference on Machine Learning*, 2016.

Daniel Russo. Simple bayesian algorithms for best arm identification. In *Conference on Learning Theory*, 2016.

Max Simchowitz, Kevin Jamieson, and Benjamin Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. In *Conference on Learning Theory*, 2017.

Vasilis Syrgkanis, Haipeng Luo, Akshay Krishnamurthy, and Robert E Schapire. Improved regret bounds for oracle-based adversarial contextual bandits. In *Advances in Neural Information Processing Systems*, 2016.

Shu Wang, Robin R Gutell, and Daniel P Miranker. Biclustering as a method for rna local multiple sequence alignment. *Bioinformatics*, 2007.

Appendix A. Non-interactive analysis

A.1. Proof of Lemma 1

Observe that $\frac{\langle v^\star - v, \hat{\mu}-\mu\rangle}{d(v^\star,v)}$ is the average of $\frac{T}{K}d(v^\star,v)$ centered sub-Gaussian random variables, each with variance parameter 1. This follows because $v^\star - v \in \{-1,0,+1\}$ is non-zero on exactly $d(v^\star,v)$ coordinates, and because $\hat{\mu}-\mu$ is the average of T/K sub-Gaussian random vectors. Therefore, by a Subgaussian tail bound and a union bound

$$\begin{aligned}
\mathbb{P}\left[\exists v \in \mathcal{V} : \frac{|\langle v^\star - v, \hat{\mu}-\mu\rangle|}{d(v^\star,v)} \geq \epsilon\right] &\leq \sum_{v\in\mathcal{V}} \mathbb{P}\left[\frac{|\langle v^\star - v, \hat{\mu}-\mu\rangle|}{d(v^\star,v)} \geq \epsilon\right] \\
&\leq 2\sum_{v\in\mathcal{V}} \exp\left(\frac{-Td(v^\star,v)\epsilon^2}{2K}\right) \\
&= 2\sum_{k=\Psi}^{K} |\mathcal{B}(k,v^\star)| \exp\left(\frac{-Tk\epsilon^2}{2K}\right) \\
&\leq 2K \exp\left(\max_{\Psi\leq k\leq K} \log|\mathcal{B}(k,v^\star)| - \frac{Tk\epsilon^2}{2K}\right).
\end{aligned}$$

Unpacking the definitions of Ψ, Φ and setting $\epsilon = \sqrt{\frac{2K}{T}\left(\Phi + \frac{\log(2K/\delta)}{\Psi}\right)}$, this bound is at most δ, which proves the lemma.

A.2. Proof of upper bound in Theorem 2

Recall the definition of the gap $\Delta_v(\mu)$, and observe that probability of error for the MLE $\hat{v} = \text{argmax}_{v\in\mathcal{V}} \langle \hat{\mu}, v\rangle$ is

$$\begin{aligned}
\mathbb{P}\left[\hat{v} \neq v^\star\right] &\leq \mathbb{P}\left[\exists v \in \mathcal{V} : \langle \hat{\mu}, v - v^\star\rangle > 0\right] = \mathbb{P}\left[\exists v \in \mathcal{V} : \langle \hat{\mu} - \mu, v - v^\star\rangle > \langle \mu, v^\star - v\rangle\right] \\
&= \mathbb{P}\left[\exists v \in \mathcal{V} : \frac{\langle \hat{\mu} - \mu, v - v^\star\rangle}{d(v, v^\star)} > \Delta_v(\mu)\right] \\
&\leq \mathbb{P}\left[\exists v \in \mathcal{V} : \frac{\langle \hat{\mu} - \mu, v - v^\star\rangle}{d(v, v^\star)} > \min_{v\neq v^\star} \Delta_v(\mu)\right].
\end{aligned}$$

Now the result follows from Lemma 1, specifically by setting the right hand side of the normalized regret inequality to be at most $\min_{v\neq v^\star} \Delta_v(\mu)$ and solving for T.

A.3. Proof of lower bound in Theorem 2

The proof here is based of Fano's inequality and follows the analysis of Krishnamurthy (2016). Let us simplify notation and define $P_v = \mathbb{P}_{\mu=\Delta(2v-\mathbf{1})}$ to be the distribution where T/K samples are drawn from each arm and $\mu = \Delta(2v - \mathbf{1})$. For any distribution π supported on $\mathcal{V}$ let P_π denote the mixture distribution where first $v^\star \sim \pi$ and then the samples are drawn from $P_{v^\star}$. With this notation, Fano's inequality (with non-uniform prior) shows that for any algorithm

$$\sup_{v^\star\in\mathcal{V}} \mathbb{P}_{v^\star}[\hat{v} \neq v^\star] \geq \mathbb{E}_{v^\star\sim\pi}\mathbb{P}_{\mu=\Delta(2v^\star-\mathbf{1})}[\hat{v} \neq v^\star] \geq 1 - \frac{\mathbb{E}_{v\sim\pi} KL(P_v||P_\pi) + \log 2}{H(\pi)}.$$

(This slightly generalizes one standard version of Fano's inequality, where π is uniform over $\mathcal{V}$, so the denominator is $\log|\mathcal{V}|$.) Let $\tilde{v} \in \mathcal{V}$ denote the candidate achieving the maximum in the definition of Φ and define the prior

$$\pi(v) \propto \exp\left(-\frac{T\Delta^2\|\tilde{v} - v\|_2^2}{K}\right).$$

With this definition, the entropy term becomes

$$\begin{aligned}
H(\pi) &= \log\left(\sum_v \exp\left(-\frac{T\Delta^2\|\tilde{v} - v\|_2^2}{K}\right)\right) + \sum_v \pi(v)\frac{T\Delta^2\|v - \tilde{v}\|_2^2}{K} \\
&= \log\left(\sum_v \exp\left(-\frac{T\Delta^2\|\tilde{v} - v\|_2^2}{K}\right)\right) + 2\sum_v \pi(v) KL(P_v||P_{\tilde{v}}).
\end{aligned}$$

Here in the last step we use the definition of the Gaussian KL, and the tensorization property for KL-divergence. As for the KL term in the numerator, it is not too hard to see that

$$\sum_v \pi(v) KL(P_v||P_\pi) \leq \sum_v \pi_v KL(P_v||P_\pi) + KL(P_\pi||P_{\tilde{v}}) = \sum_v \pi_v KL(P_v||P_{\tilde{v}}).$$

Thus, we have proved the lower bound if

$$\sum_v \pi(v) KL(P_v||P_{\tilde{v}}) + \log(2) \leq \frac{1}{2}\log\left(\sum_v \exp\left(-\frac{T\Delta^2\|\tilde{v} - v\|_2^2}{K}\right)\right) + \sum_v \pi(v) KL(P_v||P_{\tilde{v}}).$$

After simple algebraic manipulations, we get

$$\log(2\log(2)) \leq \sum_v \exp\left(-\frac{T\Delta^2}{K} d(v,\tilde{v})\right) = \sum_k \exp\left(\log|\mathcal{B}(k,\tilde{v})| - \frac{Tk\Delta^2}{K}\right).$$

Since the sum dominates the maximum, $\tilde{v}$ realizes the definition of Φ, and since $2\log(2) \leq 3$, we obtain the result. More formally, if

$$T \leq \frac{K}{\Delta^2}\left(\Phi - \log\log 3\right),$$

then

$$\begin{aligned}\sum_k \exp\left(\log|\mathcal{B}(k,\tilde{v})| - \frac{Tk\Delta^2}{K}\right) &\geq \max_k \exp\left(\log|\mathcal{B}(k,\tilde{v})| - \Phi k + \log\log 3\right)\\ &= \exp(\log\log 3) = \log 3.\end{aligned}$$

Thus if T is smaller than above, the minimax probability of error is at least $1/2$.

Remark 7 *As we have discussed, the lower bound identifies the minimax rate up to constants for examples where $\Phi \geq \log(K)/\Psi$, in the moderate confidence regime where $\delta = poly(1/K)$. Obtaining the optimal δ dependence even for non-interactive algorithms, seems quite challenging and is an intriguing technical question for future work.*

Remark 8 *We emphasize here that the lower bound applies only for non-interactive algorithms and only in the homogeneous case. A more refined instance-dependent bound is possible with our technique but is not particularly illuminating.*

Appendix B. Calculations for examples

B.1. Instantiations of Theorem 4

To instantiate Theorem 4 for the examples, we need to compute Φ, Ψ, and Δ_a for each arms a. In the homogeneous case, we always have $\Delta_a = \Delta$. We now compute Φ, Ψ for the four examples.

