Supplemental Material for Top Feasible Arm Identification

Julian Katz-Samuels   Clayton Scott

February 25, 2019

A Outline and Notation

In this section, we provide an outline of the supplemental material and define some notation. In Section B, we prove Theorem 1. In Section C, we prove Theorem 3 and the main lemmas used in its proof. In Section D, we prove Theorem 4, splitting it up into three distinct theorems. In Section E, we discuss our conjecture that there is a small gap between δ- PAC and δ- PAC-EXPLANATORY algorithms; we also prove and discuss our lower bound for δ- PAC algorithms. In section F, we prove a number of technical lemmas. In Section G, we present and discuss a version of TF-LUCB that allows for tolerance of infeasibility and suboptimality. In section H, we provide pseudocode for TF-AE and FFAF.

Define $\mathcal{S}^{D-1} = \{\boldsymbol{x} \in \mathbb{R}^D : \|\boldsymbol{x}\|_2 = 1\}$. Define a function $d : (0,1) \times (0,1) \mapsto \mathbb{R}$ such that $d(x, y) := x \log(\frac{x}{y}) + (1-x)\log(\frac{1-x}{1-y})$. Recall that if $U = \emptyset$, then we use the convention $\min_{x \in U} x = \infty$ and $\max_{x \in U} x = -\infty$.

B Lower Bound

For the proof of Theorem 1, we introduce the following notation. For a given problem $(\nu, P, \boldsymbol{r}, m)$, define

$$\begin{aligned}
&\text{FEAS}(\nu, P, \boldsymbol{r}, m) = \{i \in [K] : \boldsymbol{\mu}_i \in P\}, \ \text{INFEAS}(\nu, P, \boldsymbol{r}, m) = \text{FEAS}(\nu, P, \boldsymbol{r}, m)^c, \\
&\text{OPT}(\nu, P, \boldsymbol{r}, m) = \{i \in \text{FEAS}(\nu, P, \boldsymbol{r}, m) : \boldsymbol{r}^\top \boldsymbol{\mu}_i \geq \max^{(m)}_{j \in \text{FEAS}(\nu, P, \boldsymbol{r}, m)} \boldsymbol{r}^\top \boldsymbol{\mu}_j\}, \\
&\text{SUBOPT}(\nu, P, \boldsymbol{r}, m) = \{i \in [K] : \boldsymbol{r}^\top \boldsymbol{\mu}_i < \max^{(m)}_{j \in \text{FEAS}(\nu, P, \boldsymbol{r}, m)} \boldsymbol{r}^\top \boldsymbol{\mu}_j\}.
\end{aligned}$$

Proof of Theorem 1. **Step 1: Pick a good partition of the arms.** Fix $\delta > 0$. Let $(\nu, P, \boldsymbol{r}, m)$ satisfy the hypotheses of the theorem statement. In the interest of brevity, abbreviate

$$\begin{aligned}
&\text{FEAS} := \text{FEAS}(\nu, P, \boldsymbol{r}, m), \ \text{INFEAS} := \text{INFEAS}(\nu, P, \boldsymbol{r}, m), \\
&\text{OPT} := \text{OPT}(\nu, P, \boldsymbol{r}, m), \ \text{SUBOPT} := \text{SUBOPT}(\nu, P, \boldsymbol{r}, m).
\end{aligned}$$

Let $\mathcal{A}$ denote a δ- PAC-EXPLANATORY algorithm wrt $\mathcal{M}$ with stopping time τ.

We claim that there exists $(S, I) \in$ Valid-Partitions that satisfies the following property:

$$i \in S \Longrightarrow \Pr_\nu(i \in \widehat{\text{S}}) \geq \frac{1-\delta}{2}; \quad i \in I \Longrightarrow \Pr_\nu(i \in \widehat{\text{I}}) \geq \frac{1-\delta}{2}. \tag{1}$$

As an intermediate step, we claim that for every $i \in \text{OPT}^c$,

$$\max(\Pr_\nu(i \in \widehat{\text{I}}), \Pr_\nu(i \in \widehat{\text{S}})) \geqslant \frac{1-\delta}{2}. \tag{2}$$

To see this, fix $i \in \text{OPT}^c$. Define the events

$$\begin{aligned} B &= \{\widehat{\text{O}} = \text{OPT}, (\widehat{\text{S}}, \widehat{\text{I}}) \in \text{Valid-Partitions}\}, \\ B_1 &= B \cap \{i \in \widehat{\text{S}}\}, \\ B_2 &= B \cap \{i \in \widehat{\text{I}}\}. \end{aligned}$$

Note that $B = B_1 \cup B_2$ and $B_1 \cap B_2 = \varnothing$. Since $\mathcal{A}$ is δ- PAC-EXPLANATORY wrt $\mathcal{M}$,

$$\begin{aligned} 1-\delta &\leqslant \Pr_\nu(B) \\ &= \Pr_\nu(B_1) + \Pr_\nu(B_2) \\ &\leqslant \Pr_\nu(i \in \widehat{\text{S}}) + \Pr_\nu(i \in \widehat{\text{I}}) \\ &\leqslant 2\max(\Pr_\nu(i \in \widehat{\text{S}}), \Pr_\nu(i \in \widehat{\text{I}})). \end{aligned}$$

This establishes the claim in (2). Furthermore, note that if $i \in \text{OPT}^c \setminus \text{INFEAS} = \text{SUBOPT} \cap \text{FEAS}$, then $B_2 = \varnothing$, so that

$$\Pr_\nu(i \in \widehat{\text{S}}) \geqslant \frac{1-\delta}{2}. \tag{3}$$

Similarly, if $i \in \text{OPT}^c \setminus \text{SUBOPT}$, then $B_1 = \varnothing$, so that

$$\Pr_\nu(i \in \widehat{\text{I}}) \geqslant \frac{1-\delta}{2}. \tag{4}$$

Define

$$\begin{aligned} S &= \{i \in \text{SUBOPT} : \Pr_\nu(i \in \widehat{\text{S}}) \geqslant \frac{1-\delta}{2}\} \\ I &= \text{INFEAS} \setminus \{i \in \text{SUBOPT} : \Pr_\nu(i \in \widehat{\text{S}}) \geqslant \frac{1-\delta}{2}\}. \end{aligned}$$

We claim that $(S, I) \in$ Valid-Partitions. Clearly, $S \subset \text{SUBOPT}$, $I \subset \text{INFEAS}$, $S \cap I = \varnothing$, and $S \cup I \subset \text{OPT}^c$. Therefore, it suffices to show that $\text{OPT}^c \subset S \cup I$. Let $i \in \text{OPT}^c$. If $i \in \text{INFEAS}$, then either $i \in I$ or $i \in S$, so suppose that $i \notin \text{INFEAS}$. Then, $i \in \text{OPT}^c \setminus \text{INFEAS} = \text{SUBOPT} \cap \text{FEAS} \subset S$ by (3). Thus, the claim that $(S, I) \in$ Valid-Partitions follows.

We claim that (S, I) has the property (1). Let $i \in S$. By definition of S, $\Pr_\nu(i \in \widehat{\text{S}}) \geqslant \frac{1-\delta}{2}$. Next, let $i \in I$. If $i \in \text{SUBOPT}$, then $i \in I \subset \text{INFEAS}$ and $i \notin S$ imply that $\Pr_\nu(i \in \widehat{\text{S}}) < \frac{1-\delta}{2}$. Then, by (2) $\Pr_\nu(i \in \widehat{\text{I}}) \geqslant \frac{1-\delta}{2}$. If $i \notin \text{SUBOPT}$, then (4) implies that $\Pr_\nu(i \in \widehat{\text{I}}) \geqslant \frac{1-\delta}{2}$. Thus, the claim follows.

Next, we outline the rest of our proof. For the rest of the proof, the S and I that we constructed are fixed. Using the fact that $\tau = \sum_{i=1}^K N_i(\tau)$, we will show that for this choice of S and I,

$$\mathbb{E}_\nu[\tau] = \sum_{i=1}^K \mathbb{E}_\nu[N_i(\tau)] \geqslant \frac{1}{15}\ln(\frac{1}{2\delta})[\sum_{i \in \text{OPT}} \max([\min_{j \in S} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)]^{-2},\ \text{dist}(\boldsymbol{\mu}_i, \partial P)^{-2}) \tag{5}$$

$$+ \sum_{i \in S}[\min_{j \in \text{OPT}} \boldsymbol{r}^\top(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)]^{-2} + \sum_{i \in I} \text{dist}(\boldsymbol{\mu}_i, P)^{-2}]. \tag{6}$$

To this end, we lower bound $\mathbb{E}_\nu[N_i(\tau)]$ for each of the distinct cases ($i \in$ OPT, $i \in S$, $i \in I$). To do this, we construct a related problem by modifying one of the distributions and applying Lemma F.1. The result will follow by taking the minimum of the right-hand side of (6) over all $(S', I') \in$ Valid-Partitions.

In each of the next steps, we will define a new problem to obtain a lower bound. To avoid notational clutter, we will redefine the symbols $\boldsymbol{\mu}'_i$, ν'_i, and $\nu^{(i)}$ in each step. The context should make their meaning clear.

