Top Feasible Arm Identification

Julian Katz-Samuels
University of Michigan

Clayton Scott
University of Michigan

Abstract

We propose a new variant of the top arm identification problem, *top feasible arm identification*, where there are K arms associated with D-dimensional distributions and the goal is to find m arms that maximize some known linear function of their means subject to the constraint that their means belong to a given set $P \subset \mathbb{R}^D$. This problem has many applications since in many settings, feedback is multi-dimensional and it is of interest to perform *constrained maximization*. We present problem-dependent lower bounds for top feasible arm identification and upper bounds for several algorithms. Our most broadly applicable algorithm, TF-LUCB-B, has an upper bound that is loose by a factor of $O(D\log(K))$. Many problems of practical interest are two-dimensional and, for these, it is loose by a factor of $O(\log(K))$. Finally, we conduct experiments on synthetic and real-world datasets that demonstrate the effectiveness of our algorithms. Our algorithms are superior both in theory and in practice to a naive two-stage algorithm that first identifies the feasible arms and then applies a best arm identification algorithm to the feasible arms.

1 Introduction

In the top arm identification problem in multi-armed bandits, there are K scalar-valued distributions (also referred to as arms) and an agent plays a sequential game where, at each round, the agent chooses (or "pulls") one of the arms and observes an i.i.d. realization from it. At the end of the game, the agent outputs the set of m arms believed to have the largest means. This problem has applications in areas such as crowdsourcing, A/B testing, and clinical trials.

While top arm identification considers settings where the feedback is scalar-valued and the goal is maximization, in many applications, the feedback is multi-dimensional and it is of interest to perform *constrained maximization*. For example, in crowdsourcing, an important challenge is to identify high-quality workers that complete work at a suitable pace (e.g., below 15 seconds on average) and, in clinical trials, it is of interest to efficiently find drugs that are most likely to be effective and have an acceptably low probability of causing an adverse effect.

In this paper, we introduce *top feasible arm identification* for situations where the feedback is multi-dimensional and the goal is constrained maximization. In this problem, there are K arms and each arm i is associated with a D-dimensional distribution ν_i that has mean $\boldsymbol{\mu}_i$. At each round $t = 1, 2, \ldots$, the agent chooses an arm I_t and observes an independent random vector drawn from ν_{I_t}. For given $P \subset \mathbb{R}^D$, $\boldsymbol{r} \in \mathbb{R}^D$, $m \leqslant K$, and $\delta \in (0, 1)$, the goal of the agent is to identify m arms that maximize $\boldsymbol{r}^\top \boldsymbol{\mu}_i$ subject to the constraint $\boldsymbol{\mu}_i \in P$, with probability at least $1 - \delta$, in the fewest number of samples possible.

We make several contributions to this problem. We prove problem-dependent lower bounds for top feasible arm identification. We also propose a family of algorithms TF-LUCB, where each instance is specified by a test for feasibility TestF, and we prove a master theorem that characterizes an upper bound for TF-LUCB in terms of the subroutine TestF. Finally, we use this master theorem to prove upper bounds for several algorithms. Our most broadly applicable algorithm, TF-LUCB-B, has an upper bound that is loose by a factor of $O(D\log(K))$. Many problems of practical interest are two-dimensional and for these, it is loose by a factor of $O(\log(K))$. Notably, our algorithms are superior both in theory and in practice to a naive two-stage algorithm that first identifies the feasible arms and then applies a best arm identification algorithm to the feasible arms. The sample complexity of such a two-stage algorithm can be arbitrarily larger

Proceedings of the 22[nd] International Conference on Artificial Intelligence and Statistics (AISTATS) 2019, Naha, Okinawa, Japan. PMLR: Volume 89.

than the sample complexity of our algorithms and, indeed, in our experiments we improve on such a baseline by as much as a factor of 4.5.

2 Related Work

Top arm identification has received a lot of attention in recent years (Mannor and Tistisklis, 2004; Audibert and Bubeck, 2010; Gabillon et al., 2012; Kalyanakrishnan et al., 2012; Bubeck et al., 2013; Chen et al., 2014; Jamieson et al., 2014). Most work considers the case where arms are scalar-valued and, thus, their results cannot be applied to our problem setting. Recently, Chen et al. (2017) proposed the general sampling problem, which does encompass a variant of top feasible arm identification. Their work differs from ours in several significant ways. First, in the work of Chen et al. (2017), the agent samples from one dimension of one arm at a time, whereas in our setting pulling an arm yields a random D-dimensional vector. Second, Chen et al. (2017) assume that the arms are isotropic Gaussian, whereas we assume each arm is a multi-dimensional sub-Gaussian distribution. Finally, their algorithm (see their Algorithm 7) is impractical for moderate values of δ in the fixed confidence setting since its first stage consists of a uniform allocation strategy that terminates when the confidence bounds of all of the means are small enough to determine which of the arms are in the top feasible m with probability at least 0.99.

Auer et al. (2016) also consider a setting where arms are multi-dimensional. Their goal is to determine the Pareto front of the arms, which is quite different from the task of constrained maximization in top feasible arm identification. We also remark that they use an elimination algorithm, whereas we adapt the LUCB algorithm from Kalyanakrishnan et al. (2012) to our setting.

Recently, Katz-Samuels and Scott (2018) proposed the feasible arm identification problem, in which there are K multi-dimensional distributions and a given polyhedron, and the goal is to determine which of the distributions have means belonging to the polyhedron. By contrast, in top feasible arm identification, the goal is to find a collection of arms whose means are feasible and maximize some linear function. In short, Katz-Samuels and Scott (2018) deal with feasibility while the current paper deals with constrained maximization. Furthermore, whereas Katz-Samuels and Scott (2018) consider the fixed budget setting (in which there is a fixed number of rounds), we consider the fixed confidence setting. These differences require the development of new ideas and algorithms.

We also note that top feasible arm identification differs from best-arm identification in linear bandits (Soare et al., 2014). In best-arm identification in linear bandits, each arm i is associated with a *known* feature vector $\boldsymbol{x}_i$ and the reward of arm i has mean $\boldsymbol{x}_i^\top \boldsymbol{\theta}$ where $\boldsymbol{\theta}$ is unknown. In our setting, each arm is associated with a D-dimensional distribution and the goal is to maximize some known linear function $f : \mathbb{R}^D \longrightarrow \mathbb{R}$ subject to the constraint that $\boldsymbol{\mu}_i \in P$.