For TOP-K, we have $\Psi = 2$ and

$$\Phi = \max_k \frac{\log\left(\binom{K-s}{k}\binom{s}{s-k}\right)}{k} \leq O\left(\log(K)\right).$$

Thus the sample complexity is $T \lesssim \frac{K}{\Delta^2}$, where recall that $\lesssim$ ignores logarithmic factors.

For DISJSET, it is easy to see that $\Phi \asymp 1/(2s)$, $\Psi = 2s$, so we have $T \lesssim \frac{K}{s\Delta^2}$.

For MATCHING, $\Psi = 4$ since we must switch at least two edges to produce another perfect matching. To calculate Φ, by symmetry we may assume that the "center" v is the identity matching $\{(a_1,b_1)\}_{i=1}^{\sqrt{K}}$ where $\{a_i\}, \{b_i\}$ form the two partition cells. Then, as an upper bound, the number of matchings that differ on $4s$ edges is at most $\binom{\sqrt{K}}{s}s! \leq K^s$. (Actually this bounds the ball volume and hence the sphere volume.) Thus $\Phi \leq O(\log(K))$ and so we have $T \lesssim \frac{K}{\Delta^2}$.

For BICLIQUE, $\Psi = 2\sqrt{s}$ which arises by swapping a single node on either side of the partition. The computation of Φ is more involved. The idea is that for every vertex that we swap into the biclique, we switch $\Theta(\sqrt{s})$ edges, formally at least $\sqrt{s}/2$ edges but no more than $\sqrt{s}$. Then rather than optimizing over the radius in the decision set, we optimize over the number of vertices swapped in on both sides of the partition, which we denote s_L, s_R. For a set v, note that we can obtain any set by swapping s_L column and s_R rows of v

$$\max_k \frac{1}{k}\left(\log|\mathcal{B}(k,v)|\right) \leq \max_{s_R,s_L} \frac{2}{\sqrt{s}(s_L+s_R)} \log\left(\binom{\sqrt{K}}{s_L}\binom{\sqrt{K}}{s_R}\right) \leq O\left(\frac{\log(K)}{\sqrt{s}}\right).$$

This gives $T \lesssim \frac{K}{\sqrt{s}\Delta^2}$.

B.2. Comparison with Chen et al. (2014).

To compare with Chen et al. (2014), we must introduce some of their definitions. Translating to our terminology, they define $\Delta_a^{(C)} = \min_{v:a\in v\ominus v^\star}\langle \mu, v^\star - v\rangle$, which differs from our definition since it is not normalized. They also define exchange classes and a notion of width of the decision set. An exchange class is a collection of patches $b = (b_+, b_-)$ where $b_+, b_- \subset [K]$ and $b_+ \cup b_- = \emptyset$, with several additional properties. To describe them further define the operator $v \oplus b = (v \setminus b-) \cup b_+$ and $v \oslash b = (v \setminus b_+) \cup b_-$ where v is interpreted as a subset of $[K]$. Then a set of patches $\mathcal{B}$ is an exchange class for $\mathcal{V}$ if for every pair $v \neq v' \in \mathcal{V}$ and every $a \in v \setminus v'$, there exists a patch $b \in \mathcal{B}$ such that (1) $a \in b_-$, (2) $b_+ \subset v' \setminus v$, (3) $b_- \subset v \setminus v'$, (4) $v \oplus b \in \mathcal{V}$, and (5) $v' \oslash b \in \mathcal{V}$. Then they define the width

$$\text{width}(\mathcal{V}) = \min_{\text{exchange classes } \mathcal{B}} \max_{b\in\mathcal{B}} |b_-| + |b_+|$$

With these definitions, the fixed-confidence bound of Chen et al. (2014) is

$$\tilde{O}\left(\text{width}(\mathcal{V})^2 \sum_a \frac{1}{(\Delta_a^{(C)})^2} \log(K/\delta)\right)$$

where we have omitted a logarithmic dependence on the arm complexity parameter $\Delta_a^{(C)}$.

For homogeneous DISJSET it is easy to see that width$(\mathcal{V}) \asymp s$ and $\Delta_a^{(C)} \asymp s\Delta$. Hence their bound is $O(K\log(K)/\Delta^2)$.

For MATCHING, number the vertices on one side $a_1, \ldots, a_{\sqrt{K}}$ and on the other side $b_1, \ldots, b_{\sqrt{K}}$. Let $v^\star$ be the matching with edges $\{(a_i, b_i)\}_{i=1}^{\sqrt{K}}$. In the homogeneous case where $\mu = \Delta(2v^\star - 1)$, it is easy to see that $\Delta_a^{(C)} = \Theta(\Delta)$ since for every edge e (which correspond to the arms in the bandit problem), there exist a matching that contains this edge, that disagrees with $v^\star$ on exactly two edges. Specifically, if $e = (a_i, b_j)$ then the matching that has edge (a_k, b_k) for all $k \neq i, j$ and edges (a_i, b_j) and (a_j, b_i) has symmetric set difference exactly 4.

On the other hand we argue that the width is $\Theta(\sqrt{K})$. This is by the standard augmenting path property of the matching polytope. In particular if $v^\star$ is as above and we define another matching $v = \{(a_i, b_{i+1 \mod \sqrt{K}})\}_{i=1}^{\sqrt{K}}$, then the only patch for $v^\star, v$ is to swap all edges. Hence the bound of Chen et al. (2014), in this instance is $\tilde{O}\left(\frac{K^2}{\Delta^2}\log(K/\delta)\right)$ which is a factor of K worse than the non-interactive algorithm in this setting.

For BICLIQUE, we have width$(\mathcal{V}) \asymp s$ yet $\Delta_a^{(C)} \asymp \sqrt{s}\Delta$. For the former, consider two bicliques v, v' that disagree on all nodes on both sides of the partition. Then the smallest patch betweeen them is the trivial one (v, v') since any other potential patch covers edges between the two bicliques (which are not contained in either one). For the latter, for any edge a we can swap at most two nodes, one from each side, to cover this edge. Thus their bound is $O(Ks/\Delta^2)$.

For the worst case comparison, note that

$$\Delta_a = \min_{v:a\in v\ominus v^\star} \frac{\langle \mu, v^\star - v\rangle}{d(v^\star, v)} \geq \min_{v:a\in v\ominus v^\star} \frac{\langle \mu, v^\star - v\rangle}{D} = \frac{\Delta_a^{(C)}}{D}$$
$$\Delta_a \leq \frac{\langle \mu, v^\star - v_a\rangle}{d(v^\star, v_a)} \leq \frac{\Delta_a^{(C)}}{\Psi},$$

where v_a is the set that witnesses $\Delta_a^{(C)}$ and $D \triangleq \max_{u,v\in\mathcal{V}} d(u, v)$ is the diameter. Ignoring logarithmic factors, our bound therefore satisfies

$$T_{\text{chen14}} \frac{\Psi^2}{\text{width}^2}\Lambda \leq T \leq T_{\text{chen14}} \frac{D^2}{\text{width}^2}\Lambda.$$

B.3. Comparison with Chen et al. (2017).

As for Chen et al. (2017), their guarantee is

$$\tilde{O}\left(\text{Low}(\mathcal{V})(\log(1/\delta) + \log|\mathcal{V}|)\right),$$

ignoring some logarithmic factors. Here $\text{Low}(\mathcal{V})$ is the solution to the optimization problem

$$\text{minimize} \sum_a \tau_a \text{ s.t. } \sum_{a\in v\ominus v^\star} \frac{1}{\tau_a} \leq \langle \mu, v^\star - v\rangle^2, \forall v \neq v^\star \text{ and } \tau_a \geq 0, \forall a \in [K]. \tag{6}$$

In the homogeneous case for bipartite matching, we show that $\text{Low}(\mathcal{V}) = \Theta(K/\Delta^2)$. This proves what we want since $\log(\mathcal{V}) \asymp \sqrt{K}$ and hence the bound is a factor of $\sqrt{K}$ worse than Theorem 2.