Step 2.a: reward bound for $i \in$ OPT. Fix $i \in$ OPT. First, we show that

$$\mathbb{E}_\nu[N_i(\tau)] \geqslant \frac{2}{15} \ln(\frac{1}{2\delta})[\min_{j \in S} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \epsilon]^{-2}$$

for a sufficiently small $\epsilon > 0$. If $S = \varnothing$, $\min_{j \in S} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) = -\infty$ by definition and there is nothing to show. So, suppose that $S \neq \varnothing$. Define

$$j_0 = \arg\max_{j \in S} \boldsymbol{r}^\top \boldsymbol{\mu}_j. \tag{7}$$

Define for all $j \in [K]$

$$\boldsymbol{\mu}'_j = \begin{cases} \begin{pmatrix} \boldsymbol{\mu}_{j_0,1} - \epsilon \\ \boldsymbol{\mu}_{i,2:D} \end{pmatrix} & \text{if } j = i \\ \boldsymbol{\mu}_j & \text{if } j \neq i \end{cases}$$
$$\nu'_j = N(\boldsymbol{\mu}'_j, I_D).$$

where $\epsilon > 0$ is chosen sufficiently small such that for all $\delta \in [0, \epsilon)$, $\boldsymbol{r}^\top \boldsymbol{\mu}'_i + \delta \neq \boldsymbol{r}^\top \boldsymbol{\mu}'_j$ for all $j \neq i$ (which is possible since $\boldsymbol{r}^\top \boldsymbol{\mu}_l \neq \boldsymbol{r}^\top \boldsymbol{\mu}_k$ for all $l \neq k \in [K]$). Define $\nu^{(i)} = (\nu'_1, \ldots, \nu'_K)$ and consider the problem $(\nu^{(i)}, P, \boldsymbol{r}, m)$. We claim that $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$. Since $\boldsymbol{\mu}_i \notin \partial P$ and $\partial P = \partial(\mathbb{R} \times P') = \mathbb{R} \times \partial P'$ for some $P' \subset \mathbb{R}^{D-1}$, $\boldsymbol{\mu}'_i \notin \partial P$. Further, by construction, $\boldsymbol{r}^\top \boldsymbol{\mu}'_i \neq \boldsymbol{r}^\top \boldsymbol{\mu}'_j$ for all $j \neq i$. Thus, none of the arms have means on the boundary of P and all of the rewards of the arms are distinct, so $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$.

In the interest of brevity, abbreviate

$$\text{FEAS}_i \coloneqq \text{FEAS}(\nu^{(i)}, P, \boldsymbol{r}, m), \ \text{SUBOPT}_i \coloneqq \text{SUBOPT}(\nu^{(i)}, P, \boldsymbol{r}, m).$$

We claim that $j_0 \notin \text{SUBOPT}_i$. Suppose $j_0 \in$ FEAS. Then,

$$\begin{aligned} \boldsymbol{r}^\top \boldsymbol{\mu}'_{j_0} &= \boldsymbol{r}^\top \boldsymbol{\mu}_{j_0} \\ &= \max_{l \in S} \boldsymbol{r}^\top \boldsymbol{\mu}_l & (8) \\ &\geqslant \max_{l \in \text{SUBOPT} \cap \text{FEAS}} \boldsymbol{r}^\top \boldsymbol{\mu}_l & (9) \\ &= \max^{(m+1)}_{l \in \text{FEAS}} \boldsymbol{r}^\top \boldsymbol{\mu}_l & (10) \\ &= \max^{(m)}_{l \in \text{FEAS}_i} \boldsymbol{r}^\top \boldsymbol{\mu}'_l & (11) \end{aligned}$$

where line (8) follows from (7), line (9) follows from $S \supset$ SUBOPT $\cap$ FEAS, (10) follows from $j_0 \in$ FEAS by assumption, and (11) follows from the fact that $j_0 \in$ FEAS and the only difference between ν and $\nu^{(i)}$ is in the ith arm, which now has reward less than the j_0th arm. Thus, if $j_0 \in$ FEAS, then $j_0 \notin \text{SUBOPT}_i$.

On the other hand, if $j_0 \notin$ FEAS, then

$$\boldsymbol{r}^\top \boldsymbol{\mu}'_{j_0} = \max_{l \in S} \boldsymbol{r}^\top \boldsymbol{\mu}'_l \tag{12}$$

$$> \boldsymbol{r}^\top \boldsymbol{\mu}'_i \tag{13}$$

$$> \max_{l \in \text{SUBOPT} \cap \text{FEAS}} \boldsymbol{r}^\top \boldsymbol{\mu}'_l \tag{14}$$

where line (12) follows from (7) and $\boldsymbol{\mu}'_l = \boldsymbol{\mu}_l$ for all $l \in S$, line (13) follows from $\boldsymbol{\mu}'_i$ is defined to satisfy $\boldsymbol{r}^\top \boldsymbol{\mu}'_i > \max_{l : \boldsymbol{r}^\top \boldsymbol{\mu}'_l < \boldsymbol{r}^\top \boldsymbol{\mu}'_{j_0}} \boldsymbol{r}^\top \boldsymbol{\mu}'_l$, and line (14) follows from $S \supset$ SUBOPT $\cap$ FEAS, $j_0 \notin$ FEAS, and $\{\boldsymbol{r}^\top \boldsymbol{\mu}_l^{|} prime\}_{l \in [K]}$ distinct. (14) implies that

$$\max^{(m)}_{l \in \text{FEAS}_i} \boldsymbol{r}^\top \boldsymbol{\mu}'_l = \boldsymbol{r}^\top \boldsymbol{\mu}'_i < \boldsymbol{r}^\top \boldsymbol{\mu}'_{j_0}$$

so that $j_0 \notin \text{SUBOPT}_i$. This establishes the claim that $j_0 \notin \text{SUBOPT}_i$.

Consider the event $B = \{j_0 \in \widehat{\text{S}}\}$. Then, since $\mathcal{A}$ is δ- PAC-EXPLANATORY wrt to $\mathcal{M}$, $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$, and arm $j_0 \notin \text{SUBOPT}_i$, we have that

$$\Pr_{\nu^{(i)}}(B) \leqslant \Pr_{\nu^{(i)}}(\widehat{\text{S}} \not\subset \text{SUBOPT}_i) \leqslant \delta. \tag{15}$$

Further, by construction of S,

$$\Pr_\nu(B) \geqslant \frac{1-\delta}{2}. \tag{16}$$

Then,

$$\frac{1}{2}[\boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{j_0}) + \epsilon]^2 \mathbb{E}_\nu[N_i(\tau)] = \text{KL}(\nu_i, \nu'_i)\mathbb{E}_\nu[N_i(\tau)] \tag{17}$$

$$\geqslant d(\Pr_\nu(B), \Pr_{\nu^{(i)}}(B)) \tag{18}$$

$$\geqslant d(\Pr_\nu(B), \delta) \tag{19}$$

$$\geqslant d(\frac{1-\delta}{2}, \delta) \tag{20}$$

$$\geqslant \frac{1}{15} \ln(\frac{1}{2\delta}). \tag{21}$$

Line (17) follows by the formula for the KL-divergence of two multivariate normal distributions, (18) follows by Lemma F.1, (19) follows since $y \mapsto d(x, y)$ is decreasing when $x > y$, (15), (16), and $\delta < .1$, (20) follows since $x \mapsto d(x, y)$ is increasing when $x > y$, (15), (16), and $\delta < .1$, and (21) follows by Lemma (F.7). The claim follows by rearranging the inequality.

Step 2.b: feasibility bound for $i \in$ OPT. Next, we show that for sufficiently small $\epsilon > 0$,

$$\mathbb{E}_\nu[N_i(\tau)] \geqslant \frac{2}{15} \ln(\frac{1}{2\delta})[\,\text{dist}(\boldsymbol{\mu}_i, \partial P) + \epsilon]^{-2}.$$

Since P is nonempty and $P \neq \mathbb{R}^D$, by Lemma F.6 ∂P is nonempty. Since in addition ∂P is closed, by Lemma F.2, there exists $\boldsymbol{\tau}_i \in \text{Proj}_{\partial P}(\boldsymbol{\mu}_i)$. Since $\boldsymbol{\tau}_i \in \partial P$, by the assumptions of the Theorem on P, for all $\epsilon > 0$, $B_\epsilon(\tau) \cap (P^c)^\circ \neq \varnothing$. Thus, for any $\epsilon > 0$, there exists a direction $\boldsymbol{v} \in \mathbb{R}^D$ with $\|\boldsymbol{v}\|_2 = 1$ such that $\boldsymbol{\tau}_i + \epsilon \boldsymbol{v} \in (P^c)^\circ$. Further, since by the assumptions of the Theorem on P, $P = \mathbb{R} \times P'$ for some $P' \subset \mathbb{R}^{D-1}$, we can choose $\boldsymbol{v}$ such that $v_1 = 0$.

Define for $j \in [K]$

$$\boldsymbol{\mu}'_j = \begin{cases} \boldsymbol{\tau}_i + \epsilon \boldsymbol{v} & \text{if } j = i \\ \boldsymbol{\mu}_j & \text{if } j \neq i \end{cases}$$
$$\nu'_j = N(\boldsymbol{\mu}'_j, I_D).$$

Define $\nu^{(i)} = (\nu'_1, \ldots, \nu'_K)$ and consider the problem $(\nu^{(i)}, P, \boldsymbol{r}, m)$. We claim that $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$. Since $\boldsymbol{\mu}'_i \in (P^c)^\circ$, $\boldsymbol{\mu}'_i \notin \partial P$. Therefore, it suffices to show that $\boldsymbol{r}^\top \boldsymbol{\mu}'_i \neq \boldsymbol{r}^\top \boldsymbol{\mu}'_j$ for all $j \neq i$. To show this, it suffices to show that $\tau_{i,1} = \mu_{i,1}$ since then it follows by our choice of $\boldsymbol{v}$, $\boldsymbol{r} = \boldsymbol{e}_1$, and the fact that for all $j \neq i$, $\boldsymbol{r}^\top \boldsymbol{\mu}_j \neq \boldsymbol{r}^\top \boldsymbol{\mu}_i$. Towards a contradiction, suppose that $\mu_{i,1} \neq \tau_{i,1}$. Define

$$\tau'_{i,j} = \begin{cases} \tau_{i,j} & : j \neq 1 \\ \mu_{i,1} & : \text{otherwise} \end{cases}.$$

Recall that $P = \mathbb{R} \times P'$ for some $P' \subset \mathbb{R}^{D-1}$ and observe that $\partial P = \partial \mathbb{R} \times P' \cup \mathbb{R} \times \partial P' = \mathbb{R} \times \partial P'$. Thus, $\boldsymbol{\tau}_i \in \partial P$ implies that $\boldsymbol{\tau}'_i \in \partial P$. Further, $\|\boldsymbol{\tau}'_i - \boldsymbol{\mu}_i\|_2 < \|\boldsymbol{\tau}_i - \boldsymbol{\mu}_i\|_2$, which is a contradiction to $\boldsymbol{\tau}_i \in \partial P$. Thus, the claim follows and hence $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$.