3 Problem Statement

Notation. For $n \in \mathbb{N}$, let $[n] = \{1, \ldots, n\}$. Let U be a finite set and f be a scalar-valued function with domain containing U, and define $\max^{(l)}_{x \in U} f(x) \coloneqq$

$$\begin{cases} \max\limits_{\{x \in U : |\{y \in U : f(y) \geqslant f(x)\}| \geqslant l-1\}} f(x) & : |U| \geqslant l \\ -\infty & : \text{ otherwise} \end{cases}.$$

In words, $\max^{(l)}_{x \in U} f(x)$ is the value of the lth largest $x \in U$ under $f(\cdot)$ and if $|U| < l$, then it is $-\infty$. For a set $A \subset \mathbb{R}^D$, let ∂A denote the boundary of A, i.e., $\partial A = \bar{A} \backslash A^\circ$ (where $\bar{A}$ denotes the closure of A and A° denotes the interior of A). Let $\boldsymbol{x} \in \mathbb{R}^D$, and define $\text{dist}(\boldsymbol{x}, A) = \inf_{\boldsymbol{y} \in A} \|\boldsymbol{x} - \boldsymbol{y}\|_2$. Let $\gamma > 0$, and define $B_\gamma(\boldsymbol{x}) = \{\boldsymbol{y} : \|\boldsymbol{x} - \boldsymbol{y}\|_2 < \gamma\}$. Let x_i denote the ith entry of $\boldsymbol{x}$ and for $\boldsymbol{y}_i \in \mathbb{R}^D$, let $y_{i,j}$ denote the jth entry of $\boldsymbol{y}_i$. Let $\boldsymbol{e}_i$ denote the ith standard basis vector. We use "whp" for "with high probability" and "wrt" for "with respect to."

Problem Parameters. Suppose that there are K arms associated with distributions $\nu_1, \ldots, \nu_K$ over $\mathbb{R}^D$ that have means $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K \in \mathbb{R}^D$, and let $\nu = (\nu_1, \ldots, \nu_K)$. At each round $t = 1, 2, \ldots$, an agent chooses an arm I_t and observes an independent draw $\boldsymbol{X}_t \sim \nu_{I_t}$.

Let $P \subset \mathbb{R}^D$ denote a nonempty set such that $P \neq \mathbb{R}^D$. Let $\boldsymbol{r}$ denote a reward vector, which is fixed, known, and the same across all arms. We assume $\|\boldsymbol{r}\|_2 = 1$. We say that $\boldsymbol{r}^\top \boldsymbol{\mu}_i$ is the expected *reward* of arm i. Let m denote the number of top feasible arms desired. We denote an instance of the top feasible arm identification problem by $(\nu, P, \boldsymbol{r}, m)$. Let $\Pr_\nu$ ($\mathbb{E}_\nu$) denote the probability measure (expected value) associated with the problem instance $(\nu, P, \boldsymbol{r}, m)$.

Define $\text{FEAS} = \{i \in [K] : \boldsymbol{\mu}_i \in P\}$, $\text{INFEAS} = \text{FEAS}^c$, and

$$\text{OPT} = \{i \in \text{FEAS} : \boldsymbol{r}^\top \boldsymbol{\mu}_i \geqslant \max^{(m)}_{j \in \text{FEAS}} \boldsymbol{r}^\top \boldsymbol{\mu}_j\},$$
$$\text{SUBOPT} = \{i \in [K] : \boldsymbol{r}^\top \boldsymbol{\mu}_i < \max^{(m)}_{j \in \text{FEAS}} \boldsymbol{r}^\top \boldsymbol{\mu}_j\}.$$

We say that an arm j is *suboptimal* if $\boldsymbol{r}^\top \boldsymbol{\mu}_j < \max^{(m)}_{i \in \text{FEAS}} \boldsymbol{r}^\top \boldsymbol{\mu}_i$; we say that an arm j is *feasible* (*infeasible*) if $\boldsymbol{\mu}_j \in P$ ($\boldsymbol{\mu}_j \notin P$). We note that, in general, SUBOPT and INFEAS are not disjoint, and that when

there are fewer than m arms that are feasible ($\boldsymbol{\mu}_i \in P$), SUBOPT $= \varnothing$.

We consider the following class of problems: $\mathcal{M} \coloneqq$

$\{(\nu, P, \boldsymbol{r}, m) : (\forall i : \boldsymbol{\mu}_i \notin \partial P)$ and
$(\max^{(m)}_{i \in \text{FEAS}} \boldsymbol{r}^\top \boldsymbol{\mu}_i > \max^{(m+1)}_{j \in \text{FEAS}} \boldsymbol{r}^\top \boldsymbol{\mu}_j \vee |\,\text{FEAS}\,| \leqslant m)\}$.

In words, $\mathcal{M}$ consists of problems where the means of the arms do not belong to the boundary of P and either there are m or fewer feasible arms or the mth largest reward of a feasible arm and the $(m+1)$th largest reward of a feasible arm are distinct. It is possible to drop the assumption $(\nu, P, \boldsymbol{r}, m) \in \mathcal{M}$ by allowing for a tolerance for suboptimality or infeasibility, and we describe this extension in the supplemental material.

Goal. We consider the fixed confidence setting with a novel criterion for correctness. An algorithm $\mathcal{A}$ is associated with a policy that determines which arm $I_t \in [K]$ is chosen at time t, a finite stopping time τ wrt $I_1, \boldsymbol{X}_1, I_2, \boldsymbol{X}_2, \ldots$ (i.e., $\Pr_\nu(\tau < \infty) = 1$) that determines when the algorithm stops, and an outputted partition of the arms $(\widehat{\text{O}}, \widehat{\text{S}}, \widehat{\text{I}})$ with $\widehat{\text{O}} \cup \widehat{\text{S}} \cup \widehat{\text{I}} = [K]$.

A standard criterion of correctness for an algorithm is δ- PAC, which we now define.

Definition 1. *Let $\delta \in (0,1)$. We say an algorithm $\mathcal{A}$ is δ- PAC wrt $\mathcal{M}$ if for any problem $(\nu, P, \boldsymbol{r}, m)$ belonging to $\mathcal{M}$, $\mathcal{A}$ outputs $\widehat{O} \subset [K]$ such that $\Pr_\nu(\widehat{O} = OPT) \geqslant 1 - \delta$.*

A standard goal is to design algorithms that are δ- PAC wrt $\mathcal{M}$ and that minimize τ. We propose a novel criterion δ- PAC-EXPLANATORY and aim to design algorithms that are δ- PAC-EXPLANATORY wrt $\mathcal{M}$ and that minimize τ.