The proof here is by passing to the dual of Program (6). First we construct the Lagrangian

$$\mathcal{L}(\tau, \alpha) = \sum_a \tau_a + \sum_v \alpha_v \left(\sum_{a\in v\ominus v^\star} \frac{1}{\tau_a} - \langle \mu, v^\star - v\rangle^2\right).$$

By weak duality, the solution of the primal problem is always lower bounded by the solution of the dual problem

$$\min_\tau \max_\alpha \mathcal{L}(\tau, \alpha) \geq \max_\alpha \min_\tau \mathcal{L}(\tau, \alpha).$$

Taking the derivative with respect to τ we have

$$\frac{\partial \mathcal{L}}{\partial \tau_a} = 1 - \sum_{v:a\in v\ominus v^\star} \alpha_v \left(\frac{1}{\tau_a^2}\right) = 0 \Rightarrow \tau_a = \sqrt{\sum_{v:a\in v\ominus v^\star} \alpha_v},$$

and plugging back into the Lagrangian gives

$$\max_{\alpha_v \succeq 0} \sum_a \sqrt{\sum_{v:a\in v\ominus v^\star} \alpha_v} + \sum_v \alpha_v \left(\sum_{a\in v\ominus v^\star} \frac{1}{\sqrt{\sum_{v':a\in v'\ominus v^\star} \alpha_{v'}}} - \langle \mu, v^\star - v\rangle^2\right)$$
$$= \max_{\alpha_v \succeq 0} 2\sum_a \sqrt{\sum_{v:a\in v\ominus v^\star} \alpha_v} - \sum_v \alpha_v \langle \mu, v^\star - v\rangle^2. \tag{7}$$

By weak duality, any feasible solution here provides a lower bound on $\text{Low}(\mathcal{V})$. For matchings, we construct a feasible solution in a similar way to the construction we used to analyze the bound of Chen et al. (2014). Let $v^\star$ be the matching $\{(a_i, b_i)\}_{i=1}^{\sqrt{K}}$. For every edge (a_i, b_j), there is a unique matching v that disagrees with $v^\star$ on exactly 4 edges, and for these matchings we will set α_v to some constant value α. We set $\alpha_v = 0$ otherwise. This ensures that for every $a \notin v^\star$, $\sum_{v:a\in v\ominus v^\star} \alpha_v = \alpha$. On the other hand, for $a \in v^\star$, we get $\sum_{v:a\in v\ominus v^\star} \alpha_v = (\sqrt{K}-1)\alpha \geq \alpha$, since we can swap out this edge with one of $\sqrt{K}-1$ other edges, iterating over all other nodes on the other side of the partition. In other words, for every arm $a \in [K]$, the first term is at least $\sqrt{\alpha}$, while no more than K α_vs are non-zero. In total, a lower bound on the dual program is given by

$$\max_{\alpha\geq 0} 2K\sqrt{\alpha} - 4K\Delta^2\alpha.$$

This simpler program is optimized with $\alpha = 1/(16\Delta^4)$ and plugging back in reveals that

$$\text{Low}(\mathcal{V}) = \Omega(K/\Delta^2).$$

This is all we need for our comparison, since $\text{Low}(\mathcal{V})\log(|\mathcal{V}|) = \Omega(K^{3/2}/\Delta^2)$ in this case.

For BICLIQUE, let us assume that $\sqrt{s}$ divides $\sqrt{K}$. Considering the dual program (7), we set α_v in the following way: We define the set that contains all the hypotheses that can be obtained by swapping the first row or the first column of $v^\star$ with another row or column to be set V_c. Note that $|V_c| = 2(\sqrt{K}-\sqrt{s})$. For $v \in V_c$, we set $\alpha_v = \alpha_1$. We define a maximum set of disjoint hypotheses that does not share any rows or columns with $v^\star$ to be V_q. Note that $|V_q| = (\frac{\sqrt{K}-\sqrt{s}}{\sqrt{s}})^2$. For $v \in V_q$ we set $\alpha_v = \alpha_2$. We discard all the other sets and set $\alpha_v = 0$. For the remaining sets, note that for each arm $a \notin v^\star$, there is only one hypothesis v such that that $a \in v^\star \ominus v$, let us call it v_a. Thus the dual program (7) can be lower bounded as follow

$$(7) \geq \max_{\alpha_1,\alpha_2} \sum_{\{a|v_a\in V_c\}} \sqrt{\alpha_1} + \sum_{\{a|v_a\in V_q\}} \sqrt{\alpha_2} - \sum_{v\in V_c} \alpha_1 s\Delta^2 - \sum_{v\in V_q} \alpha_2 s^2\Delta^2 \tag{8}$$

Note that $|\{a \mid v_a \in V_c\}| = 2(\sqrt{K}-\sqrt{s})\sqrt{s}$ and $|\{a \mid v_a \in V_q\}| = (\sqrt{K}-\sqrt{s})^2$. Solving for α_1, α_2 gives $\alpha_1 = 1/s\Delta^4$, $\alpha_2 = 1/s^2\Delta^4$, and plugging these back into the dual gives

$$\text{Low}(\mathcal{V}) = \Omega\left(\frac{1}{\Delta^2}\left(\sqrt{K}-\sqrt{s}+\frac{(\sqrt{K}-\sqrt{s})^2}{s}\right)\right).$$

Recall that their sample complexity is $\text{Low}(\mathcal{V})\log(\mathcal{V})$, where $\log(\mathcal{V}) \asymp \sqrt{s}$. This means that their sample complexity is lower bounded as

$$\frac{(\sqrt{K}-\sqrt{s})\sqrt{s}}{\Delta^2} + \frac{(\sqrt{K}-\sqrt{s})^2}{\sqrt{s}\Delta^2} = \Omega\left(\frac{1}{\Delta^2}\left(\sqrt{Ks}+\frac{K}{\sqrt{s}}\right)\right)$$

For the worst case comparison, notice that if we set $\tau_i = \infty, \forall i \neq a$, we obtain a lower bound for τ_a in each of the constraints. Specifically, we get $\tau_a = \max_{v,a\in v^\star\ominus v} \frac{1}{\langle \mu, v^\star - v\rangle^2} > \frac{1}{\langle \mu, v^\star - v_a\rangle^2}$, where v_a is the set that witness our complexity for arm a. Thus $\tau_a \geq \frac{1}{D^2\Delta_a^2}$, so the worst case ratio between our bound and their bound is $\frac{D^2}{\log|\mathcal{V}|}\Lambda$. On the other hand, it is easy to see that with $\tau_a = \frac{1}{\Delta_a^2}$, using our definition of arm complexity, their program is feasible, and so we have

$$\text{Low}(\mathcal{V}) \leq \sum_a \frac{1}{\Delta_a^2},$$

which readily yields a lower bound on our complexity, in terms of theirs.

B.4. Comparison with Gabillon et al. (2016).

In Gabillon et al. (2016), the authors introduce a improved gap by defining the complement of a set. Intuitively the complement is the easiest set to compare with. For any set $v \neq v^\star$, the gap is

$$\Delta_v^{(G)} = \max_{v', \langle \mu, v'-v\rangle > 0} \frac{\langle \mu, v'-v\rangle}{d(v', v)},$$

and the set that achieves this maximum is the *complement* of v. A tie breaks in favor of the sets that are closer to v. The gap of an arm a is $\Delta_a^{(G)} = \min_{v, a\in v\ominus v^\star} \Delta_v^{(G)}$, and their sample complexity is

$$O\left(\sum_a \frac{1}{(\Delta_a^{(G)})^2}\log(K/\delta)\right),$$

which is similar to the sample complexity of Chen et al. (2014) except the width is absorbed into the new gap definition. As a consequence, this bound is never worse than Chen et al. (2014).

In the homogeneous BICLIQUE example, it is easy to see $\Delta_a^{(G)} = \Delta$, since taking $v' = v^\star$ will always achieve the maximum. Hence the bound becomes $O(\frac{K}{\Delta^2}\log(K/\delta))$, which is $\Omega(\sqrt{s})$ worse than the bound in Theorem 2.

For general μ, note that for any set v,

$$\Delta_v^{(G)} \geq \Delta_v = \frac{\langle \mu, v^* - v\rangle}{d(v^*, v)} \geq \frac{\langle \mu, C_v - v\rangle}{d(C_v, v)}\frac{d(C_v, v)}{d(v^*, v)} \geq \Delta_v^{(G)}\frac{\Psi}{D},$$

where C_v is the complement of v. Hence the bounds satisfy

$$T_{\text{Gabillon16}} \cdot \Lambda \leq T \leq T_{\text{Gabillon16}} \cdot \frac{D^2}{\Psi^2}\Lambda.$$

Appendix C. Proofs

In this section we provide the proofs of Theorem 3 and Theorem 4. Several lemmas and their proofs are provided in Appendix D.