In the interest of brevity, abbreviate

$$\text{FEAS}_i \coloneqq \text{FEAS}(\nu^{(i)}, P, \boldsymbol{r}, m), \ \text{OPT}_i \coloneqq \text{OPT}(\nu_P^{(i)}, P, \boldsymbol{r}, m).$$

Define the event $B = \{i \in \widehat{\text{O}}\}$. Then, $i \notin \text{FEAS}_i$, so that the event B implies that the algorithm $\mathcal{A}$ makes a mistake. Since $\mathcal{A}$ is δ-PAC-EXPLANATORY wrt $\mathcal{M}$, $\Pr_{\nu^{(i)}}(B) \leqslant \Pr_{\nu^{(i)}}(\widehat{\text{O}} \not\subseteq \text{OPT}_i) \leqslant \delta$. Further, since $i \in \text{OPT}$ and $\mathcal{A}$ is δ-PAC-EXPLANATORY wrt $\mathcal{M}$,

$$\Pr_\nu(B) \geqslant \Pr_\nu(\widehat{\text{O}} = \text{OPT}) \geqslant 1 - \delta \geqslant \frac{1-\delta}{2}.$$

Thus,

$$\frac{1}{2}(\,\text{dist}(\boldsymbol{\mu}_i, \partial P) + \epsilon)^2 \mathbb{E}_\nu[N_i(\tau)] = \frac{1}{2}(\|\boldsymbol{\tau}_i - \boldsymbol{\mu}_i\|_2 + \epsilon)^2 \mathbb{E}_\nu[N_i(\tau)] \tag{22}$$
$$\geqslant \frac{1}{2}\|\boldsymbol{\tau}_i + \epsilon \boldsymbol{v} - \boldsymbol{\mu}_i\|_2^2 \, \mathbb{E}_\nu[N_i(\tau)] \tag{23}$$
$$= \text{KL}(\nu_i, \nu'_i) \mathbb{E}_\nu[N_i(\tau)] \tag{24}$$
$$\geqslant \frac{1}{15} \ln(\frac{1}{2\delta}). \tag{25}$$

Line (22) follows by the definition of $\boldsymbol{\tau}_i$, line (23) follows by the triangle inequality and $\|v\|_2 = 1$, line (24) follows by the definition of the KL divergence for multivariate normal distributions, and line (25) follows by a similar series of inequalities as (17)-(21).

Step 3: $i \in S$**.** If $S = \varnothing$, then there is nothing to show in this step. So, suppose that $S \neq \varnothing$. Then, $S \neq \varnothing$ implies that there are at least m feasible arms. Let $j_0 \in [K]$ such that $\boldsymbol{\mu}_{j_0} \in P$ and

$$\boldsymbol{r}^\top \boldsymbol{\mu}_{j_0} = \min_{l \in \text{OPT}} \boldsymbol{r}^\top \boldsymbol{\mu}_l.$$

Define for $j \in [K]$

$$\boldsymbol{\mu}'_j = \begin{cases} \begin{pmatrix} \mu_{i,1} + \mu_{j_0,1} + \epsilon \\ \boldsymbol{\mu}_{i,2:D} \end{pmatrix} & \text{if } j = i \\ \boldsymbol{\mu}_j & \text{if } j \neq i \end{cases}$$
$$\nu'_j = N(\boldsymbol{\mu}'_j, I_D).$$

where $\epsilon > 0$ is chosen sufficiently small so that for any $\delta \in [0, \epsilon)$, $\boldsymbol{r}^\top \boldsymbol{\mu}'_i - \delta \neq \boldsymbol{r}^\top \boldsymbol{\mu}'_j$ for all $j \neq i$ (which is possible since $\boldsymbol{r}^\top \boldsymbol{\mu}_l \neq \boldsymbol{r}^\top \boldsymbol{\mu}_k$ for all $l \neq k \in [K]$). Define $\nu^{(i)} = (\nu'_1, \ldots, \nu'_K)$ and consider the problem $(\nu^{(i)}, P, \boldsymbol{r}, m)$. It follows that $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$ by a similar argument that showed in Step 2.a that when $i \in$ OPT, $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$.

In the interest of brevity, abbreviate

$$\text{SUBOPT}_i \coloneqq \text{SUBOPT}(\nu^{(i)}, P, \boldsymbol{r}, m).$$

Define $B = \{i \in \widehat{\text{S}}\}$. Note that arm $i \notin \text{SUBOPT}_i$ by construction. Thus, since $\mathcal{A}$ is δ- PAC-EXPLANATORY wrt $\mathcal{M}$, we have that $\Pr_{\nu^{(i)}}(B) \leqslant \delta$. Further, by construction of S, $\Pr_\nu(B) \geqslant \frac{1-\delta}{2}$. Therefore, by a similar series of inequalities as (17)-(21), it follows that

$$\frac{1}{15} \ln(\frac{1}{2\delta}) \leqslant \frac{1}{2} [\boldsymbol{r}^\top (\boldsymbol{\mu}_{j_0} - \boldsymbol{\mu}_i) + \epsilon]^2 \mathbb{E}_\nu[N_i(\tau)]. \tag{26}$$

Step 4: $i \in I$**.** Since $P \neq \mathbb{R}^D$ and P is nonempty, by Lemma F.6 ∂P is nonempty. Since in addition ∂P is closed, by Lemma F.2, there exists $\boldsymbol{\tau}_i \in \text{Proj}_{\partial P}(\boldsymbol{\mu}_i)$. By the assumptions of the Theorem on P, since $\boldsymbol{\tau}_i \in \partial P$, for every $\epsilon > 0$, $B_\epsilon(\boldsymbol{\tau}_i) \cap P^\circ \neq \varnothing$. Thus, for sufficiently small $\epsilon > 0$, there exists a direction $\boldsymbol{v} \in \mathbb{R}^D$ with $\|\boldsymbol{v}\|_2 = 1$ such that $\boldsymbol{\tau}_i + \epsilon \boldsymbol{v} \in P^\circ$. Since by the assumptions of the Theorem on P, $P = \mathbb{R} \times P'$ for some $P' \subset \mathbb{R}^{D-1}$, we can choose $\boldsymbol{v}$ such that $v_1 = 0$. Define for $j \in [K]$

$$\boldsymbol{\mu}'_j = \begin{cases} \boldsymbol{\tau}_i + \epsilon \boldsymbol{v} & \text{if } j = i \\ \boldsymbol{\mu}_j & \text{if } j \neq i \end{cases}$$
$$\nu'_j = N(\boldsymbol{\mu}'_j, I_D).$$

Define $\nu^{(i)} = (\nu'_1, \ldots, \nu'_K)$ and consider the problem $(\nu^{(i)}, P, \boldsymbol{r}, m)$. It follows that $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$ by a similar argument that showed in step 2.b that when $i \in$ OPT, $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$.

In the interest of brevity, abbreviate

$$\text{INFEAS}_i \coloneqq \text{INFEAS}(\nu^{(i)}, P, \boldsymbol{r}, m).$$

Define the event $B = \{i \in \widehat{\text{I}}\}$. Observe that $i \notin \text{INFEAS}_i$. Thus, since $\mathcal{A}$ is δ- PAC-EXPLANATORY wrt $\mathcal{M}$,

$$\Pr_{\nu^{(i)}}(B) \leqslant \Pr_{\nu^{(i)}}(\widehat{\text{I}} \not\subset \text{INFEAS}_i) \leqslant \delta.$$

Further, by construction of I, $\Pr_\nu(i \in \widehat{\text{I}}) \geqslant \frac{1-\delta}{2}$. Therefore, by a similar series of inequalities as (22)-(25), it follows that

$$\frac{1}{15} \ln(\frac{1}{2\delta}) \leqslant \frac{1}{2} (\,\text{dist}(\boldsymbol{\mu}_i, P) + \epsilon)^2 \mathbb{E}_\nu[N_i(\tau)]. \tag{27}$$

Step 5: Putting it together. Using $\mathbb{E}_\nu[\tau] = \sum_{i=1}^K \mathbb{E}_\nu[N_i(\tau)]$ and inequalities (21), (25), (26), and (27), we establish for all sufficiently small $\epsilon > 0$,

$$\begin{aligned}\mathbb{E}_\nu[\tau] \geqslant \frac{2}{15}\ln(\frac{1}{2\delta})[\sum_{i\in \text{OPT}} \max([\min_{j\in S} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \epsilon]^{-2}, [\,\text{dist}(\boldsymbol{\mu}_i, \partial P) + \epsilon]^{-2}) \\ + \sum_{i\in S}[\min_{j\in \text{OPT}} \boldsymbol{r}^\top(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i) + \epsilon]^{-2} + \sum_{i\in I}[\,\text{dist}(\boldsymbol{\mu}_i, P) + \epsilon]^{-2}].\end{aligned}$$

Since this bound holds for all $\epsilon > 0$ sufficiently small, letting $\epsilon \longrightarrow 0$ on the RHS of the above inequality establishes (6).

□

C Proof of Theorem 3

To begin, we introduce some notation. Fix $(S, I) \in$ Valid-Partitions. We will bound the number of samples required to identify each arm as belonging to either OPT, S, or I. Define

$$d(S) = \frac{\min_{i\in\text{OPT}} \boldsymbol{r}^\top \boldsymbol{\mu}_i + \max_{j\in S} \boldsymbol{r}^\top \boldsymbol{\mu}_j}{2}.$$

If either $|\,\text{FEAS}\,| < m$ or $S = \varnothing$, then define $d(S) \coloneqq -\infty$. Next, we introduce a notion, which captures when arm i needs to be pulled more. Define for all $i \in [K]$,

$$\begin{aligned}\text{NEEDY}_i^t(S, I) = [\{i \in \text{OPT}\} \wedge (\{i \in G_t\} \vee \{\boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(t)} - U_{\boldsymbol{r}}(N_i(t), \delta) \leqslant d(S)\})] \\ \vee [\{i \in S\} \wedge \{\boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(t)} + U_{\boldsymbol{r}}(N_i(t), \delta) \geqslant d(S)\}] \\ \vee [\{i \in I\} \wedge \{i \in G_t\}]\end{aligned}$$

In words, if arm i is optimal, then it needs to be pulled more if either it has not been determined whp that $\boldsymbol{\mu}_i \in P$ or the lower bound on its reward is below $d(S)$. If i is in S, then it needs to be pulled more if the upper bound on its reward is above $d(S)$, and if i is in I, then it needs to be pulled more if it has not been determined that $\boldsymbol{\mu}_i \notin P$.