Definition 2. *Let $\delta \in (0,1)$. We say an algorithm $\mathcal{A}$ is δ- PAC-EXPLANATORY wrt $\mathcal{M}$ if for any problem $(\nu, P, \boldsymbol{r}, m)$ belonging to $\mathcal{M}$, $\mathcal{A}$ outputs a triple $(\widehat{O}, \widehat{S}, \widehat{I})$ of disjoint sets such that $\widehat{O} \cup \widehat{S} \cup \widehat{I} = [K]$ and*

$$\Pr_\nu(\widehat{O} = OPT \text{ and } (\widehat{S}, \widehat{I}) \in \textit{Valid-Partitions}) \geqslant 1 - \delta$$

where Valid-Partitions $\coloneqq$

$$\{(S, I) : S \subset SUBOPT, I \subset INFEAS, \\ S \cap I = \varnothing, S \cup I = OPT^c\}.$$

To identify arms in OPT, an agent must rule out every $i \in \text{OPT}^c$ as suboptimal or infeasible. When SUBOPT $\cap$ INFEAS $\neq \varnothing$, there are arms that can be ruled out in multiple ways. Valid-Partitions captures the various *correct* ways to partition the arms in OPT^c to distinguish them from OPT. Thus, our notion, δ- PAC-EXPLANATORY, is slightly stronger than δ- PAC since it essentially requires that whp *(i)* an algorithm output the correct top m feasible arms and *(ii)* that it provide a correct reason for rejecting each arm (either that it is suboptimal or infeasible). We remark that it is natural to require only one reason for rejecting an arm because once an algorithm identifies an arm as infeasible (suboptimal), there is no reason to keep pulling it to determine whether it is suboptimal (infeasible). Furthermore, in most problems and for most algorithms, if an arm is infeasible and suboptimal, showing one of these is easier than showing the other.

This notion is practically relevant since in many applications it is of interest to provide a reason to reject an arm. For example, in crowdsourcing, it might be necessary to provide a worker with a reason for why she was not hired. In clinical trials, it might be useful for the clinician to know why a drug is rejected. Furthermore, as we discuss in the supplemental material, we conjecture that there is a small gap between δ- PAC and δ- PAC-EXPLANATORY algorithms.

Sub-Gaussian Assumption. We assume that each ν_i is a multi-dimensional sub-Gaussian distribution, which we now define. Let X be a scalar random variable and $\boldsymbol{X} \in \mathbb{R}^D$ a random vector. We say that X is *sub-Gaussian* if $\mathbb{E}\exp(\frac{X^2}{\sigma^2}) \leqslant 2$ for some $\sigma > 0$ and $\boldsymbol{X} \in \mathbb{R}^D$ is sub-Gaussian if for all $\boldsymbol{a} \in \mathbb{R}^D$, $\boldsymbol{X}^\top \boldsymbol{a}$ is sub-Gaussian. The sub-Gaussian norms of X and $\boldsymbol{X}$ are defined respectively as:

$$\|X\|_{\psi_2} = \inf\{\sigma > 0 : \mathbb{E}\exp(\frac{X^2}{\sigma^2}) \leqslant 2\}, \\ \|\boldsymbol{X}\|_{\psi_2} = \sup_{\boldsymbol{a} \in \mathbb{R}^D : \|\boldsymbol{a}\|_2 = 1} \left\|\boldsymbol{X}^\top \boldsymbol{a}\right\|_{\psi_2}.$$

X is said to be *σ-sub-Gaussian* if $\|X\|_{\psi_2} \leqslant \sigma$ and $\boldsymbol{X}$ is said to be *σ-sub-Gaussian* if $\|\boldsymbol{X}\|_{\psi_2} \leqslant \sigma$. For the remainder of this paper, we assume that $\nu_1, \ldots, \nu_K$ are σ-sub-Gaussian. See Vershynin (2012); Vershynin et al. (2017) for more details.

4 Lower Bounds

Theorem 1 gives our lower bound for δ- PAC-EXPLANATORY algorithms.

Theorem 1. *Let $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K \in \mathbb{R}^D$ such that $\forall i \neq j \in [K] : \mu_{i,1} \neq \mu_{j,1}$. Define $\nu_i = N(\boldsymbol{\mu}_i, I_D)$ for all $i \in [K]$. Suppose $P = \mathbb{R} \times P'$ for some $P' \subset \mathbb{R}^{D-1}$ and $\forall \boldsymbol{x} \in \partial P, \forall \epsilon > 0 : B_\epsilon(\boldsymbol{x}) \cap P^\circ \neq \varnothing$ and $B_\epsilon(\boldsymbol{x}) \cap (P^c)^\circ \neq \varnothing$. Let $\boldsymbol{r} = \boldsymbol{e}_1$. Assume $(\nu, P, \boldsymbol{e}_1, m) \in \mathcal{M}$ and let $\delta \in (0, 0.1)$. For any $(S, I) \in$ Valid-Partitions, define $L(S, I) \coloneqq$*

$$\sum_{i \in OPT} \max([\min_{j \in S} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)]^{-2}, \ \text{dist}(\boldsymbol{\mu}_i, \partial P)^{-2}) \\ + \sum_{i \in S}[\min_{j \in OPT} \boldsymbol{r}^\top(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)]^{-2} + \sum_{i \in I} \text{dist}(\boldsymbol{\mu}_i, P)^{-2}.$$

Then, any algorithm $\mathcal{A}$ that is δ- PAC-EXPLANATORY wrt $\mathcal{M}$ has a stopping time τ on the problem $(\nu, P, \boldsymbol{e}_1, m)$ that satisfies

$$\mathbb{E}_\nu[\tau] \geqslant \min_{(S,I) \in \textit{Valid-Partitions}} \frac{2}{15} \ln(\frac{1}{2\delta}) L(S, I).$$

The conditions $P = \mathbb{R} \times P'$ and $\boldsymbol{r} = \boldsymbol{e}_1$ decouple the reward and feasibility of the arms and hold in many applications. The other conditions on P remove pathological cases such as isolated points and nowhere dense sets with positive measure.

The lower bound is the solution of a constrained minimization problem over all the ways to distinguish the arms in OPT^c from OPT, i.e., $(S, I) \in$ Valid-Partitions. If we fix some $(S, I) \in$ Valid-Partitions, there are three main terms in the lower bound reflecting the difficulty of identifying arms as belonging to either OPT, S, or I, respectively. Essentially, optimal arms must be shown to be feasible and to have reward greater than all arms in S, arms in S must be shown to have reward less than arms in OPT, and arms in I must be shown to be infeasible.

The key observation in the proof is that for a given problem $(\nu, P, \boldsymbol{r}, m)$, we can associate with an algorithm $\mathcal{A}$ a particular $(S, I) \in$ Valid-Partitions such that for every $i \in S$ ($i \in I$), it is likely that $\mathcal{A}$ puts $i \in \widehat{\text{S}}$ ($i \in \widehat{\text{I}}$). Then, using the notion of δ- PAC-EXPLANATORY, it suffices to analyze the difficulty of identifying each arm as belonging either to OPT, S, or I. The result follows by minimizing over $(S, I) \in$ Valid-Partitions.