Proof of Theorem 3. We repeatedly use the following identity for the ℓ_1 norm: For any $u \in \{0,1\}^K$ and any $x \in [0,1]^K$,

$$\|x - u\|_1 = \langle x + u, \mathbf{1}\rangle - 2\langle x, u\rangle. \tag{9}$$

This identity reveals that the disagreement region, $\mathcal{V}_t$, is polyhedral and hence Program (4) is just a linear feasibility problem. Now suppose that Program (4) is feasible and that $x^\star \in \text{conv}(\mathcal{V})$ is a feasible point. Then for every distribution $p \in \Delta(\mathcal{V})$, $x^\star$ satisfies the linear combination of the constraints weighted by $p \in \Delta(\mathcal{V})$, which is precisely what we check by solving Problem (5) and examining the objective value in line 11 in each iteration of the algorithm. Hence by contraposition, if the algorithm ever detects infeasibility, it must be correct.

For the other direction, we use the regret bound for Follow-the-Perturbed Leader (Kalai and Vempala, 2005). Succinctly, when the learner makes decisions $d_t \in \mathcal{D} \subset \mathbb{R}^d$ and the adversary chooses losses $\ell_t \in \mathcal{S} \subset \mathbb{R}^d$, FTPL with parameter $\epsilon \leq 1$ guarantees

$$\mathbb{E}\sum_{t=1}^{T}\langle d_t, \ell_t\rangle - \min_{d\in\mathcal{D}}\sum_{t=1}^{T}\langle d, \ell_t\rangle \leq \epsilon RAT + D/\epsilon,$$

where $D = \max_{d,d'\in\mathcal{D}}\|d - d'\|_1$, $R = \max_{d\in\mathcal{D},\ell\in\mathcal{S}}|\langle d, s\rangle|$, and $A = \max_{\ell\in S}\|\ell\|_1$. Setting $\epsilon = \sqrt{D/(RAT)}$ gives $2\sqrt{DRAT}$ regret. The algorithm chooses d_t by sampling $\sigma_t \sim \mathrm{Unif}([0, 1/\epsilon]^d)$ and playing $d_t = \mathrm{argmin}_{d\in\mathcal{D}}\langle d, \sigma_t + \sum_{\tau=1}^{t-1}\ell_\tau\rangle$. This induces a distribution over decisions d_t, which we denote by $p_t \in \mathbb{R}^d$ and the expectation accounts for this randomness. It will be important for us that FTPL can accommodate adaptive adversaries, and hence the loss ℓ_t can depend on p_t but not on the random decision d_t.

In our case, we have $\mathcal{D} = \mathrm{conv}(\mathcal{V})$, and we write $\ell_t = \Delta\mathbf{1} - 2\Delta x_t - \hat{\mu}$ where x_t is the solution to Program (5) in the t^{th} iteration. This makes $D \leq K$. Recall that $\hat{\mu}$ is the empirical average of y_t. By Chernoff bound and a union bound, with probability at least $1 - \delta_1$, $\|\hat{\mu} - \bar{\mu}\|_1 < K\sqrt{2\log(2K/\delta_1)}$. Since $\bar{\mu} \in [-1, 1]^K$, $x_t \in [0, 1]^K$, $\Delta \in [0, 1]$, we get $A, R \leq 5K\sqrt{\log(2K/\delta_1)}$ in our reduction. So with $\epsilon = \sqrt{1/(25KT\log(2K/\delta_1))}$ the regret is upper bounded by $2K\sqrt{25KT\log(2K/\delta_1)}$. Note that while x_t and hence ℓ_t depends on the random choices of the learner through $\tilde{p}_t$, we will actually apply the regret bound only on the expectation, which we denote by p_t, which can be equivalently viewed as the adversary sampling to generate $\tilde{p}_t$ and ℓ_t. To translate from p_t to $\tilde{p}_t$ we need one final lemma, which we prove in Appendix D.

Lemma 9 *Let $p_t = \mathbb{E}_{\sigma_t}\tilde{p}_t$ and let ℓ_t be any vector, which may depend on $\tilde{p}_t$. Then with probability at least $1 - \delta$, simultaneously for all rounds $t \in [T]$*

$$\left|\sum_{u\in\mathcal{V}}(\tilde{p}_t(u) - p_t(u))\langle u, \ell_t\rangle\right| \leq 3K\sqrt{\frac{\log(2KT/\delta)}{2m}}.$$

Now, we condition on the event in Lemma 9 and use the fact that x_t optimizes Program (5) (which is defined by $\tilde{p}_t$) and passes the check in line 11. Applying the FTPL regret bound, we get

$$\begin{aligned}
0 &\leq \sum_{t=1}^{T}\frac{1}{m}\sum_{i=1}^{m}\Delta\langle x_t, \mathbf{1} - 2u_{t,i}\rangle + \langle x_t, \hat{\mu}\rangle + \sum_{i=1}^{m}\langle u_{t,i}, \Delta\mathbf{1} - \hat{\mu}\rangle \\
&= \sum_{t=1}^{T}\sum_{u\in\mathcal{V}}\tilde{p}_t(u)\langle u, \ell_t\rangle + \langle x_t, \hat{\mu} + \mathbf{1}\Delta\rangle \\
&\leq \sum_{t=1}^{T}\langle x_t, \Delta\mathbf{1} + \hat{\mu}\rangle + \min_{u}\sum_{t=1}^{T}\langle u, \ell_t\rangle + 2K\sqrt{25KT\log(2K/\delta_1)} + 3TK\sqrt{\frac{\log(2KT/\delta_2)}{2m}}.
\end{aligned}$$

Note that we apply the regret bound on p_t, the expected decision of the algorithm, rather than on $\tilde{p}_t$, the randomized one. Setting $\delta_1 = \delta_2 = \delta/2$, dividing through by T, and using (9), we get

$$\forall u \in \mathcal{V}, \langle u - \bar{x}, \hat{\mu} \rangle \leq \Delta \|\bar{x} - u\|_1 + 2K\sqrt{25K\log(4K/\delta)/T} + 3K\sqrt{\frac{\log(4KT/\delta)}{2m}},$$

where $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$. The theorem follows by our choices for T and m. In particular, the number of oracle calls is $Tm = O\left(K^6/\Delta^4 \cdot \log^2(K/\delta)\log(K\log(K/\Delta)/\delta)\right)$.

Proof of Theorem 4. The key lemma in the proof of Theorem 4 is a uniform concentration inequality on the empirical mean $\hat{\mu}$ used by the algorithm. To state the inequality let $\bar{\mu}_t(a) \in \mathbb{R}^d$ be the conditional mean of $y_t(a)$, conditioning on all randomness up to round t, including the execution of DISAGREE. This means that $\bar{\mu}_t(a)$ is either $\mu(a)$ or $2\hat{v}_t(a) - 1$, depending on the outcome of the disagreement check. Recall the definition of Δ_t in Algorithm 1. We first derive a concentration inequality relating $\hat{\mu}_t$ to the empirical means $\bar{\mu}_t$:

Lemma 10 *With the above definitions, for any $\delta \in (0,1)$, with probability at least $1 - \delta/2$*

$$\forall t > 0, \forall v \in \mathcal{V}, \left|\frac{1}{t}\sum_{i=1}^{t}\langle v^\star - v, \bar{\mu}_i - y_i\rangle\right| \leq d(v^\star, v)\Delta_t.$$

This concentration inequality is not challenging to prove, but is much sharper than ones used in prior work. The key difference is that our inequality is a *regret inequality* in the sense that it only bounds differences with the true optimum $v^\star$, while the prior results bound differences between all pairs of hypotheses. Our definition of the version space $\mathcal{V}(\hat{\mu}, \Delta)$ enables using this concentration inequality, which leads to our sample complexity guarantees.

Define the event $\mathcal{E}$ to be the event that Lemma 10 holds and also that the disagreement computation succeeds at all rounds for all arms, which by Theorem 3 happens with probability $1 - \sum_{t>0}\frac{\delta}{t^2\pi^2} \geq 1 - \delta/2$. Under this event, we establish two facts:

1. $\forall t, v^\star \in \mathcal{V}_t$ where $\mathcal{V}_t = \mathcal{V}(\hat{\mu}_t, \Delta_t)$ is the version space at round t (Lemma 11).
2. If $\Delta_t \leq \Delta_a/3$, then arm a will never be queried again (Lemma 13).