Next, we state the two main lemmas that we use in the proof of Theorem 3.

Lemma C.1. *Fix $\delta > 0$ and a problem $(\nu, P, \boldsymbol{r}, m) \in \mathcal{M}$. Fix $(S, I) \in$ Valid-Partitions. Suppose that for all $i \in [K]$ and for all $t \geqslant 1$,* (i) *it holds that*

$$|\boldsymbol{r}^\top(\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{i,t})| \leqslant U_{\boldsymbol{r}}(t, \delta), \tag{28}$$

and (ii) *TestF(i, t) = True implies that $\boldsymbol{\mu}_i \in P$ and TestF(i, t) = False implies that $\boldsymbol{\mu}_i \notin P$. Then, for all t prior to termination (i.e., $t < \tau$), NEEDY$_{l_t}^t(S, I) \vee$ NEEDY$_{h_t}^t(S, I)$ is true.*

Lemma C.1 essentially says that provided *(i)* $U_{\boldsymbol{r}}(t, \delta)$ bounds the deviation $|\boldsymbol{r}^\top(\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{i,t})|$ and *(ii)* TestF does not make a mistake, then every round prior to termination, at least one of the pulled arms is "needy."

The second main lemma states that provided that *(i)* $U_{\boldsymbol{r}}(t, \delta)$ bounds the deviation $|\boldsymbol{r}^\top(\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{i,t})|$ and *(ii)* TestF does not make a mistake, the algorithm returns a correct answer, i.e., returns $(\widehat{\text{O}}, \widehat{\text{S}}, \widehat{\text{I}})$ such that $\widehat{\text{O}} = \text{OPT}$, $\widehat{\text{S}} \subset \text{SUBOPT}$, and $\widehat{\text{I}} \subset \text{INFEAS}$.

Lemma C.2. *Fix $\delta > 0$ and a problem $(\nu, P, \boldsymbol{r}, m) \in \mathcal{M}$. Suppose that for all $i \in [K]$ and $t \in \mathbb{N}$,* (i) *it holds that*

$$|\boldsymbol{r}^\top(\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{i,t})| \leqslant U_{\boldsymbol{r}}(t, \delta), \tag{29}$$

and (ii) *TestF(i,t) = True implies that $\boldsymbol{\mu}_i \in P$ and TestF(i,t) = False implies that $\boldsymbol{\mu}_i \notin P$. Then, TF-LUCB($\delta$) returns $(\widehat{O}, \widehat{S}, \widehat{I})$ such that $\widehat{O}$ = OPT, and $(\widehat{S}, \widehat{I}) \in$ Valid-Partitions.*

The proofs of the two lemmas are given in Section C.1.

Next, we prove Theorem 3. The proof has three main steps. First, we show that whp for every arm i *(i)* TestF does not make a mistake about the feasibility of arm i, *(ii)* after arm i has been pulled $\eta(\nu_i, P)$ times, TestF determines whether arm i is feasible, and *(iii)* $U_{\boldsymbol{r}}(t, \delta)$ controls the deviation of the empirical mean reward to the expected reward for arm i. Second, we apply Lemma C.2 to conclude that the algorithm returns the correct answer. Finally, we upper bound the sample complexity, τ, of the algorithm by essentially upper bounding how many times an arm must be pulled before no longer being "needy."

Proof of Theorem 3. **Step 1: Defining the event.** Let $(S, I) \in$ Valid-Partitions that achieves the minimum in the upper bound (1) stated in Theorem 3. For the sake of brevity, we write NEEDY_i^t and d instead of $\text{NEEDY}_i^t(S, I)$ and $d(S)$, respectively.

If $\boldsymbol{\mu}_i \in P$, let

$$\begin{aligned} B_i =& \{\forall t \in \mathbb{N} : \text{TestF}(i,t) \neq \text{ False}\} \cap \{\forall t \geqslant \eta(\nu_i, P) : \text{TestF}(i,t) = \text{ True}\} \\ &\cap \{\forall t \in \mathbb{N} : |\boldsymbol{r}^\top(\widehat{\boldsymbol{\mu}}_{i,t} - \boldsymbol{\mu}_i)| \leqslant U_{\boldsymbol{r}}(t, \delta)\}. \end{aligned}$$

If $\boldsymbol{\mu}_i \notin P$, let

$$\begin{aligned} B_i =& \{\forall t \in \mathbb{N} : \text{TestF}(i,t) \neq \text{ True}\} \cap \{\forall t \geqslant \eta(\nu_i, P) : \text{TestF}(i,t) = \text{ False}\} \\ &\cap \{\forall t \in \mathbb{N} : |\boldsymbol{r}^\top(\widehat{\boldsymbol{\mu}}_{i,t} - \boldsymbol{\mu}_i)| \leqslant U_{\boldsymbol{r}}(t, \delta)\}. \end{aligned}$$

In words, when $\boldsymbol{\mu}_i \in P$, B_i says that *(i)* TestF does not make the mistake of concluding that arm i is infeasible, *(ii)* after arm i has been pulled $\eta(\nu_i, P)$ times, TestF determines that arm i is feasible, and *(iii)* $U_{\boldsymbol{r}}(t, \delta)$ controls the deviation of the empirical mean reward to the expected reward of arm i. For $\boldsymbol{\mu}_i \notin P$, B_i is the analogous event.

Observe that since $\|\boldsymbol{r}\|_2 = 1$ and ν_i is σ-sub-Gaussian, if $\boldsymbol{X} \sim \nu_i$, then

$$\left\|\boldsymbol{r}^\top \boldsymbol{X}\right\|_{\psi_2} \leqslant \|\boldsymbol{X}\|_{\psi_2} \leqslant \sigma$$

so that $\boldsymbol{r}^\top \boldsymbol{X}$ is σ-sub-Gaussian.

Then, by the union bound,

$$\Pr(\cup_{i=1}^K B_i^c) \tag{30}$$

$$\leqslant \sum_{i\in[K]} \Pr(B_i^c) \tag{31}$$

$$\leqslant \sum_{i\in\text{FEAS}} \Pr([\{\forall t \in \mathbb{N} : \text{TestF}(i,t) \neq \text{ False}\} \cap \{\forall t \geqslant \eta(\nu_i, P) : \text{TestF}(i,t) = \text{ True}\}]^c) \tag{32}$$

$$+ \sum_{i\in\text{INFEAS}} \Pr([\{\forall t \in \mathbb{N} : \text{TestF}(i,t) \neq \text{ True}\} \cap \{\forall t \geqslant \eta(\nu_i, P) : \text{TestF}(i,t) = \text{ False}\}]^c) \tag{33}$$

$$+ \sum_{i\in[K]} \Pr(\exists t \in \mathbb{N} : |\boldsymbol{r}^\top(\widehat{\boldsymbol{\mu}}_{i,t} - \boldsymbol{\mu}_i)| > U_{\boldsymbol{r}}(t,\delta)) \tag{34}$$

$$\leqslant \sum_{i\in[K]} 2\frac{\delta}{2K} \tag{35}$$

$$= \delta, \tag{36}$$

where line (35) follows by Lemma F.10 and the assumption on TestF 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 uses at most $\eta(\xi, R)$ samples. For the rest of the proof, we assume $\cap_{i\in[K]} E_i$.

Step 2: Correctness. On event $\cap_{i\in[K]} B_i$, the conditions of Lemma C.2 are satisfied, so that TF-LUCB returns $(\widehat{\text{O}}, \widehat{\text{S}}, \widehat{\text{I}})$ such that $\widehat{\text{O}} = \text{OPT}$, $\widehat{\text{S}} \subset \text{SUBOPT}$ and $\widehat{\text{I}} \subset \widehat{\text{I}}$.

Step 3: Sample Complexity. Next, we bound the sample complexity of TF-LUCB, i.e., prove (1) in the statement of Theorem 3. If $i \in \text{OPT}$, let ρ_i denote the smallest integer such that $\forall t \geqslant \rho_i$

$$U_{\boldsymbol{r}}(t,\delta) < \frac{\min_{j\in S} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}{4}. \tag{37}$$

We claim that for all $i \in \text{OPT}$ and $s \in \mathbb{N}$, if $N_i(s) \geqslant \max(\rho_i, \eta(\nu_i, P))$, then $\text{NEEDY}_i^s = 0$. Let $i \in \text{OPT}$. Let $N_i(s) \geqslant \max(\rho_i, \eta(\nu_i, P))$. Then, on event B_i, TestF$(i, N_i(s)) =$ True, which implies that $i \notin G_s$. Further,

$$\boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(s)} - U_{\boldsymbol{r}}(N_i(s),\delta) \geqslant \boldsymbol{r}^\top \boldsymbol{\mu}_i - 2\,U_{\boldsymbol{r}}(N_i(s),\delta) \tag{38}$$

$$\geqslant \boldsymbol{r}^\top \boldsymbol{\mu}_i - \frac{\min_{j\in S} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}{2} \tag{39}$$

$$\geqslant d \tag{40}$$

where line (38) follows by event B_i and line (39) follows by (37). Thus, $\text{NEEDY}_i^s = 0$.