We also state a similar lower bound for algorithms that are δ- PAC wrt $\mathcal{M}$.

Theorem 2. *Assume the conditions of Theorem 1. Define* $r_o = \min_{j \in OPT} \boldsymbol{r}^\top \boldsymbol{\mu}_j$, $r_s \coloneqq \max_{j \in OPT^c \cap FEAS} \boldsymbol{r}^\top \boldsymbol{\mu}_j$, *and* $L' \coloneqq$

$$\sum_{i \in INFEAS \cap SUBOPT} \min([r_o - \boldsymbol{r}^\top \boldsymbol{\mu}_i]^{-2},\ \text{dist}(\boldsymbol{\mu}_i, P)^{-2})$$
$$+ \sum_{i \in OPT} \max([\boldsymbol{r}^\top \boldsymbol{\mu}_i - r_s]^{-2},\ \text{dist}(\boldsymbol{\mu}_i, \partial P)^{-2})$$
$$+ \sum_{i \in OPT^c \cap FEAS} [\min_{j \in OPT} \boldsymbol{r}^\top (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)]^{-2}$$
$$+ \sum_{i \in INFEAS \cap SUBOPT^c} \text{dist}(\boldsymbol{\mu}_i, P)^{-2}.$$

Then, any algorithm $\mathcal{A}$ that is δ- PAC wrt $\mathcal{M}$ has a stopping time τ on the problem $(\nu, P, \boldsymbol{e}_1, m)$ that satisfies

$$\mathbb{E}_\nu[\tau] \geqslant \ln(\frac{1}{2.4\delta}) L'.$$

The bound in Theorem 2 suggests that δ- PAC algorithms must show that arms in OPT are feasible and have reward greater than every arm in $\text{OPT}^c \cap \text{FEAS}$, arms in $\text{OPT}^c \cap \text{FEAS}$ have reward less than arms in OPT, arms in $\text{INFEAS} \cap \text{SUBOPT}^c$ are infeasible, and, finally, arms in $\text{INFEAS} \cap \text{SUBOPT}$ are either infeasible or suboptimal.

Since any δ- PAC-EXPLANATORY algorithm wrt $\mathcal{M}$ is δ- PAC wrt $\mathcal{M}$, we expect the lower bound in Theorem 1 to be at least as large as the lower bound in Theorem 2, and this is indeed the case. The main difference between the bounds occurs in the terms corresponding to $i \in \text{OPT}$. Essentially, in Theorem 1, it is required to show that every arm in OPT has reward greater than all arms that are ruled out as suboptimal (i.e., belong to S), whereas in Theorem 2, these arms must only be shown to have reward greater than arms in $\text{FEAS} \cap \text{OPT}^c$. See the supplemental material for a more detailed discussion.

5 TF-LUCB: A Family of Algorithms for Top Feasible Arm Identification

In this section, we introduce an algorithm for the top feasible arm identification problem. To begin, we define some notation. Let $\widehat{\boldsymbol{\mu}}_{i,s}$ denote the empirical mean of arm i after s samples. Let $N_i(t) = \sum_{s=1}^{t-1} \mathbf{1}\{I_s = i\}$ denote the number of times that arm i has been selected up to round t. Let

$$U(t, \delta) = \sigma \sqrt{\frac{2\log(1/\delta) + 6\log\log(1/\delta) + 3\log\log(et)}{t}}$$

denote a confidence bound, which holds uniformly over time (see Lemma F.10 in the supplemental material) (Kaufmann et al., 2016). For the sake of simplicity, we assume henceforth that $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K \in B_{\frac{1}{2}}(\mathbf{0})$ and $P \subset B_{\frac{1}{2}}(\mathbf{0})$.

Challenge. As suggested by Theorem 2, a major challenge in designing a nearly optimal algorithm is how to rule out with nearly *optimal* sample complexity an arm i that is infeasible and whose reward $\boldsymbol{r}^\top \boldsymbol{\mu}_i$ is too small to be among the top m feasible arms (i.e., belongs to $\text{INFEAS} \cap \text{SUBOPT}$). In short, a nearly optimal algorithm must determine which is easier to show: that arm i is infeasible or that it has too small reward. Either of these can be arbitrarily more difficult to show than the other; for example, consider an infeasible arm with mean very close to the set P and a very small reward relative to the other arms. In this case, it is quite easy to show suboptimality, but very difficult to show infeasibility.

Algorithm. TF-LUCB is a family of algorithms, where each instance is specified by a subroutine TestF. TestF(i, s) considers the first s pulls of arm i and returns True if arm i is feasible whp, returns False if i is in-

feasible whp, and otherwise returns ?, indicating "don't know." When the context makes it clear which distribution is involved, we simply write TestF(s). TestF essentially solves what we will call the *set membership problem*, which we now define. In this problem, there is a distribution ξ over $\mathbb{R}^D$ with mean $\boldsymbol{\mu} \in \mathbb{R}^D$ and a set $P \subset \mathbb{R}^D$. At round $t = 1, 2, \ldots$ an algorithm $\mathcal{B}$ observes $\boldsymbol{X}_t \sim \xi$. An algorithm $\mathcal{B}$ is associated with a stopping time τ wrt $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$, and after τ rounds outputs True if it concludes that $\boldsymbol{\mu} \in P$ and False if it concludes that $\boldsymbol{\mu} \notin P$. We define the following class of set membership problems:

$$\mathcal{N} = \{(\xi, P) : \xi \text{ is a distribution over } \mathbb{R}^D \text{ with mean } \boldsymbol{\mu} \in B_{\frac{1}{2}}(\mathbf{0}), P \subset B_{\frac{1}{2}}(\mathbf{0}), P \neq \varnothing, \boldsymbol{\mu} \notin \partial P\}.$$

We defer our discussion of specific algorithms for the set membership problem until the next section.