The correctness of the algorithm follows from the first fact. In detail, the algorithm only terminates at round t if for all arms, Algorithm 2 detects infeasibility. By Theorem 3, this means that $\mathcal{V}_t \cap \mathcal{V} = \{\hat{v}_t\}$, and, by Lemma 11, we must have $v^\star \in \mathcal{V}_t$. Thus conditioned on $\mathcal{E}$, the algorithm returns $v^\star$.

For the sample complexity, from the second fact and the definition of Δ_t, arm a will not be sampled once

$$t \geq \frac{72}{\Delta_a^2}\left(\Phi + \frac{\log(K\pi^2 t^2/\delta)}{\Psi}\right).$$

A sufficient condition for this transcendental inequality to hold is (see Fact 14):

$$T_a \geq \frac{144}{\Delta_a^2}\left(\Phi + \frac{2\log(144/(\Delta_a^2\Psi)) + 2\log(K\pi^2/\delta)}{\Psi}\right).$$

The sample complexity is at most $\sum_a T_a$, which proves the theorem.

Appendix D. Proofs for the lemmas

Proof [Proof of Lemma 9] Let V be a $\mathbb{R}^{K\times|\mathcal{V}|}$ matrix whose columns are the vectors $v \in \mathcal{V}$. Recall that $p_t \in \Delta(\mathcal{V})$ is a distribution over the perturbed leader at round t. Let $S_i \in \{0,1\}^{|\mathcal{V}|}$ be the indicator vector of the i^{th} sample. Clearly, $\mathbb{E}[S_i] = p_t$ and $\hat{p}_t = \frac{1}{m}\sum_{i=1}^m S_i$. We have

$$\left|\sum_{u\in\mathcal{V}} (\hat{p}_t(u) - p_t(u))\,\langle u, \ell_t\rangle\right| = |\langle V\hat{p}_t - Vp_t, \ell_t\rangle| \le \|V\hat{p}_t - Vp_t\|_\infty \|\ell_t\|_1.$$

Let $()_j$ denote the j-th coordinate of a vector. By Hoeffding's inequality and union bound we have

$$\mathbb{P}\left[\forall t \in [T], \forall j \in [K],\ |(V\hat{p}_t)_j - (Vp_t)_j| \ge \epsilon\right] \le 2KT\exp\left(-2m\epsilon^2\right),$$

so that with probability at least $1-\delta$

$$\forall t \in [T],\ \|V\hat{p}_t - Vp_t\|_\infty \le \sqrt{\frac{\log\left(2KT/\delta\right)}{2m}}.$$

This proves the lemma. ■

Proof [Proof of Lemma 10] Let $\mathcal{F}_t$ be the σ-algebra conditioning on all randomness up to and including the execution of DISAGREE for all arms $a \in [K]$ at round t. Thus $y_t(a)$ is $\mathcal{F}_t$ measurable and with $Z_t = \sum_{i=1}^t (\bar{\mu}_i - y_i)$ it is not hard to see that $\{Z_t\}_{t=1}^T$ forms a vector-valued martingale adapted to the filtration $\{\mathcal{F}_t\}_{t=0}^T$:

$$\mathbb{E}[Z_t|\mathcal{F}_t] = \mathbb{E}[(\bar{\mu}_t - y_t) + Z_{t-1}|Z_{t-1}] = Z_{t-1}.$$

Observe also that $\bar{\mu}_t(a) - y_t(a)$ is a 0-mean Subgaussian random variable with variance parameter at most 1. Thus, for any $v \neq v^\star$, Subgaussian martingale concentration gives

$$\mathbb{P}\left[\left|\sum_{a\in v\ominus v^\star} Z_t(a)/t\right| \ge \epsilon\right] \le 2\exp\left\{-\frac{t\epsilon^2}{8d(v,v^\star)}\right\}.$$

With a union bound, we get

$$\mathbb{P}\left[\exists t, \exists v\in\mathcal{V}, \left|\sum_{a\in v\ominus v^\star} Z_t(a)/t\right| \ge \epsilon_t(v,v^\star,\delta)\right] \le 2\sum_{t>0}\sum_{v\in\mathcal{V}} \exp\left\{-\frac{t\epsilon_t(v,v^\star,\delta)^2}{8d(v,v^\star)}\right\}.$$

Following the argument in the proof of Theorem 2, this right hand side will be at most $\delta/2$ if

$$\epsilon_t(v^\star, v, \delta) = \sqrt{\frac{8d(v^\star,v)}{t}\log\left(\frac{Kt^2\pi^2|\mathcal{B}(d(v^\star,v),v^\star)|}{\delta}\right)}.$$

We set $\Delta_t = \sqrt{\frac{8}{t}\left(\Phi + \log(K\pi^2t^2/\delta)/\Psi\right)}$ so that for all $v \in \mathcal{V}$, $\Delta_t d(v,v^\star) > \epsilon_t(v,v^\star,\delta)$, which concludes the proof. ■

Lemma 11 *Recall the definition of $\mathcal{V}_t = \mathcal{V}(\hat{\mu}_t, \Delta_t)$ at round t, with $\mathcal{V}(\hat{\mu}, \Delta)$ defined in (3). Then in event $\mathcal{E}$, we have that $\forall t, v^\star \in \mathcal{V}_t$.*

Proof The proof is by induction. First, we know that if $v^\star \in \mathcal{V}_{t-1}$ then $\langle v^\star - v, \bar{\mu}_t \rangle \geq \langle v^\star - v, \mu \rangle$. This follows since if arm a is queried then $\bar{\mu}_t(a) = \mu(a)$ and if arm a is not queried, we know that $v^\star(a) = \hat{v}_t(a)$ and our hallucination sets $\bar{\mu}_t(a) = 2\hat{v}_t(a) - 1 = 2v^\star(a) - 1$. So if $v^\star(a) = 1$ and $v(a) = 0$, we have $\bar{\mu}_t(a) = 1 \geq \mu(a)$. If $v^\star(a) = 0$ and $v(a) = 1$, we have $\bar{\mu}_t(a) = -1 \leq \mu(a)$. So in both cases we have $(v^\star(a) - v(a))\bar{\mu}_t(a) \geq (v^\star(a) - v(a))\mu(a)$. Thus, if $\forall i \in [t-1], v^\star \in \mathcal{V}_i$ (which is our inductive hypothesis), then by Lemma 10 $\forall v \in \mathcal{V}$

$$\langle v - v^\star, \hat{\mu}_t \rangle \leq \left\langle v - v^\star, \frac{1}{t}\sum_{i=1}^{t} \bar{\mu}_t \right\rangle + \Delta_t d(v, v^\star) \leq \langle v - v^\star, \mu \rangle + \Delta_t d(v, v^\star) \leq \Delta_t d(v, v^\star).$$

By definition of $\mathcal{V}_t$, this proves that $v^\star \in \mathcal{V}_t$. Clearly the base case holds since $v^\star \in \mathcal{V}_0 = \mathcal{V}$. ■

Lemma 12 *Let $x \in conv(\mathcal{V}) = \sum_i \alpha_i v_i$, where $v_i \in \mathcal{V}$, $\sum_i \alpha_i = 1$, $\alpha_i \geq 0$ and let $v \in \mathcal{V}$. Then,*

$$\|x - v\|_1 = \sum_i \alpha_i \|v_i - v\|_1. \tag{10}$$

Proof This follows by integrality of $v \in \mathcal{V}$ and (9). In particular, for integral v, $\|x - v\|$ is actually linear so we can bring the $\sum_i \alpha_i$ outside the ℓ_1 norm. ■

Lemma 13 *Under event $\mathcal{E}$, once t is such that $\Delta_t < \Delta_a/3$, arm a will not be sample again.*

Proof We consider here the case where $v^\star(a) = 1$. For $v^\star(a) = 0$ the analysis is similar. Assume for the sake of contradiction that a is sampled, which means that DIS$(a, 1 - \hat{v}_t(a), \Delta_t, \hat{\mu}_t)$ returns TRUE. If $v^\star(a) = 1$, then $\forall v \in \mathcal{V}$ with $v(a) = 0$ we have

$$\langle v^\star - v, \hat{\mu}_t \rangle \geq \left\langle v^\star - v, \frac{1}{t}\sum_{\tau=1}^{t} \bar{\mu}_\tau \right\rangle - \Delta_t d(v^\star, v) \geq \langle v^\star - v, \mu \rangle - \Delta_t d(v^\star, v) > 2d(v^\star, v)\Delta_t$$