If $i \in S$, let ρ_i denote the smallest integer such that $\forall t \geqslant \rho_i$

$$U_{\boldsymbol{r}}(t,\delta) < \frac{\min_{j\in\text{OPT}} \boldsymbol{r}^\top \boldsymbol{\mu}_j - \boldsymbol{r}^\top \boldsymbol{\mu}_i}{4}. \tag{41}$$

We claim that for all $i \in S$ and $s \in \mathbb{N}$, if $N_i(s) \geqslant \rho_i$, then $\text{NEEDY}_i^s = 0$. Observe that

$$\boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(s)} + U_{\boldsymbol{r}}(N_i(s),\delta) \leqslant \boldsymbol{r}^\top \boldsymbol{\mu}_i + 2\,U_{\boldsymbol{r}}(N_i(s),\delta) \tag{42}$$

$$\leqslant \boldsymbol{r}^\top \boldsymbol{\mu}_i + \frac{\min_{j\in\text{OPT}} \boldsymbol{r}^\top(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)}{2} \tag{43}$$

$$\leqslant d \tag{44}$$

where line (42) follows by event B_i, and (43) follows by (41). Thus, $\text{NEEDY}_i^s = 0$.

Finally, let $i \in I$. Then, $N_i(s) \geqslant \eta(\nu_i, P)$ implies by event B_i that $\text{TestF}(i,t) =$ False, so that $i \notin G_s$. Thus, $\text{NEEDY}_i^s = 0$.

Then,

$$\tau - 1 \leqslant \sum_{t=1}^{\infty} \mathbf{1}\{\text{NEEDY}_{h_t}^t = 1 \text{ or } \text{NEEDY}_{l_t}^t = 1\} \tag{45}$$

$$\leqslant \sum_{t=1}^{\infty} \sum_{i=1}^{K} \mathbf{1}\{h_t = i \text{ or } l_t = i\} \mathbf{1}\{\text{NEEDY}_i^t = 1\} \tag{46}$$

$$\leqslant \sum_{t=1}^{\infty} \sum_{i \in \text{OPT}} \big[\mathbf{1}\{h_t = i \text{ or } l_t = i\} \mathbf{1}\{N_i(t) \leqslant \max(\rho_i, \eta(\nu_i, P))\} \tag{47}$$

$$+ \sum_{i \in S} \mathbf{1}\{h_t = i \text{ or } l_t = i\} \mathbf{1}\{N_i(t) \leqslant \rho_i\} \tag{48}$$

$$+ \sum_{i \in I} \mathbf{1}\{h_t = i \text{ or } l_t = i\} \mathbf{1}\{N_i(t) \leqslant \eta(\nu_i, P)\}\big] \tag{49}$$

$$\leqslant \sum_{i \in \text{OPT}} \max(\rho_i, \eta(\nu_i, P)) + \sum_{i \in S} \rho_i + \sum_{i \in I} \eta(\nu_i, P). \tag{50}$$

Line (45) follows by Lemma C.1; line (47) follows by the contrapositive of the claim that for $i \in \text{OPT}$ and $s \in \mathbb{N}$, if $N_i(s) \geqslant \max(\rho_i, \eta(\nu_i, P))$, then $\text{NEEDY}_i^s = 0$; lines (48) and (49) follow by the contrapositives of the analogous claims for $i \in S$ and $i \in I$; line (50) follows by exchanging the summations via Tonelli's theorem for series and if $h_t = i$ or $l_t = i$, then $N_i(t+1) = N_t(t) + 1$.

By Lemma F.11, for $i \in \text{OPT}$,

$$\rho_i \leqslant c\sigma^2 [\min_{j \in S} \boldsymbol{r}^\top (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)]^{-2}) \log(\log([\min_{j \in S} \boldsymbol{r}^\top (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)]^{-2}) \frac{K}{\delta}).$$

where c is a universal positive constant. By Lemma F.11, for $i \in S$,

$$\rho_i \leqslant c\sigma^2 [\min_{j \in \text{OPT}} \boldsymbol{r}^\top (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)]^{-2} \log(\log([\min_{j \in \text{OPT}} \boldsymbol{r}^\top (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)]^{-2}) \frac{K}{\delta})$$

where c is a universal positive constant. The result follows. □

C.1 Main Lemmas

Define the sets

$$\text{ABOVE}_t(S) = \{i \in [K] : \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i, N_i(t)} - U_{\boldsymbol{r}}(N_i(t), \delta) > d(S)\}$$
$$\text{BELOW}_t(S) = \{i \in [K] : \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i, N_i(t)} + U_{\boldsymbol{r}}(N_i(t), \delta) < d(S)\}$$
$$\text{MIDDLE}_t(S) = [K] \backslash (\text{ABOVE}_t(S) \cup \text{BELOW}_t(S))$$

Recall that $d(S)$ is the average of the smallest reward among the arms in OPT and the largest reward among the arms in S. Note that $d(S)$ is not known to the agent. Hence, $\text{ABOVE}_t(S)$ are the arms that at time t it is clear that whp their rewards are greater than the rewards of the arms in S and, similarly, $\text{BELOW}_t(S)$

are the arms that at time t it is clear that whp their rewards are less than the rewards of the arms in OPT. $\text{MIDDLE}_t(S)$ are the arms for which more evidence must be collected about their rewards to determine whether their reward is greater than or less than $d(S)$.

Proof of Lemma C.1. Fix $(S, I) \in$ Valid-Partitions. Let t be some round prior to termination, i.e., $t < \tau$. For the sake of brevity, we write NEEDY_i^t, d, ABOVE_t, BELOW_t, and MIDDLE_t instead of $\text{NEEDY}_i^t(S, I)$, $d(S)$, $\text{ABOVE}_t(S)$, $\text{BELOW}_t(S)$, and $\text{MIDDLE}_t(S)$ respectively.

Case 1: $|\,\textbf{FEAS}\,| < m$. Then, SUBOPT $= \emptyset$ so that $S = \emptyset$. We claim that $h_t \in G_t$. Towards a contradiction, suppose that $h_t \notin G_t$. Since $h_t \in \text{TOP}_t \subset E_t$, if $h_t \notin G_t$, then $h_t \in F_t$. Then, by lines 11 and 13 of the algorithm, $\text{TOP}_t \subset F_t$. Either *(i)* $|\,\text{TOP}_t\,| < m$ or *(ii)* $|\,\text{TOP}_t\,| = m$. Suppose $|\,\text{TOP}_t\,| < m$. Then, the definition of TOP_t implies that

$$E_t = \underset{t}{\text{TOP}} \subset F_t \subset E_t,$$

so that $\text{TOP}_t = F_t = E_t$. Thus, that t is the last round, i.e., $t = \tau$, which is a contradiction. Next, assume that $|\,\text{TOP}_t\,| = m$. Since by assumption $|\,\text{FEAS}\,| < m$ there exists $i \in$ INFEAS such that $\text{TestF}(i, t) =$ True, which is a contradiction. Thus, $h_t \in G_t$.

Since $S = \emptyset$, $h_t \in \text{OPT} \cup I$, which implies $\text{NEEDY}_{h_t}^t = 1$.

Case 2: $|\,\textbf{FEAS}\,| \geqslant m$. We split the rest of the proof up into cases, where in each case we show either that $\text{NEEDY}_{h_t}^t = 1$, $\text{NEEDY}_{l_t}^t = 1$, or there is a contradiction. We briefly make two useful observations that follow from the assumption that $\text{TestF}(i, t) =$ True implies that $\boldsymbol{\mu}_i \in P$ and $\text{TestF}(i, t) =$ False implies that $\boldsymbol{\mu}_i \notin P$. First, the assumption implies that FEAS $\subset E_t$ for all t, so that $m \leqslant |\,\text{FEAS}\,| \leqslant |E_t|$ and furthermore by the definition of TOP_t, $|\,\text{TOP}_t\,| \geqslant m$. Second, if $i \in I \subset$ INFEAS, the assumption implies that $\text{TestF}(i, t) \neq$ True, so that $i \in G_t$ for all $t \in \mathbb{N}$.

- Suppose $\text{TOP}_t^c \cap E_t = \emptyset$. Then, $|E_t^c| \geqslant n - m$, which implies that

  $$m \leqslant |\,\text{FEAS}\,| \leqslant |E_t| \leqslant m.$$

  Since $E_t = \text{TOP}_t$ by definition of TOP_t, FEAS $\subset E_t = \text{TOP}_t$, so that $\text{TOP}_t =$ FEAS $=$ OPT. Either $\text{TOP}_t \subset F_t$ or $\text{TOP}_t \not\subset F_t$. If $\text{TOP}_t \subset F_t$, then

  $$E_t = \underset{t}{\text{TOP}} \subset F_t \subset E_t,$$

  so that $\text{TOP}_t = F_t = E_t$. Thus, that t is the last round, i.e., $t = \tau$, which is a contradiction. If $\text{TOP}_t \not\subset F_t$, then $h_t \in G_t$ by line 13 of the algorithm, so that $\text{NEEDY}_{h_t}^t = 1$. For the remainder of the proof, we will assume $\text{TOP}_t^c \cap E_t \neq \emptyset$.

- Suppose $h_t \in \text{BELOW}_t$ and $l_t \in \text{ABOVE}_t$. Then,

  $$\begin{aligned} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{h_t, N_{h_t}(t)} &\leqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{h_t, N_{h_t}(t)} + U_{\boldsymbol{r}}(N_{h_t}(t), \delta) && (51) \\ &< d && (52) \\ &< \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{l_t, N_{l_t}(t)} - U_{\boldsymbol{r}}(N_{l_t}(t), \delta) && (53) \\ &\leqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{l_t, N_{l_t}(t)} && (54) \end{aligned}$$

  where line (52) follows since $h_t \in \text{BELOW}_t$ and line (53) follows since $l_t \in \text{ABOVE}_t$. Thus, $\boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{h_t, N_{h_t}(t)} < \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{l_t, N_{l_t}(t)}$. However, $h_t \in \text{TOP}_t$ and $l_t \in \text{TOP}_t \cap E_t$ imply $\boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{h_t, N_{h_t}(t)} \geqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{l_t, N_{l_t}(t)}$ and thus we have a contradiction.