Given a subroutine TestF, TF-LUCB is an adaptation of LUCB (Lower Upper Confidence Bound) from Kalyanakrishnan et al. (2012) to the top feasible arm identification problem. TF-LUCB maintains three sets: arms F_t that are feasible whp, arms G_t that have not been determined whp to be feasible or infeasible, and arms $E_t \coloneqq F_t \cup G_t$ that have not been ruled out as infeasible whp. At round t, TF-LUCB considers TOP_t, the set of m arms that have not been ruled out as infeasible whp (i.e., belong to E_t) and have the top m estimated rewards. TF-LUCB uses $U_{\boldsymbol{r}}(t, \delta) \coloneqq U(t, \frac{\delta}{2K})$ for a confidence bound on the reward associated with an arm. If all of the arms in TOP_t are feasible whp, then it pulls an arm h_t in TOP_t with the smallest lower confidence bound. If only some of the arms in TOP_t are determined to be feasible whp, then to avoid oversampling optimal arms, it chooses the arm h_t instead by picking the arm in $\text{TOP}_t \cap G_t$ with the smallest lower confidence bound, i.e., an arm in the top empirical m for which it is still not determined whp whether it is feasible. We note that because $\text{TOP}_t \cap E_t^c = \varnothing$ by definition of TOP_t, when $\text{TOP}_t \nsubseteq F_t$, $\text{TOP}_t \cap G_t$ is nonempty so that the argmax operator in line 14 is well-defined. If there are arms outside of TOP_t that have not been ruled out as infeasible, then the algorithm pulls an additional arm l_t among these (in $\text{TOP}_t^c \cap E_t$) that maximizes an upper confidence bound on its reward. The algorithm terminates when it determines whp that each arm in TOP_t is feasible and has mean larger than arms in $\text{TOP}_t^c \cap E_t$, or that the arms in TOP_t are feasible and all other arms are infeasible.

For the sake of brevity, define the function $F(x, y) = x^{-2}\log(\log(x^{-2})y)$. Theorem 3 shows that TF-LUCB is δ- PAC-EXPLANATORY with a bound on τ that nearly matches the lower bound.

Theorem 3. *Let $\delta \in (0, 1)$ and $(\nu, P, \boldsymbol{r}, m) \in \mathcal{M}$. Suppose that for any set membership problem $(\xi, R) \in \mathcal{N}$ where ξ is σ-sub-Gaussian and has mean $\boldsymbol{\mu}$, with probability at least $1 - \frac{\delta}{2K}$, TestF returns True only if $\boldsymbol{\mu} \in R$ and False only if $\boldsymbol{\mu} \in R^c$, and TestF uses at most $\eta(\xi, R)$ samples, where $\eta(\xi, R)$ is a deterministic function of ξ and R. For any $(S, I) \in$ Valid-Partitions, define $\mathcal{U}(S, I) \coloneqq$*

$$\sum_{i\in S} F(\min_{j\in OPT} \boldsymbol{r}^\top(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i), \frac{K}{\delta}) + \sum_{i\in I} \eta(\nu_i, P) + \sum_{i\in OPT} \max(F(\min_{j\in S} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j), \frac{K}{\delta}), \eta(\nu_i, P)).$$

Then, with probability at least $1 - \delta$, TF-LUCB returns $(\widehat{O}, \widehat{S}, \widehat{I})$ such that $\widehat{O} = OPT$, $(\widehat{S}, \widehat{I}) \in$ Valid-Partitions, and

$$\tau \leqslant \min_{(S,I)\in \textit{Valid-Partitions}} c\sigma^2 \mathcal{U}(S, I). \tag{1}$$

where c is a universal positive constant.

This upper bound has a very similar structure to the lower bound in Theorem 1. It is the solution of a constrained minimization problem over $(S, I) \in$ Valid-Partitions. One can interpret this form as saying that TF-LUCB finds the easiest way to solve a given instance of the top feasible arm identification problem. Ignoring doubly logarithmic factors, the upper bound on the reward-associated terms is loose by a factor of $\log(K)$.[1] Theorem 3 can be interpreted as a reduction of the top feasible arm identification problem to the set membership problem and in the next section we will discuss how various algorithms for the set membership problem affect the sample complexity of TF-LUCB.

In light of Theorem 3, it is instructive to consider a two-stage algorithm that first identifies the collection of feasible arms and then applies a best arm identification algorithm to the feasible arms. The drawback of this two-stage approach is that there may be suboptimal infeasible arms that are much easier to rule out as suboptimal rather than infeasible. Essentially, such a two-stage algorithm solves a problem instance by picking the $(S', I') \in$ Valid-Partitions such that $I' =$ INFEAS, whereas TF-LUCB adapts to the problem instance to choose the best $(S, I) \in$ Valid-Partitions. Thus, the sample complexity of such a two-stage algorithm is at least the sample complexity of TF-LUCB and can be arbitrarily larger than the sample complexity of TF-LUCB. To see this, consider a problem with an arm whose mean is very close to the boundary of P, but has very small reward relative to the other arms.

The proof of Theorem 3 considers the $(S, I) \in$ Valid-Partitions that minimizes (1) and analyzes the

[1] We note that this logarithmic factor could be improved by adapting LUCB++ (Simchowitz et al., 2017) instead of LUCB.

Algorithm 1 TF-LUCB: Top-m Feasible Lower Upper Confidence Bound algorithm

Input: TestF, sub-Gaussian norm bound σ, confidence δ
for $t = 1, 2, \ldots$ **do**
  $F_t \longleftarrow \{i \in [K] : \text{TestF}(i, N_i(t)) = \text{ True}\}$ // arms that are feasible whp
  $G_t \longleftarrow \{i \in [K] : \text{TestF}(i, N_i(t)) = ?\}$ // arms whose feasibility is unclear whp
  $E_t \longleftarrow F_t \cup G_t$ // arms that are not ruled out as infeasible whp
  $\text{TOP}_t \longleftarrow \arg\max_{Z \subset E_t, |Z| = \min(m, |E_t|)} \sum_{i \in Z} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i, N_i(t)}$
  if $\text{TOP}_t = F_t$ and $F_t = E_t$
    return $(\text{TOP}_t, \text{TOP}_t^c \cap E_t, E_t^c)$
  if $\text{TOP}_t \subset F_t$ and $\min_{i \in \text{TOP}_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i, N_i(t)} - U_{\boldsymbol{r}}(N_i(t), \delta) \geqslant \max_{j \in \text{TOP}_t^c \cap E_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{j, N_j(t)} + U_{\boldsymbol{r}}(N_j(t), \delta)$
    return $(\text{TOP}_t, \text{TOP}_t^c \cap E_t, E_t^c)$
  if $\text{TOP}_t \subset F_t$
    $h_t = \arg\min_{i \in \text{TOP}_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i, N_i(t)} - U_{\boldsymbol{r}}(N_i(t), \delta)$
  if $\text{TOP}_t \not\subset F_t$
    $h_t = \arg\min_{i \in \text{TOP}_t \cap G_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i, N_i(t)} - U_{\boldsymbol{r}}(N_i(t), \delta)$
  if $\text{TOP}_t^c \cap E_t \neq \varnothing$
    $l_t = \arg\max_{j \in \text{TOP}_t^c \cap E_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{j, N_j(t)} + U_{\boldsymbol{r}}(N_j(t), \delta)$
    Pull arm l_t
  Pull arm h_t

sample complexity of TF-LUCB to identify each arm as belonging to either OPT, S, or I. It is shown that at each round t, either h_t or l_t is a needy arm wrt to the sets OPT, S, and I (defined precisely in the supplemental material) in the sense that either it is necessary to determine whether it is feasible or it is necessary to improve our estimate of its reward.