The first inequality is Lemma 10, the second uses the property of the hallucinated samples that we used in Lemma 11. The last inequality is due to $\Delta_a \leq \frac{\langle v^\star - v, \mu \rangle}{d(v^\star, v)}$ and our assumption that $\Delta_t < \Delta_a/3$. This implies that $\hat{v}_t(a) = 1$, which means that we execute DISAGREE to check if any surviving hypothesis $v \in \mathcal{V}_t$ has $v(a) = 0$. Since we sampled arm a, this means there exists $x \in \text{conv}(\mathcal{V})$ such that

$$\forall u \in \mathcal{V} \langle u - x, \hat{\mu}_t \rangle \leq \Delta_t \|u - x\|_1 + \Delta_t.$$

This follows by Theorem 3 which holds under the event $\mathcal{E}$. Now write $x = \sum_i \alpha_i v_i$ where α is a distribution and $v_i \in \mathcal{V}$. Since $x(a) = 0$, we must have $v_i(a) = 0$ for all i. This means that

$$\langle v^\star - x, \hat{\mu}_t \rangle = \sum_i \alpha_i \langle v^\star - v_i, \hat{\mu}_t \rangle > 2\sum_i \alpha_i d(v^\star, v_i)\Delta_t \geq \Delta_t \|v^\star - x\|_1 + \Delta_t.$$

Algorithm 3 Fixed budget algorithm for combinatorial identification

Input: $\mathcal{V}$, set of arm $[K]$, $\{n_t\}_d$

Set $t \leftarrow 1, A_1 \leftarrow \emptyset, R_1 \leftarrow \emptyset$

for $t = 1, 2, 3, ..., K$ **do**

  Sample arms in $[K]\backslash(A_t \cup R_t)$ for $n_t - n_{t-1}$ times. For $a \in A_t$ use sample value 1 and for $a \in R_t$ use sample value -1 (i.e., hallucinate samples).

  Update $\hat{\mu}_t$ and find $\hat{v}_t = \arg\max_{v \in \mathcal{V}} \langle v, \hat{\mu}_t \rangle$.

  $\hat{a}_t = \text{argmax}_{a \in [K]\backslash(A_t \cup R_t)} \hat{\Delta}_{t,a}$.

  If $\hat{a}_t \in \hat{v}_t$, then $A_{t+1} = A_t \cup \{\hat{a}_t\}, R_{t+1} = R_t$, else $A_{t+1} = A_t, R_{t+1} = R_t \cup \{\hat{a}_t\}$.

end for

return A_{K+1}

The last inequality is due to Lemma 12 and the fact that $\forall i, d(v^\star, v_i) \geq 1$. This contradicts the guarantee in Theorem 3, which means that $\textsc{Dis}(a, 1 - \hat{v}_t(a), \Delta_t, \hat{\mu}_t)$ cannot return $\textsc{True}$. ■

We use the Lemma 8 from Antos et al. (2010).

Fact 14 (Lemma 8 from Antos et al. (2010)) *Let $a > 0$, for any $t \geq \frac{2}{a}\max\{(\log\frac{1}{a} - b), 0\}$, we have $at + b \geq \log t$.*

Appendix E. Proof of Theorem 5

In the fixed budget setting, we follow a classic rejection strategy used by many algorithms in other settings (e.g., Successive Rejects Audibert and Bubeck (2010), SAR Bubeck et al. (2013), CSAR Chen et al. (2014) and also the algorithm of Gabillon et al. (2016)).

We require several new definitions. First recall that our definition of the gap for arm a is Δ_a. Let $\Delta^{(j)}$ be the j^{th} largest element in $\{\Delta_a\}_{a \in [K]}$. Then the main complexity measure is $\tilde{H} = \max_j (K + 1 - j)(\Delta^{(j)})^{-2}$. For short hand we define the partial harmonic sum $\widetilde{\log}(t) = \sum_{i=1}^{t} 1/i$. Assume that the total budget is T, and define

$$n_t = \left\lceil \frac{T - K}{\widetilde{\log}(K)(K + 1 - t)} \right\rceil, \qquad n_0 = 0$$

which will be related to the number of queries issued in each round of our algorithm. As before, let $\hat{\mu}_t$ be the empirical mean at round t of the algorithm and let $\hat{v}_t = \text{argmax}_{v \in \mathcal{V}} \langle v, \hat{\mu}_t \rangle$ be the empirical maximizer. Define the empirical gaps at round t for hypotheses and arms respectively as

$$\hat{\Delta}_{t,v} = \frac{\langle \hat{\mu}_t, \hat{v}_t - v \rangle}{d(\hat{v}_t, v)}, \qquad \hat{\Delta}_{t,a} = \min_{a \in \hat{v}_t \ominus v} \hat{\Delta}_{t,v}.$$

With these definitions, we are now ready to describe the fixed budget algorithms, with pseudocode in Algorithm 3. The algorithm maintains a set of "accepted" and "rejected" arms, A_t and R_t in the pseudocode at round t, and once an arm is marked "accept" or "reject" it is never queried again. At each round t we issue several queries to all surviving arms, ensuring that each arm has n_t total queries, and then we find the arm with the largest empirical gap $\hat{\Delta}_{t,a}$ and accept it if it is included in the ERM $\hat{v}_t$. Otherwise we reject. Note that the algorithm is not oracle efficient, since computing the empirical arm gaps is not amenable to linear optimization.

Proof [Proof of Theorem 5] First, note that in each round we eliminate one arm and sample the rest for $n_t - n_{t-1}$ times. Thus after round t we have sampled each surviving arm n_t times, and exactly one arm is sampled n_i times for each $i \in [K]$. Thus the total number of samples is

$$\sum_{t=1}^{K} n_t = \sum_{t=1}^{K} \left\lceil \frac{T-K}{\widetilde{\log}(K)(K+1-t)} \right\rceil \leq \sum_{t=1}^{K} \frac{T-K}{\widetilde{\log}(K)(K+1-t)} + 1 = T.$$

Second, define $\bar{\mu}_t$ as before to be the mean of the all samples up to and including round t, taking into account the hallucination. $\bar{\mu}_t(a)$ is an average of n_t terms where if at round $i \leq t$ we place $a \in A_i$, then the last $n_t - n_i$ terms are just 1. Similarly if at round n_i we place $a \in R_i$ then the last $n_t - n_i$ terms are -1. Otherwise all terms are simply $\mu(a)$. Formally,

$$\bar{\mu}_t(a) = \frac{1}{n_t} \sum_{\tau=1}^{K} (n_\tau - n_{\tau-1}) \left[\mu \mathbf{1}\{a \notin R_\tau \cup A_\tau\} + \mathbf{1}\{a \in A_\tau\} - \mathbf{1}\{a \in R_\tau\}\right].$$

Note that this is different but related to our definition in the fixed confidence proof. We define the high probability event:

$$\mathcal{E} \triangleq \{\forall t \in [K], \forall v \in \mathcal{V}, |\langle v - v^\star, \hat{\mu}_t - \bar{\mu}_t \rangle| < c d(v, v^\star) \Delta^{(t)}\},$$

where $c < 1$ is a constant that we will set later. Now we show that $\mathcal{E}$ holds with high probability:

$$\begin{aligned}
\mathbb{P}[\bar{\mathcal{E}}] &\leq \sum_t \sum_{v \in \mathcal{V}} \exp\left\{-\frac{c^2 d(v,v^\star)(\Delta^{(t)})^2(T-K)}{\tilde{\log}(K)(K+1-t)}\right\} \leq K \sum_{v \in \mathcal{V}} \exp\left\{-\frac{c^2(T-K)d(v,v^\star)}{\tilde{\log}(K)\tilde{H}}\right\} \\
&\leq K \sum_{k \in [K]} \exp\left\{-\frac{c^2(T-K)k}{\tilde{\log}(K)\tilde{H}} + \log|\mathcal{B}(k, v^\star)|\right\} \leq K^2 \exp\left\{\Psi\left(\Phi - \frac{(T-K)c^2}{\tilde{\log}(K)\tilde{H}}\right)\right\}.
\end{aligned}$$

We proceed to show that, conditioned on event $\mathcal{E}$, $A_{K+1} = v^\star$. At round t, define

$$a_t^\star = \underset{a \in [K] \setminus (A_t \cup R_t)}{\operatorname{argmax}} \Delta_a,$$

where A_t and R_t are the accepted and reject arms at the beginning of round t and Δ_a is the true arm complexity. Further assuming (inductively) that $A_t \subset v^\star$ and $R_t \cup v^\star = \emptyset$, we establish five facts:

Fact 1. At the beginning of round t, $a_t^\star$ satisfies $\Delta_{a_t^\star} \geq \Delta^{(t)}$. If this statement does not hold at round t, then we must have eliminated all of the the t arms $\Delta^{(1)}, \ldots, \Delta^{(t)}$. However, since we eliminate exactly one arm in each round, we can only eliminate $t-1$ arms before round t, which produces a contradiction since $a_t^\star$ is the maximizer.