- Suppose that $h_t \in \text{BELOW}_t$ and $l_t \in \text{BELOW}_t$; we will derive a contradiction. We claim that $\text{OPT} \cap \text{TOP}_t^c = \varnothing$. Suppose that there exists $i \in \text{OPT} \cap \text{TOP}_t^c$. Since $\text{TestF}(i,t) \neq \text{False}$ for all $i \in \text{OPT}$, $i \in E_t$. Then,

$$\boldsymbol{r}^\top \boldsymbol{\mu}_i \leqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(t)} + U_{\boldsymbol{r}}(N_i(t),\delta) \tag{55}$$
$$\leqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{l_t,N_{l_t}(t)} + U_{\boldsymbol{r}}(N_{l_t}(t),\delta) \tag{56}$$
$$< d, \tag{57}$$

  where (55) follows by (28), (56) follows by $i \in E_t$, and (57) follows by $l_t \in \text{BELOW}_t$. $\boldsymbol{r}^\top \boldsymbol{\mu}_i < d$ is a contradiction, so that $\text{OPT} \cap \text{TOP}_t^c = \varnothing$. Thus, for all $i \in \text{TOP}_t^c$, either $\boldsymbol{r}^\top \boldsymbol{\mu}_i < d$ or $\boldsymbol{\mu}_i \notin P$. Furthermore, observe that

$$\boldsymbol{r}^\top \boldsymbol{\mu}_{h_t} \leqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{h_t,N_{h_t}(t)} + U_{\boldsymbol{r}}(N_{h_t}(t),\delta) \tag{58}$$
$$\leqslant d \tag{59}$$

  where line (58) follows by (28) and line (59) follows by $h_t \in \text{BELOW}_t$. Thus, there are at least $K - m + 1$ arms that are either suboptimal or infeasible. But, this is a contradiction since by assumption $|\text{FEAS}| \geqslant m$, there are exactly $K - m$ arms that are suboptimal or infeasible.

- Suppose $h_t \in \text{ABOVE}_t$ and $\text{TOP}_t \not\subset F_t$. Since $\text{TOP}_t \not\subset F_t$, $h_t \in G_t$ so that if $h_t \in \text{OPT} \cup I$, then $\text{NEEDY}_{h_t}^t = 1$. So, suppose that $h_t \in S$. If $h_t \in S$, then

$$\boldsymbol{r}^\top \boldsymbol{\mu}_{h_t} \geqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{h_t,N_{h_t}(t)} - U_{\boldsymbol{r}}(N_{h_t}(t),\delta) > d$$

  where the first inequality follows by (28) and the second inequality follows by $h_t \in \text{ABOVE}_t$. But, $\boldsymbol{r}^\top \boldsymbol{\mu}_{h_t} > d$ is a contradiction since $h_t \in S$.

- Suppose $h_t \in \text{ABOVE}_t$, $\text{TOP}_t \subset F_t$, and $l_t \in \text{BELOW}_t$. Then, $\text{TOP}_t \subset F_t$, $h_t \in \text{ABOVE}_t$, and $l_t \in \text{BELOW}_t$ imply that the termination condition is satisfied so that $t = \tau$, which is a contradiction.

- Suppose $h_t \in \text{ABOVE}_t$, $\text{TOP}_t \subset F_t$, and $l_t \in \text{ABOVE}_t$. First, we claim that $\text{TOP}_t \subset \text{OPT}$. Let $i \in \text{TOP}_t$. Then, $i \in F_t$, which implies that $\text{TestF}(i,t) = \text{True}$, so that $i \notin I$. Further, $h_t \in \text{ABOVE}_t$ implies that

$$\boldsymbol{r}^\top \boldsymbol{\mu}_i \geqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(t)} - U_{\boldsymbol{r}}(N_i(t),\delta) > d,$$

  where the first inequality follows by (28) and the second inequality follows by $h_t \in \text{ABOVE}_t$. Therefore, $i \notin S$. Thus, $i \in \text{OPT}$, proving that $\text{TOP}_t \subset \text{OPT}$.

  There are three cases: either $l_t \in \text{OPT}$, $l_t \in S$, or $l_t \in I$. $l_t \in \text{OPT}$ implies that there are $m+1$ optimal feasible arms since $|\text{TOP}_t| \geqslant m$ as established earlier and $\text{OPT} \supset \text{TOP}_t$, which is a contradiction. Since $l_t \in \text{ABOVE}_t$, we have by (28),

$$\boldsymbol{r}^\top \boldsymbol{\mu}_{l_t} \geqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{l_t,N_{l_t}(t)} - U_{\boldsymbol{r}}(N_{l_t}(t),\delta) > d,$$

  which implies that $l_t \notin S$. Thus, $l_t \in I$. Since $l_t \in G_t$ as established earlier, we have that $\text{NEEDY}_{l_t}^t = 1$.

- If $l_t \in \text{MIDDLE}_t$, then $l_t \notin \text{ABOVE}_t \cup \text{BELOW}_t$ so if $l_t \in \text{OPT} \cup S$, then $\text{NEEDY}_{l_t}^t = 1$. Further, if $l_t \in I$, then as argued previously $l_t \in G_t$, so that $\text{NEEDY}_{l_t}^t = 1$.

- If $h_t \in \text{MIDDLE}_t$, the argument is identical to the previous case.

□

Proof of Lemma C.2. First, we observe that $\widehat{\text{O}} = \text{TOP}_\tau$, $\widehat{\text{S}} = (\text{TOP}_\tau \cup E_\tau^c)^c$, and $\widehat{\text{I}} = E_\tau^c$. Note that $\widehat{\text{S}} \cap \widehat{\text{I}} = \emptyset$ by definition of the algorithm.

Step 1: $\text{TOP}_\tau = \text{OPT}$.

To begin, we make two useful observations. *(i)* We claim that $\text{TOP}_\tau \subset \text{FEAS}$. Let $i \in \text{TOP}_\tau$. Then, since at termination, $\text{TOP}_\tau \subset F_\tau$, we have that $\text{TestF}(i, \tau) = \text{True}$. Then, by the hypothesis, $\boldsymbol{\mu}_i \in P$, so that $i \in \text{FEAS}$. *(ii)* We claim that $\text{OPT} \subset E_\tau$. Let $i \in \text{OPT}$. Since by assumption $\text{TestF}(i, \tau) \neq \text{False}$ for all i such that $\boldsymbol{\mu}_i \in P$, it follows that $i \in E_\tau$, establishing $\text{OPT} \subset E_\tau$.

Case 1: $|\,\textbf{FEAS}\,| < m$. Notice that since there are fewer than m feasible arms, $\text{SUBOPT} = \emptyset$ and we have that $\text{OPT} = \text{FEAS}$.

By our observation *(i)*, $\text{TOP}_\tau \subset \text{FEAS} = \text{OPT}$.

Next, we show that $\text{OPT} \subset \text{TOP}_\tau$. Let $i \in \text{OPT}$. Then, by observation *(ii)* $i \in E_\tau$. Since $\text{TOP}_\tau \subset \text{OPT}$ and $|\,\text{OPT}\,| < m$, $|\,\text{TOP}_\tau\,| < m$. Since $|\,\text{TOP}_\tau\,| < m$, the definition of TOP_τ in line 6 of the algorithm implies that $|\,\text{TOP}_\tau\,| = |E_\tau|$ and $\text{TOP}_\tau \subset E_\tau$. Therefore, $\text{TOP}_\tau = E_\tau$ so $i \in \text{TOP}_\tau$, which establishes the claim.

Case 2: $|\,\textbf{FEAS}\,| \geqslant m$. By observation *(i)*, $\text{TOP}_\tau \subset \text{FEAS}$, which implies that $\text{TOP}_\tau \cap \text{INFEAS} = \emptyset$. Next, we show that $\text{TOP}_\tau \cap \text{SUBOPT} = \emptyset$. Towards a contradiction, suppose that there exists $i \in \text{TOP}_\tau \cap \text{SUBOPT}$. Then, since $|\,\text{OPT}\,| = m$ and $|\,\text{TOP}_\tau\,| = m$ by *(ii)*, there exists $j \in \text{OPT} \cap \text{TOP}_\tau^c$. Since $\text{OPT} \subset E_\tau$ by observation *(ii)*, $j \in E_\tau$. Then, by line 6 defining TOP_τ, $|E_\tau| > m$, so the algorithm must terminate with the stopping condition: $\text{TOP}_t \subset F_t$ and $\min_{i \in \text{TOP}_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(t)} - U_{\boldsymbol{r}}(N_i(t), \delta) \geqslant \max_{j \in \text{TOP}_t^c \cap E_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{j,N_j(t)} + U_{\boldsymbol{r}}(N_j(t), \delta)$. By the stopping condition, we have that

$$
\begin{aligned}
\boldsymbol{r}^\top \boldsymbol{\mu}_i &\geqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(\tau)} - U_{\boldsymbol{r}}(N_i(\tau), \delta) && (60)\\
&\geqslant \min_{l \in \text{TOP}_\tau} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{l,N_k(\tau)} - U_{\boldsymbol{r}}(N_l(\tau), \delta) && (61)\\
&\geqslant \max_{k \in \text{TOP}_\tau^c \cap E_\tau} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{k,N_k(\tau)} + U_{\boldsymbol{r}}(N_k(\tau), \delta) && (62)\\
&\geqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{j,N_j(\tau)} + U_{\boldsymbol{r}}(N_j(\tau), \delta) && (63)\\
&\geqslant \boldsymbol{r}^\top \boldsymbol{\mu}_j && (64)
\end{aligned}
$$

where lines (60) and (64) follow by (29) and (62) follows by the stopping condition. Thus, $\boldsymbol{r}^\top \boldsymbol{\mu}_i \geqslant \boldsymbol{r}^\top \boldsymbol{\mu}_j$, which is contradicts the assumption $(\nu, P, \boldsymbol{r}, m) \in \mathcal{M}$. Therefore, the claim $\text{TOP}_\tau \cap \text{SUBOPT} = \emptyset$ follows.