Allowing for a tolerance: It is possible to extend TF-LUCB to allow for a tolerance on suboptimality or infeasibility (see supplemental material). For example, if a *suboptimality gap* of $\epsilon > 0$ is permitted, then $\widehat{\text{O}}$ is correct if it satisfies $\forall i \in \widehat{\text{O}}$, $i \in \text{FEAS}$ and $\boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_i + \epsilon \geqslant \min_{j \in \text{OPT}} \boldsymbol{r}^\top \boldsymbol{\mu}_j$.

6 Three Instances of TF-LUCB

In this section, we consider three distinct general classes of sets and apply Theorem 3 to derive upper bounds for algorithms for each of these. To begin, we consider a general set P. Since there are in general no known computationally efficient algorithms for such a general setting, we then consider the computationally tractable and very rich class of polyhedra. For this setting, let $P = \{\boldsymbol{x} \in \mathbb{R}^D : A\boldsymbol{x} \leqslant \boldsymbol{b}\}$ denote a polyhedron, where $A \in \mathbb{R}^{M \times D}$ and $\boldsymbol{b} \in \mathbb{R}^M$. Let $\boldsymbol{a}_j^\top$ denote the jth row of A. By dividing each constraint j by $\|\boldsymbol{a}_j\|_2$, we can assume without loss of generality that $\|\boldsymbol{a}_j\|_2 = 1$ for all $j \in [M]$. Finally, we consider the common case where the polyhedron has orthogonal constraints, i.e., $\boldsymbol{a}_i^\top \boldsymbol{a}_j = 0$ for all $i \neq j \in [M]$, which arises for example when there is one constraint per coordinate. Note that in this case, it follows that $M \leqslant D$.

For a general set, we propose the TestF subroutine: TestF-B (see Algorithm 2). It controls $\|\widehat{\boldsymbol{\mu}}_{i,t} - \boldsymbol{\mu}_i\|_2$ with a confidence bound $U_{\text{ball}}(t, \delta) \coloneqq 2U(t, \frac{\delta}{5^D 2K})$ that is constructed based on an ϵ-net argument. TestF-B returns True if the ball centered at $\widehat{\boldsymbol{\mu}}_{i,t}$ with radius $U_{\text{ball}}(t, \delta)$ does not intersect P^c, False if this ball does not intersect P, and otherwise returns ?. The variant of TF-LUCB that uses TestF-B is called *TF-LUCB-B*.

For a polyhedron, we propose the subroutine TestF-CB, which also uses the confidence bound $U_{\text{con}}(t, \frac{\delta}{2}) \coloneqq U(t, \frac{\delta}{4KM})$ (see Algorithm 3). If it determines that $\boldsymbol{\mu}_i$ satisfies all of the constraints whp, it returns True, if it determines that the ball centered at $\widehat{\boldsymbol{\mu}}_{i,t}$ with radius $U_{\text{ball}}(t, \frac{\delta}{2})$ does not intersect P whp, it returns False, and otherwise it returns ?. The variant of TF-LUCB that uses TestF-CB is called *TF-LUCB-CB*.

Finally, a polyhedron with orthogonal constraints, we propose the subroutine TestF-C, which uses the confidence bound $U_{\text{con}}(t, \delta)$ (see Algorithm 4). If it determines that $\boldsymbol{\mu}_i$ satisfies all of the constraints whp, it returns True, if it determines that $\boldsymbol{\mu}_i$ violates one of the constraints whp, it returns False, and otherwise it returns ?.

The following theorem establishes upper bounds for TF-LUCB-B, TF-LUCB-CB, and TF-LUCB-C.

Theorem 4. *Let $\delta \in (0, 1)$ and $(\nu, P, \boldsymbol{r}, m) \in \mathcal{M}$. Then, with probability at least $1 - \delta$,*

- *TF-LUCB-B returns $(\widehat{O}, \widehat{S}, \widehat{I})$ such that $\widehat{O} = OPT$, $(\widehat{S}, \widehat{I}) \in$ Valid-Partitions, τ is bounded as in (1), and $\eta(\nu_i, P)$ is bounded as in Table 1.*
- *If P is a polyhedron, TF-LUCB-CB returns $(\widehat{O}, \widehat{S}, \widehat{I})$ such that $\widehat{O} = OPT$, $(\widehat{S}, \widehat{I}) \in$*

Algorithm 2 TestF-B:

Input: arm index i, number of pulls t
if $\text{dist}(\widehat{\boldsymbol{\mu}}_{i,t}, P^c) > U_{\text{ball}}(t,\delta)$
  return True
if $\text{dist}(\widehat{\boldsymbol{\mu}}_{i,t}, P) > U_{\text{ball}}(t,\delta)$
  return False
else
  return ?

Algorithm 3 TestF-CB:

Input: arm index i, number of pulls t
if $A\widehat{\boldsymbol{\mu}}_{i,t} + U_{\text{con}}(t, \frac{\delta}{2})\mathbf{1} \leqslant \boldsymbol{b}$
  return True
if $\text{dist}(\widehat{\boldsymbol{\mu}}_{i,t}, P) > U_{\text{ball}}(t, \frac{\delta}{2})$
  return False
else
  return ?

Algorithm 4 TestF-C:

Input: arm index i, number of pulls t
if $A\widehat{\boldsymbol{\mu}}_{i,t} + U_{\text{con}}(t,\delta)\mathbf{1} \leqslant \boldsymbol{b}$
  return True
if $A\widehat{\boldsymbol{\mu}}_{i,t} - U_{\text{con}}(t,\delta)\mathbf{1} \not\leqslant \boldsymbol{b}$
  return False
else
  return ?

	TF-LUCB-B	TF-LUCB-CB	TF-LUCB-C
$i \in$ OPT	$DF(\text{ dist}(\boldsymbol{\mu}_i, \partial P), \frac{K}{\delta})$	$F(\text{ dist}(\boldsymbol{\mu}_i, \partial P), \frac{KM}{\delta}))$	$F(\text{ dist}(\boldsymbol{\mu}_i, \partial P), \frac{KM}{\delta})$
$i \in$ INFEAS	$DF(\text{ dist}(\boldsymbol{\mu}_i, \partial P), \frac{K}{\delta})$	$DF(\text{ dist}(\boldsymbol{\mu}_i, \partial P), \frac{K}{\delta})$	$v_i F(\text{ dist}(\boldsymbol{\mu}_i, P), \frac{KM}{\delta})$

Table 1: Upper bounds on $\eta(\nu_i, P)$. For the case where P is a polyhedron, let $v_i = |\{j : \boldsymbol{a}_j^\top \boldsymbol{\mu}_i > b_j\}|$.