Fact 2. Under the inductive hypothesis, for all $v \in \mathcal{V}$, we have $\langle \bar{\mu}_t, v^\star - v \rangle \geq \langle \mu, v^\star - v \rangle$. This is similar to the argument we used in the fixed confidence proof. For any arm a, if $a \notin A_t \cup R_t$ then the corresponding terms are equal. If $a \in A_t$ then since by induction we know $a \in v^\star$, the term for $v^\star$ is as high as possible and analogously if $a \in R_t$ the term for $v^\star$ is as low as possible.

Fact 3. $a_t^\star \in \hat{v}_t \iff a_t^\star \in v^\star$. Assume for the sake of contradiction that $a_t^\star \in \hat{v}_t$ and $a_t^\star \notin v^\star$. The proof is the same for the other case. We have

$$\Delta_{a_t^\star} = \min_{a_t^\star \in v^\star \ominus v} \frac{\langle \mu, v^\star - v\rangle}{d(v^\star, v)} \leq \frac{\langle \mu, v^\star - \hat{v}_t\rangle}{d(v^\star, \hat{v}_t)}.$$

Thus we have $\frac{\langle \mu, \hat{v}_t - v^\star\rangle}{d(v^\star, \hat{v}_t)} \leq -\Delta_{a_t^\star}$. By the previous fact we know $\Delta_{a_t} \geq \Delta^{(t)}$ since $a_t^\star$ is the maximizer. Now, conditioned on $\mathcal{E}$:

$$\frac{\langle \hat{\mu}_t, \hat{v}_t - v^\star\rangle}{d(\hat{v}_t, v^\star)} < \frac{\langle \bar{\mu}_t, \hat{v}_t - v^\star\rangle}{d(\hat{v}_t, v^\star)} + c\Delta^{(t)} \leq \frac{\langle \mu, \hat{v}_t - v^\star\rangle}{d(\hat{v}_t, v^\star)} + c\Delta^{(t)} < \Delta^{(t)} - \Delta_{a_t^\star} \leq 0.$$

The first inequality is by event $\mathcal{E}$, the second is by Fact 2 and the final one is by Fact 1 and the definition of $a_t^\star$. This results in a contradiction.

Fact 4. Let $\tilde{v}_{t,a_t^\star}$ be the set that witnesses $\hat{\Delta}_{t,a_t^\star}$, i.e. $\tilde{v}_{t,a_t^\star} = \mathrm{argmin}_{v: a_t^\star \in v \ominus \hat{v}_t} \hat{\Delta}_{t,v}$. We have that $\langle \hat{\mu}_t, v^\star - \tilde{v}_{t,a_t^\star}\rangle > 0$. To see why, note that $a_t^\star \in \hat{v}_t \ominus \tilde{v}_{t,a_t^\star}$ and by Fact 3 we have $a_t^\star \in v^\star \ominus \tilde{v}_{t,a_t^\star}$. Conditioning on $\mathcal{E}$ and using the fact that the true gap $\Delta_{a_t^\star}$ involves minimizing over $v \in \mathcal{V}$ we get

$$\frac{\langle \hat{\mu}_t, v^\star - \tilde{v}_{t,a_t^\star}\rangle}{d(v^\star, \tilde{v}_{t,a_t^\star})} \geq \frac{\langle \bar{\mu}_t, v^\star - \tilde{v}_{t,a_t^\star}\rangle}{d(v^\star, \tilde{v}_{t,a_t^\star})} - c\Delta^{(t)} \geq \frac{\langle \mu, v^\star - \tilde{v}_{t,a_t^\star}\rangle}{d(v^\star, \tilde{v}_{t,a_t^\star})} - c\Delta^{(t)} > \Delta_{a_t^\star} - \Delta^{(t)} \geq 0.$$

The last step here uses Fact 1.

Fact 5. $\hat{a}_t \in \hat{v}_t \iff \hat{a}_t \in v^\star$. Assume for the sake of contradiction that $\hat{a}_t \in \hat{v}_t, \hat{a}_t \notin v^\star$. We have

$$\hat{\Delta}_{t,\hat{a}_t} = \min_{\hat{a}_t \in \hat{v}_t \Delta v} \frac{\langle \hat{\mu}_t, \hat{v}_t - v\rangle}{d(\hat{v}_t, v)} \leq \frac{\langle \hat{\mu}_t, \hat{v}_t - v^\star\rangle}{d(\hat{v}_t, v^\star)}.$$

As above, let $\tilde{v}_{t,a_t^\star}$ be the set that witnesses $\hat{\Delta}_{t,a_t^\star}$. Since $\hat{a}_t$ maximizes $\hat{\Delta}_{t,a}$ over all surviving arms a and since $a_t^\star$ is surviving by definition, we have

$$\begin{aligned}
\hat{\Delta}_{t,\hat{a}_t} &\geq \hat{\Delta}_{t,a_t^\star} = \min_{a_t^\star \in \hat{v}_t \ominus v} \frac{\langle \hat{\mu}_t, \hat{v}_t - v\rangle}{d(\hat{v}_t, v)} = \frac{\langle \hat{\mu}_t, \hat{v}_t - \tilde{v}_{t,a_t^\star}\rangle}{d(\hat{v}_t, \tilde{v}_{t,a_t^\star})} \\
&\geq \frac{\langle \hat{\mu}_t, \hat{v}_t - v^\star\rangle + \langle \hat{\mu}_t, v^\star - \tilde{v}_{t,a_t^\star}\rangle}{d(\hat{v}_t, v^\star) + d(v^\star, \tilde{v}_{t,a_t^\star})} \geq \min\left\{\frac{\langle \hat{\mu}_t, \hat{v}_t - v^\star\rangle}{d(\hat{v}_t, v^\star)}, \frac{\langle \hat{\mu}_t, v^\star - \tilde{v}_{t,a_t^\star}\rangle}{d(v^\star, \tilde{v}_{t,a_t^\star})}\right\} \\
&\triangleq \min\{a, b\}.
\end{aligned}$$

The last inequality holds since both terms in the numerator are non-negative as we have shown above in Fact 4. Since we previously upper bounded $\hat{\Delta}_{t,\hat{a}_t}$ by what we are now calling a, we have $a \geq \min\{a, b\}$. If $a \leq b$, then all of the inequalities are actually equalities, so we must have $a = b$. The other case is that $a > b$, so we can address both cases by considering $a \geq b$. Expanding the definition and applying the concentration inequality, we have

$$b \triangleq \frac{\langle \hat{\mu}_t, v^\star - \tilde{v}_{t,a_t^\star}\rangle}{d(v^\star, \tilde{v}_{t,a_t^\star})} \geq \frac{\langle \mu, v^\star - \tilde{v}_{t,a_t^\star}\rangle}{d(v^\star, \tilde{v}_{t,a_t^\star})} - c\Delta^{(t)} \geq \Delta_{a_t^\star} - c\Delta^{(t)}.$$

On the other hand,

$$a \triangleq \frac{\langle \hat{\mu}_t, \hat{v}_t - v^\star\rangle}{d(\hat{v}_t, v^\star)} \leq \frac{\langle \mu, \hat{v}_t - v^\star\rangle}{d(\hat{v}_t, v^\star)} + c\Delta^{(t)} \leq c\Delta^{(t)}.$$

Both of these calculations also require Fact 2. Setting $c = 1/3$, we have

$$\Delta_{a_t^\star} \leq 2c\Delta^{(t)} < \Delta^{(t)},$$

which contradicts Fact 1 and the definition of a_t.