Note that $\text{TOP}_\tau \cap \text{INFEAS} = \emptyset$ and $\text{TOP}_\tau \cap \text{SUBOPT} = \emptyset$ imply that $\text{TOP}_\tau \subset \text{OPT}$. Since $\text{OPT} \subset E_\tau$ and $|\,\text{FEAS}\,| \geqslant m$, $|\,\text{TOP}_\tau\,| = m$. Thus, it follows that $\text{TOP}_\tau = \text{OPT}$ and correctness follows.

Step 2: $\widehat{\textbf{S}} \subset$ SUBOPT and $\widehat{\textbf{I}} \subset$ INFEAS. First, we show that $\widehat{\text{S}} \subset \text{SUBOPT}$. If $\widehat{\text{S}} = \emptyset$, there is nothing to show so suppose that $\widehat{\text{S}} \neq \emptyset$. Let $i \in \widehat{\text{S}}$. Since $i \in \widehat{\text{S}} = \text{TOP}_\tau^c \cap E_\tau$, we cannot have that $\text{TOP}_\tau = F_\tau$ and $F_\tau = E_\tau$. So, the algorithm terminates with the stopping condition: $\text{TOP}_t \subset F_t$ and $\min_{i \in \text{TOP}_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(t)} - U_{\boldsymbol{r}}(N_i(t), \delta) \geqslant \max_{j \in \text{TOP}_t^c \cap E_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{j,N_j(t)} + U_{\boldsymbol{r}}(N_j(t), \delta)$. Then, using the stopping condition,

$$
\begin{aligned}
\boldsymbol{r}^\top \boldsymbol{\mu}_i &\leqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(\tau)} + U_{\boldsymbol{r}}(N_j(\tau), \delta) && (65)\\
&\leqslant \min_{k \in \text{TOP}_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{k,N_k(t)} - U_{\boldsymbol{r}}(N_k(t), \delta) && (66)\\
&\leqslant \min_{k \in \text{OPT}} \boldsymbol{r}^\top \boldsymbol{\mu}_k && (67)
\end{aligned}
$$

where lines (65) and (67) follow by (29) and line (66) follows by the stopping condition. Thus, $i \in$ SUBOPT by the assumption $(\nu, P, \boldsymbol{r}, m) \in \mathcal{M}$.

Next, we show that $\widehat{\mathrm{I}} \subset$ INFEAS. Let $i \in \widehat{\mathrm{I}}$. Then, TestF$(i, N_\tau(i)) =$ False. By hypothesis, this implies that $\boldsymbol{\mu}_i \notin P$, so $i \in$ INFEAS.

□

D Upper Bounds for Three Instances of TF-LUCB

In the following three sections, we prove Theorem 4. We prove a separate theorem for each statement in Theorem 4: namely, Theorem D.1, Theorem D.2, and Theorem D.3. Each proof has a similar structure: *(i)* define a good event that holds whp, *(ii)* show that on this event, the TestF subroutine in question does not return the wrong answer, and *(iii)* show that after enough samples have been taken from the distribution, the TestF subroutine in question determines whether the mean of the distribution belongs to the set.

We introduce the following definition.

Definition D.1. *Let $Z \subset \mathbb{R}^D$ and $\epsilon > 0$. $\mathcal{N} \subset Z$ is an ϵ-net of Z if for all $\boldsymbol{x} \in Z$, there exists $\boldsymbol{y} \in \mathcal{N}$ such that $\|\boldsymbol{x} - \boldsymbol{y}\|_2 \leqslant \epsilon$. Let $\mathcal{N} \subset Z$ be an ϵ-net of Z. We say that $\mathcal{N}$ is* minimal *if, for any other ϵ-net $\mathcal{O}$ of Z, it holds that $|\mathcal{O}| \geqslant |\mathcal{N}|$.*

D.1 Proof of Upper Bound for TF-LUCB-B

Theorem D.1. *Let $\delta > 0$ and $(\nu, P, \boldsymbol{r}, m) \in \mathcal{M}$. With probability at least $1 - \delta$, TF-LUCB-B returns $(\widehat{O}, \widehat{S}, \widehat{I})$ such that $\widehat{O} = OPT$, $\widehat{S} \subset SUBOPT$, $\widehat{I} \subset INFEAS$, and*

$$\tau \leqslant \min_{(S,I) \in \textit{Valid-Partitions}} c\sigma^2 \Big[\sum_{i \in S} F(\min_{j \in OPT} \boldsymbol{r}^\top(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i), \frac{K}{\delta}) + \sum_{i \in I} DF(\operatorname{dist}(\boldsymbol{\mu}_i, \partial P), \frac{K}{\delta}) \quad (68)$$

$$+ \sum_{i \in OPT} \max(F(\min_{j \in S} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j), \frac{K}{\delta}), DF(\operatorname{dist}(\boldsymbol{\mu}_i, \partial P), \frac{K}{\delta})) \Big]. \quad (69)$$

where c is a universal positive constant.

Proof. By Theorem 3, it suffices to show that for any $(\xi, R) \in \mathcal{N}$ where ξ is σ-sub-Gaussian and has mean $\boldsymbol{\mu} \in \mathbb{R}^D$, with probability at least $1 - \frac{\delta}{2K}$, TestF-B returns True only if $\boldsymbol{\mu} \in R$ and returns False only if $\boldsymbol{\mu} \notin R$ and after at most

$$c\sigma^2 D \operatorname{dist}(\boldsymbol{\mu}, \partial R)^{-2} \log(\log(\operatorname{dist}(\boldsymbol{\mu}, \partial R)^{-2}) \frac{K}{\delta})$$

pulls for some universal positive constant c, it returns either True or False.

Step 1: Define the event. Let $\widehat{\boldsymbol{\mu}}_t$ denote the empirical mean of ξ after t samples. Define the event $B = \{\forall t \in \mathbb{N} : \|\widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}\|_2 \leqslant U_{\text{ball}}(t, \delta)\}$. Let $\mathcal{N}$ be a minimal $\frac{1}{2}$-net of $\mathcal{S}^{D-1}$.

Observe that since for any $\boldsymbol{y} \in \mathcal{N}$, $\|\boldsymbol{y}\|_2 = 1$ and ν_i is σ-sub-Gaussian, if $\boldsymbol{X} \sim \nu_i$, then

$$\left\|\boldsymbol{y}^\top \boldsymbol{X}\right\|_{\psi_2} \leqslant \|\boldsymbol{X}\|_{\psi_2} \leqslant \sigma$$

so that $\boldsymbol{y}^\top \boldsymbol{X}$ is σ-sub-Gaussian.

Then,

$$\Pr(B^c) = \Pr(\exists t \in \mathbb{N} : \|\widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}\|_2 > U_{\text{ball}}(t, \delta)) \tag{70}$$

$$= \Pr(\exists t \in \mathbb{N}, \exists \boldsymbol{y} \in \mathcal{N} : |\boldsymbol{y}^\top(\widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu})| > \frac{1}{2} U_{\text{ball}}(t, d)) \tag{71}$$

$$\leqslant \sum_{\boldsymbol{y} \in \mathcal{N}} \Pr(\exists t \in \mathbb{N} : |\boldsymbol{y}^\top(\widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu})| > \frac{1}{2} U_{\text{ball}}(t, d)) \tag{72}$$

$$\leqslant 5^D \frac{\delta}{5^D 2K} \tag{73}$$

$$\leqslant \frac{\delta}{2K}, \tag{74}$$

where line (71) follows by Lemma F.4 and line (73) follows by Lemma F.10 and since Lemma F.5 implies that $|\mathcal{N}| \leqslant 5^D$. So, $\Pr(B) \geqslant 1 - \frac{\delta}{2K}$. For the remainder of the proof, we suppose that B occurs.

Step 2: An incorrect answer is never returned. First, we consider the case $\boldsymbol{\mu} \in R$. First, we show that TestF-B returns only either True or ?. Towards a contradiction, suppose that TestF-B(t) = False. Then, since $\boldsymbol{\mu} \in R$ and event B,

$$U_{\text{ball}}(t, \delta) < \text{ dist}(\widehat{\boldsymbol{\mu}}_t, R) \leqslant \|\widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}\|_2 \leqslant U_{\text{ball}}(t, \delta),$$

which is a contradiction. Thus, TestF-B returns either True or ?.

Next, consider the case $\boldsymbol{\mu} \in R^c$; the proof is very similar to the case $\boldsymbol{\mu} \in R$. Towards a contradiction, suppose that TestF-B(t) = True. Then, since $\boldsymbol{\mu} \in R^c$ and event B,

$$U_{\text{ball}}(t, \delta) < \text{ dist}(\widehat{\boldsymbol{\mu}}_t, R^c) \leqslant \|\widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}\|_2 \leqslant U_{\text{ball}}(t, \delta),$$

which is a contradiction. Thus, TestF-B returns either False or ?.

Step 3: Bound the sample complexity. Next, we show that TestF-B(t) = returns either True or False for all

$$t \geqslant c\sigma^2 D \text{ dist}(\boldsymbol{\mu}, \partial R)^{-2} \log(\log(\frac{\text{dist}(\boldsymbol{\mu}, \partial R)^{-2})K}{\delta})$$

where c is a universal positive constant. Let ρ denote the smallest integer such that

$$U_{\text{ball}}(\rho, \delta) < \frac{\text{dist}(\boldsymbol{\mu}, \partial R)}{2}.$$

By Lemma F.11, $\rho \leqslant c\sigma^2 D \text{ dist}(\boldsymbol{\mu}, \partial R)^{-2} \log(\frac{\log(\text{ dist}(\boldsymbol{\mu}, \partial R)^{-2})2K}{\delta})$ for some universal positive constant c. Let $t \geqslant \rho$. Towards a contradiction, suppose that TestF-B(i, t) = ?. Then, $\text{dist}(\widehat{\boldsymbol{\mu}}_t, R) \leqslant U_{\text{ball}}(t, \delta)$ and $\text{dist}(\widehat{\boldsymbol{\mu}}_t, R^c) \leqslant U_{\text{ball}}(t, \delta)$ so that by Lemma F.9, there exists $\boldsymbol{x} \in \partial R$ such that $\|\widehat{\boldsymbol{\mu}}_t - \boldsymbol{x}\|_2 \leqslant U(t, \delta)$. Then, by the triangle inequality and event B,

$$\begin{aligned}
\|\boldsymbol{\mu} - \boldsymbol{x}\|_2 &\leqslant \|\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t\|_2 + \|\widehat{\boldsymbol{\mu}}_t - \boldsymbol{x}\|_2 \\
&\leqslant U_{\text{ball}}(t, \delta) + U_{\text{ball}}(t, \delta) \\
&< \text{ dist}(\boldsymbol{\mu}, \partial R) \\
&\leqslant \|\boldsymbol{\mu} - \boldsymbol{x}\|_2,
\end{aligned}$$

which is a contradiction. Thus, for all $t \geqslant \rho$, TestF-B(t) returns True or False. The result follows.