Valid-Partitions, τ is bounded as in (1)*, and $\eta(\nu_i, P)$ is bounded as in Table 1.*

- *If P is a polyhedron with orthogonal constraints, TF-LUCB-C returns $(\widehat{O}, \widehat{S}, \widehat{I})$ such that $\widehat{O} = OPT$, $(\widehat{S}, \widehat{I}) \in$ Valid-Partitions, τ is bounded as in* (1)*, and $\eta(\nu_i, P)$ is bounded as in Table 1.*

Ignoring doubly logarithmic factors, the terms related to determining feasibility for TF-LUCB-B are loose by a factor of $D\log(K)$ relative to our lower bound. When D is $O(\log K)$, then the bound is loose by a polylogarithmic factor. Since in many applications the dimension of the feedback is not very large, this bound is practically relevant. TF-LUCB-CB only requires $F(\text{ dist}(\boldsymbol{\mu}_i, \partial P), \frac{KM}{\delta})$ samples to show that an arm $i \in$ OPT is feasible, which is a significant improvement over the corresponding term for TF-LUCB-B if M is polynomial in D. TF-LUCB-C differs from TF-LUCB-CB in the term for showing infeasibility. The term for determining that arms in I are infeasible is loose by a factor $v_i \log(KM)$, which can be much smaller than $D\log(K)$. In the common setting where the arms are two-dimensional with one coordinate encoding reward and the other a constraint, the upper bound is only loose by a logarithmic factor. See the supplemental material for an upper bound for TF-LUCB-C for the case of a general polyhedron.

7 Experiments

In this section, we demonstrate experimentally the effectiveness of our algorithms. We consider the task of identifying OPT $\subset [K]$.

Synthetic Datasets: In each of the experiments, we use $\delta = 0.1$, the last coordinate determines the reward ($\boldsymbol{r} = (0, \ldots, 0, 1)^\top$), and the rest of the coordinates determine whether $\boldsymbol{x} \in P$. We consider two kinds of reward structures: linearly varying rewards $\boldsymbol{r}^\top \boldsymbol{\mu}_i = .95(1 - \frac{i}{100})$ and polynomially varying rewards $\boldsymbol{r}^\top \boldsymbol{\mu}_i = .95(1 - (\frac{i}{100})^{.3})$. In each trial, we randomly permute the rewards among the arms in the sense that we take a random permutation $\sigma : [K] \longrightarrow [K]$ and set $\mu_{i,D}$ to $\mu_{\sigma(i),D}$.

In one set of experiments, we use 6-dimensional multivariate Gaussian distributions as arms with covariance matrix $\frac{1}{4}I$. We use a simplex $P = \{\boldsymbol{x} \in \mathbb{R}^6 : \sum_{i=1}^5 x_i \leqslant 2, x_i \geqslant 0 \forall i \in [5]\}$. We consider one setting where there are four groups of arms $\boldsymbol{\mu}_{1:15,1:5} = (.1)^{\otimes 5}, \boldsymbol{\mu}_{16:30,1:5} = (.35)^{\otimes 5}, \boldsymbol{\mu}_{31:45,1:5} = (.45)^{\otimes 5}, \boldsymbol{\mu}_{46:60,1:5} = (-.1)^{\otimes 5}$. Only the arms in [30] are feasible. In another setting, we consider arms with arithmetically changing values. In this setting, for $i \in [30]$, $\boldsymbol{\mu}_{i,1:5} = [(.1 + (\frac{2-0.05}{5} - .1)\frac{i}{30}]^{\otimes 5}$, for $i \in [45]\backslash[30]$, $\boldsymbol{\mu}_{i,1:5} = [2.05/5 + (3/5 - 2.05/5)\frac{i-30}{15}]^{\otimes 5}$, and for $i \in [60]\backslash[45]$, $\boldsymbol{\mu}_{i,1:5} = [-0.05 + (-.3 + 0.05)\frac{i-45}{15}]^{\otimes 5}$. Only the arms in [30] are feasible. We use $\sqrt{1/4}$ as the sub-Gaussian norm for the arms.

In another set of experiments, we use 5-dimensional Bernoulli distributions. We use an ordered polyhedron $P = \{\boldsymbol{x} \in \mathbb{R}^5 : \boldsymbol{x}_i \leqslant \boldsymbol{x}_{i+1} \forall i \in [3]\}$. We consider a setting with three groups: $\boldsymbol{\mu}_{1:30,1:4} = (0.05, 0.35, 0.65, 0.95)^\top$, $\boldsymbol{\mu}_{31:40,1:4} = (0.95, 0.65, 0.35, 0.05)^\top$, and $\boldsymbol{\mu}_{41:50,1:4} = (.7, .6, .5, .4)^\top$. Only the arms in [30] are feasible. We use 1 as the sub-Gaussian norm of the arms.

Crowdsourcing Application: We consider the task of finding the most accurate crowdsourcing workers subject to the constraint that they complete tasks at a suitable average speed. We use a crowdsourcing dataset collected by Venanzi et al. (2016) in which Amazon Mechanical Turk workers determine what kind of a

Experiment	TF-LUCB-C	TF-LUCB-CB	TF-AE-C	FFAF-C	FFAF-CB
Simplex Arithmetic Linear	1.00	1.45	2.84	1.56	3.05
Simplex Arithmetic Polynomial	1.00	1.48	3.12	1.59	3.23
Simplex Groups Linear	1.00	1.25	2.78	1.29	2.14
Simplex Groups Polynomial	1.00	1.32	2.97	1.32	2.14
Ordered Groups Linear	1.00	1.04	1.93	1.15	1.16
Ordered Groups Polynomial	1.00	1.05	2.02	1.43	1.33
Crowdsourcing	1.00	N/A	2.15	2.88	N/A
Medical	1.00	N/A	1.12	4.52	N/A

Table 2: Number of samples required, relative to TF-LUCB-C, averaged over 50 trials.

statement a tweet makes regarding the weather: *(i)* positive, *(ii)* neutral, *(iii)* negative, *(iv)* unrelated, or *(v)* can't tell. We only consider workers that have answered at least 100 questions, leaving a total of 21 workers. Here, $\mu_{i,1}$ is the probability of being correct and $\mu_{i,2}$ is the average amount of time required. We seek the top 3 most accurate workers who on average answer questions within 15 seconds. Whenever an algorithm pulls an arm corresponding to a worker, it samples a datapoint associated with that worker uniformly at random with replacement. We use the standard deviation of the speed measurements (135.86 sec) as the sub-Gaussian norm for the coordinate corresponding to the speed and 1 as the sub-Gaussian norm for the other coordinate. We use $\delta = 0.1$ and allow for a suboptimality gap of 0.05.