Wrapping up. To conclude the proof, we proceed by induction. Clearly the base case that $A_0 \subset v^\star$ and $R_0 \cap v^\star = \emptyset$ is true. Now conditioning on $\mathcal{E}$ and assuming the inductive hypothesis, we have that by Fact 5, the arm $\hat{a}_t \in \hat{v}_t \iff \hat{a}_t \in v^\star$. This directly proves the inductive step since the algorithm's rule for accepting an arm agrees with $v^\star$. ■

Appendix F. Proof of Theorem 6

Recall that in the main concentration argument in Lemma 10, we proved that

$$\mathbb{P}\big[\exists t \in \mathbb{N}, \exists v \in \mathcal{V}, |\langle v^\star - v, \hat{\mu}_t - \frac{1}{t}\sum_{i=1}^{t}\bar{\mu}_i\rangle| \geq \epsilon_t(v, v^\star, \delta)\big] \leq 2\sum_{t\in\mathbb{N}}\sum_{v\in\mathcal{V}} \exp\left\{-\frac{t\epsilon_t(v, v^\star, \delta)^2}{8d(v, v^\star)}\right\}.$$

Setting

$$\epsilon_t(v, v^\star, \delta) = \sqrt{\frac{8d(v, v^\star)}{t}\log\frac{|\mathcal{B}(d(v, v^\star), v^\star)|\pi^2 K t^2}{3\delta}},$$

we have that the probability of this event is at most δ. Previously we set each hypothesis to have the same confidence interval Δ_t which provided an upper bound on $\epsilon_t(v, v^\star, \delta)$ for all v. This enabled us to write the disagreement region as a polyhedral set in $\mathcal{V}$, but to obtain a more refined bound, we would like to use $\epsilon_t(v, v^\star, \delta)$ directly. However, note that $\epsilon_t(v, v', \delta) \neq \epsilon_t(v', v, \delta)$ unless the hypothesis space is symmetric. We symmetrize ϵ_t by defining

$$D(v, v') \triangleq \max\{\log|\mathcal{B}(d(v, v'), v)|, \log|\mathcal{B}(d(v, v'), v')|\} = D(v', v),$$

and the symmetric confidence interval

$$\epsilon_t'(v, v', \delta) \triangleq \sqrt{\frac{8d(v, v')}{t}\left(\log\frac{\pi^2 K t^2}{3\delta} + D(v, v')\right)}. \tag{11}$$

Define the hypothesis complexity measures, for $v \neq v^\star$

$$H_v^{(1)} = \frac{d(v, v^\star)}{\langle \mu, v^\star - v\rangle^2}, \qquad H_v^{(2)} = \frac{d(v, v^\star)D(v, v^\star)}{\langle \mu, v^\star - v\rangle^2}.$$

The arm complexity measures, defined previously, are $H_a^{(1)} = \max_{a\in v\ominus v^\star} H_v^{(1)}$ and $H_a^{(2)} = \max_{v:a\in v\ominus v^\star} H_v^{(2)}$. The main difference here is that we are not normalizing by $d(v, v^\star)^2$ as we did in the proof of Theorem 4 but rather just $d(v, v^\star)$. In some sense we replace the term depending on Ψ with $H_a^{(1)}$ and the term depending on Φ with $H_a^{(2)}$.

Algorithm 4 Inefficient fixed confidence algorithm

Input: $\mathcal{V}$, set of arms $[K]$, δ
Set $\mathcal{V}_1 = \mathcal{V}$
for $t = 1, 2, 3, \ldots$ **do**
  $\mathcal{A}_t = \emptyset$
  for $a \in [K]$ **do**
    if $\exists v, v' \in \mathcal{V}_t$ such that $v(a) \neq v'(a)$ **then**
      $\mathcal{A}_t = \mathcal{A}_t \cup a$, query a, set $y_t(a) \sim \mathcal{N}(\mu(a), 1)$
    end if
  end for
  Update $\hat{\mu}_t = \frac{1}{t}\sum_{\tau=1}^{t} y_t$.
  $\mathcal{R}_t \leftarrow \{v \in \mathcal{V}_t \mid \exists u \in \mathcal{V}_t, u \neq v, \langle u - v, \hat{\mu}_t \rangle > \epsilon'_t(u, v, \delta)\}$
  Update $\mathcal{V}_{t+1} \leftarrow \mathcal{V}_t \setminus \mathcal{R}_t$
  If $|\mathcal{V}_{t+1}| = 1$ **return** the single element $v \in \mathcal{V}_{t+1}$.
end for

To prove Theorem 6, we construct an inefficient fixed confidence algorithm, with pseudocode in Algorithm 4. The algorithm is essentially identical to Algorithm 1, except we use the new definition ϵ' in the confidence bounds defining the version space, which forces us to do explicit enumeration. One other minor difference is that we are now explicitly enforcing monotonicity of the version space, so we need not use hallucination as we did before. We now turn to the proof.

Proof [Proof of Theorem 6] In a similar way to Lemma 10 we can prove that

$$\mathbb{P}\left[\forall t, \forall v \in \mathcal{V}_t, |\langle v^\star - v, \hat{\mu}_t - \mu \rangle| > \epsilon'_t(v^\star, v, \delta)\right] \leq 2 \sum_t \sum_{v \in \mathcal{V}} \exp\left\{-\frac{t\epsilon'_t(v, v^\star, \delta)^2}{8d(v, v^\star)}\right\}.$$

The important thing here is that if $v \in \mathcal{V}_t$ then we must query every $a \in v \ominus v^\star$ and moreover since we are explicitly enforcing monotonicity (i.e. $\mathcal{V}_t \subset \mathcal{V}_{t-1}$), we also queried all of these arms in all previous rounds. Thus we are obtaining unbiased samples to evaluate these mean differences. Using the definition of ϵ'_t in (11), this probability is at most δ.

Next we prove that when the algorithm terminates, the output is $v^\star$. We work conditional on the $1-\delta$ event that the concentration inequality holds. We argue that $v^\star$ is never eliminated, or formally $v^\star \notin \mathcal{R}_t$ for all t. To see why observe that $\forall v \in \mathcal{V}_{t-1} \neq v^\star$, we have

$$\langle \hat{\mu}_t, v - v^\star \rangle \leq \langle \mu, v - v^\star \rangle + \epsilon'_t(v, v^\star, \delta).$$

This means that no surviving $v \in \mathcal{V}_t$ can eliminate $v^\star$. This verifies correctness of the algorithm, since $v^\star$ is never eliminated, so it must be the single element in $\mathcal{V}_t$ when the algorithm terminates.

We now turn to the sample complexity. We argue here that if $t > 32H_a^{(1)} \log(\pi^2 K t^2/(3\delta)) + 32H_a^{(2)}$ then from round t onwards, arm a will not be sampled again. This condition on t implies that for all $v \in \mathcal{V}$ such that $a \in v \ominus v^\star$, we have

$$\epsilon'_t(v^\star, v, \delta) \triangleq \sqrt{\frac{8d(v, v^\star)}{t}\left(\log(\pi^2 K t^2/(3\delta)) + D(v, v^\star)\right)} < \langle \mu, v^\star - v \rangle / 2,$$

by the definitions of $H_a^{(1)}$ and $H_a^{(2)}$. Using this simpler fact we argue that a cannot be sampled again. Working toward a contradiction, assume that a is sampled at round $t+1$, which means there exists two hypotheses $v_1, v_2 \in \mathcal{V}_{t+1}$ such that $v_1(a) \neq v_2(a)$. Since $v^\star \in \mathcal{V}_{t+1}$ this implies that there exists $v \in \mathcal{V}_{t+1}$ such that $v^\star(a) \neq v(a)$. But we clearly have

$$\langle v^\star - v, \hat{\mu}_t \rangle \geq \langle v^\star - v, \mu \rangle - \epsilon_t'(v^\star, v, \delta) > \epsilon_t'(v^\star, v, \delta)$$

which is a contradiction since v must have been eliminated at round t. This proves that $a \notin \mathcal{A}_{t+1}$, and since $\mathcal{V}_{t+1}$ is monotonically shrinking, so is $\mathcal{A}_t$, which means that a is never sampled again.

To summarize, we have now shown that for each arm a, the arm will be sampled at most t_a times, where t_a is the smallest integer satisfying

$$t_a \geq 32 H_a^{(1)} \log(\pi^2 K t_a^2/(3\delta)) + 32 H_a^{(2)}.$$

The final result now follows from an application of Fact 14. ∎