Clinical Trials Application: We examine the problem in clinical trials of finding the most effective drugs that also meet some safety threshold. We use data from Genovese et al. (2013) (see ARCR20 in week 16 in Table 2 and Table 3), which studies the drug secukinumab for treating rheumatoid arthritis. Each arm corresponds to a dosage level (25mg, 75mg, 150mg, 300mg, placebo) and has two attributes: the probability of being effective, $\mu_{i,1}$, and the probability of causing an infection or infestation, $\mu_{i,2}$. The dosage levels 25mg, 75mg, 150mg, and 300mg have averages $\boldsymbol{\mu}_1 = (.34, .259)^\top$, $\boldsymbol{\mu}_2 = (.469, .184)^\top$, $\boldsymbol{\mu}_3 = (.465, .209)^\top$, $\boldsymbol{\mu}_4 = (.537, .293)^\top$, respectively, and the placebo has average $\boldsymbol{\mu}_5 = (.36, .36)^\top$. In our experiment, whenever arm i is chosen two Bernoulli random variables with means given by $\boldsymbol{\mu}_i$ are drawn. We assume that a drug is acceptable if the probability of an infection is below .25, we set $m = 1$, and we allow for a suboptimality gap of 0.05. Thus, the correct answer is either arm 2 or arm 3. We use $\delta = 0.05$. We use 1 as the sub-Gaussian norm.

Algorithms: We consider our algorithms TF-LUCB-C and TF-LUCB-CB. We also consider Find-Feasible-Arms-First (FFAF), which is a two-stage algorithm that first determines which of the arms are feasible and then applies LUCB to the feasible arms to find the top arms. FFAF-CB uses TestF-CB to test feasibility, whereas FFAF-C uses TestF-C. We also implement an action elimination algorithm (TF-AE-C) that samples remaining arms in a round-robin fashion, eliminating an arm if it is determined using confidence bounds to be either suboptimal or infeasible. We only consider a variant that uses TestF-C since TF-AE-C has poor performance. For the experiments where $D = 2$, we only run the constraint based algorithms since the ϵ-net approach uses strictly worse (by a constant factor) confidence bounds.

Discussion of Results: Table 2 displays our results as the number of samples required, relative to TF-LUCB-C. All algorithms find a correct set of arms on every trial. TF-LUCB-C has the best sample complexity in all of the experiments, beating the FFAF algorithms by a substantial margin in many of them. In particular, FFAF-C requires nearly five times as many samples as TF-LUCB-C on the medical dataset and nearly three times as many samples on the crowdsourcing dataset. The performance gap between TF-LUCB and FFAF depends on the relative difficulty of showing arms to be suboptimal vs. infeasible. In particular, FFAF-C has poor performance on the real-world datasets because on the crowdsourcing dataset the sub-Gaussian norm for showing feasibility is large and on the medical dataset one of the suboptimal infeasible arms is very close to the boundary. TF-AE-C performs so poorly because each suboptimal feasible arm must be pulled until at least m arms are shown to be feasible and have larger reward than it.

8 Conclusion

We introduced a novel problem, top feasible arm identification: the first general pure exploration multi-armed bandit problem on constrained optimization. We argued that it has many real-world applications since in many settings there is multi-dimensional feedback and a natural goal is constrained optimization based on this feedback (e.g., safety and effectiveness in clinical trials); thus, we argue that our algorithms are of significant practical interest.

Acknowledgements

We thank the anonymous reviewers for their very helpful comments. This work was supported in part by NSF grants 1422157 and 1838179.

References

J.-Y. Audibert and S. Bubeck. Best arm identification in multi-armed bandits. *Conference on Learning Theory*, 2010.

P. Auer, C.-K. Chiang, R. Ortner, and M. Drugan. Pareto front identification from stochastic bandit feedback. *Artificial Intelligence and Statistics*, pages 939–947, 2016.

S. Bubeck, T. Wang, and N. Viswanathan. Multiple identifications in multi-armed bandits. *International Conference on Machine Learning*, pages 258–265, 2013.

L. Chen, A. Gupta, J. Li, M. Qiao, and R. Wang. Nearly optimal sampling algorithms for combinatorial pure exploration. *Proceedings of Machine Learning Research*, 65:1–53, 2017.

S. Chen, T. Lin, I. King, M. Lyu, and W. Chen. Combinatorial pure exploration of multi-armed bandits. *Advances in Neural Information Processing Systems*, pages 379–387, 2014.

V. Gabillon, M. Ghavamzadeh, and A. Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. *Advances in Neural Information Processing Systems*, pages 3212–3220, 2012.

M. Genovese, P. Durez, H. Richards, J. Supronik, E. Dokoupilova, V. Mazurov, and J. Aelion. Efficacy and safety of secukinumab in patients with rheumatoid arthritis: a phase ii, dose-finding, double-blind, randomised, placebo controlled study. *Annals of the rheumatic diseases*, 72:863–869, 2013.

K. Jamieson, M. Malloy, R. Nowak, and S. Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. *Conference on Learning Theory*, pages 424–439, 2014.

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

J. Katz-Samuels and C. Scott. Feasible arm identification. In J. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2535–2543, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR. URL `http://proceedings.mlr.press/v80/katz-samuels18a.html`.

E. Kaufmann, O. Cappé, and A. Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17:1–42, 2016.

S. Mannor and J. Tistisklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, pages 623–648, 2004.

M. Simchowitz, K. Jamieson, and B. Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. *Conference on Learning Theory*, pages 1794–1834, 2017.

M. Soare, A. Lazaric, and R. Munos. Best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, pages 828–836, 2014.

M. Venanzi, J. Guiver, P. Kohli, and N. Jennings. Time-sensitive bayesian information aggregation for crowdsourcing systems. *Journal of Artificial Intelligence Research*, pages 517–545, 2016.

R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y. Eldar and G. Kutyniok, editors, *Compressed Sensing: Theory and Applications*, pages 210–268. Cambridge University Press, 2012.

R. Vershynin, P. Hsu, C. Ma, J. Nelson, E. Schnoor, D. Stoger, T. Sullivan, and T. Tao. High-dimensional probability: An introduction with applications in data science. 2017.