□

D.2 Proof of Upper Bound for TF-LUCB-CB

Define

$$\mathcal{N}_{poly} = \{(\xi, R) \in \mathcal{N} : \ R \text{ is a polyhedron}\}.$$

Theorem D.2. *Let $\delta > 0$, $P = \{\boldsymbol{x} \in \mathbb{R}^D : A\boldsymbol{x} \leqslant \boldsymbol{b}\}$, and $(\nu, P, \boldsymbol{r}, m) \in \mathcal{M}$. With probability at least $1-\delta$, TF-LUCB-CB returns $(\widehat{O}, \widehat{S}, \widehat{I})$ such that $\widehat{O} = OPT$, $\widehat{S} \subset SUBOPT$, $\widehat{I} \subset INFEAS$, and*

$$\tau \leqslant \min_{(S,I) \in \textit{Valid-Partitions}} c\sigma^2 \Big[\sum_{i \in S} F(\min_{j \in OPT} \boldsymbol{r}^\top (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i), \frac{K}{\delta}) + \sum_{i \in I} DF(\text{ dist}(\boldsymbol{\mu}_i, \partial P), \frac{K}{\delta}) \tag{75}$$

$$+ \sum_{i \in OPT} \max(F(\min_{j \in S} \boldsymbol{r}^\top (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j), \frac{K}{\delta}), F(\text{ dist}(\boldsymbol{\mu}_i, \partial P), \frac{KM}{\delta})) \Big]. \tag{76}$$

where c is a universal positive constant.

Proof. By Theorem 3, it suffices to show that for any $(\xi, R) \in \mathcal{N}_{poly}$ where ξ is σ-sub-Gaussian and has mean $\boldsymbol{\mu} \in \mathbb{R}^D$, with probability at least $1 - \frac{\delta}{2K}$, if $\boldsymbol{\mu} \in R$, then TestF-CB only returns either ? or True and for all

$$t \geqslant c\sigma^2 \text{ dist}(\boldsymbol{\mu}, \partial R)^{-2} \log(\log(\text{ dist}(\boldsymbol{\mu}, \partial R)^{-2}) \frac{K}{\delta})$$

where c is a universal positive constant, TestF-CB(t) returns True, and if $\boldsymbol{\mu} \notin R$, then TestF-CB only returns either ? or False and for all

$$t \geqslant c\sigma^2 D \text{ dist}(\boldsymbol{\mu}, \partial R)^{-2} \log(\log(\text{ dist}(\boldsymbol{\mu}, \partial R)^{-2}) \frac{KM}{\delta})$$

where c is a universal positive constant, TestF-CB(t) returns False.

Step 1: Define the event. For the sake of brevity, let $U_{\text{ball}}(t) := U_{\text{ball}}(t, \frac{\delta}{2})$ and $U_{\text{con}}(t) := U_{\text{con}}(t, \frac{\delta}{2})$. Let $\widehat{\boldsymbol{\mu}}_t$ denote the empirical mean of ξ after t samples. Define the event

$$\begin{aligned} B =& \{\forall t \in \mathbb{N} : \|\widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}\|_2 \leqslant U_{\text{ball}}(t)\} \\ &\cap \{\forall t \in \mathbb{N}, \forall s \in [M] : |\boldsymbol{a}_s^\top \widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}| \leqslant U_{\text{con}}(t)\}. \end{aligned}$$

Let $\mathcal{N}$ be a minimal $\frac{1}{2}$-net of $\mathcal{S}^{D-1}$. Observe that since for any $\boldsymbol{y} \in \mathcal{N}$, $\|\boldsymbol{y}\|_2 = 1$, for any $j \in [M]$, $\|\boldsymbol{a}_j\|_2 = 1$ and ν_i is σ-sub-Gaussian, if $\boldsymbol{X} \sim \nu_i$, then for $\boldsymbol{z} \in \mathcal{N} \cup \{\boldsymbol{a}_j : j \in [M]\}$

$$\|\boldsymbol{z}^\top \boldsymbol{X}\|_{\psi_2} \leqslant \|\boldsymbol{X}\|_{\psi_2} \leqslant \sigma$$

so that $\boldsymbol{z}^\top \boldsymbol{X}$ is σ-sub-Gaussian.

By the union bound, Lemma F.10, and a similar argument as in (74),

$$\Pr(B^c) \leqslant \frac{\delta}{4K} + \frac{\delta}{4K} = \frac{\delta}{2K}.$$

For the remainder of the proof, suppose that B occurs.

Step 2: $\boldsymbol{\mu} \in R$. Suppose $\boldsymbol{\mu} \in R$. First, we show that TestF-CB returns only either True or ?. Towards a contradiction, suppose that TestF-CB(t) = False. Then, since $\boldsymbol{\mu} \in R$ and event B,

$$U_{\text{ball}}(t) < \text{ dist}(\widehat{\boldsymbol{\mu}}_t, R) \leqslant \|\widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}\|_2 \leqslant U_{\text{ball}}(t),$$

which is a contradiction. Thus, TestF-CB returns either True or ?.

Next, we show that TestF-CB(t) = True for all

$$t \geqslant c\sigma^2 \ \text{dist}(\boldsymbol{\mu}, \partial R)^{-2} \log(\log(\ \text{dist}(\boldsymbol{\mu}, \partial R)^{-2})\frac{K}{\delta})$$

where c is a universal positive constant. Let ρ denote the smallest integer such that

$$U_{\text{con}}(\rho) < \frac{\text{dist}(\boldsymbol{\mu}, \partial R)}{2} = \frac{\min_{s \in [M]} b_s - \boldsymbol{a}_s^\top \boldsymbol{\mu}}{2}.$$

where the equality follows by Lemma F.3. By Lemma F.11,

$$\rho \leqslant c\sigma^2 \ \text{dist}(\boldsymbol{\mu}, \partial R)^{-2} \log(\frac{\log(\ \text{dist}(\boldsymbol{\mu}, \partial R)^{-2})KM}{\delta})$$

for some universal positive constant c. Let $t \geqslant \rho$. Fix $r \in [M]$. Then, by event B,

$$\begin{aligned}
\boldsymbol{a}_r^\top \widehat{\boldsymbol{\mu}}_t + U_{\text{con}}(t, \delta) &\leqslant \boldsymbol{a}_r^\top \boldsymbol{\mu} + 2\, U_{\text{con}}(t) \\
&\leqslant \boldsymbol{a}_r^\top \boldsymbol{\mu} + b_r - \boldsymbol{a}_r^\top \boldsymbol{\mu} \\
&= b_r.
\end{aligned}$$

Thus, TestF-CB(t) = True.

Step 3: $\boldsymbol{\mu} \in R^c$. Suppose $\boldsymbol{\mu} \in R^c$. Towards a contradiction, suppose that TestF-CB(t) returns True. Then, for all $s \in [M]$, $\boldsymbol{a}_s^\top \widehat{\boldsymbol{\mu}}_t + U_{\text{con}}(t) \leqslant b_s$. Then, by the event B,

$$b_s \geqslant \boldsymbol{a}_s^\top \widehat{\boldsymbol{\mu}}_t + U_{\text{con}}(t) \geqslant \boldsymbol{a}_s^\top \boldsymbol{\mu}$$

which contradicts the assumption that $\boldsymbol{\mu} \notin R$. Thus, TestF-CB$(t)$ only returns ? or False.

Next, we show that TestF-CB(t) returns False for all

$$t \geqslant cD \ \text{dist}(\boldsymbol{\mu}, \partial R)^{-2} \log(\log(\frac{\text{dist}(\boldsymbol{\mu}, \partial R)^{-2})K}{\delta})$$

where c is a universal positive constant. Let ρ denote the smallest integer such that

$$U_{\text{ball}}(\rho) < \frac{\text{dist}(\boldsymbol{\mu}, \partial R)}{2}.$$

By Lemma F.11, $\rho \leqslant c\sigma^2 D \ \text{dist}(\boldsymbol{\mu}, \partial R)^{-2} \log(\frac{\log(\ \text{dist}(\boldsymbol{\mu}, \partial R)^{-2})2K}{\delta})$ for some universal positive constant c. Let $t \geqslant \rho$. Towards a contradiction, suppose that TestF-B(t) = ?. Then, $\text{dist}(\widehat{\boldsymbol{\mu}}_t, R) \leqslant U_{\text{ball}}(t)$ and $\text{dist}(\widehat{\boldsymbol{\mu}}_t, R^c) \leqslant U_{\text{ball}}(t)$, so there exists $\boldsymbol{x} \in \partial R$ such that $\|\widehat{\boldsymbol{\mu}}_t - \boldsymbol{x}\|_2 \leqslant U_{\text{ball}}(t)$. Then, by the triangle inequality and event B,

$$\begin{aligned}
\|\boldsymbol{\mu} - \boldsymbol{x}\|_2 &\leqslant \|\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}_t\|_2 + \|\widehat{\boldsymbol{\mu}}_t - \boldsymbol{x}\|_2 \\
&\leqslant U_{\text{ball}}(t) + U_{\text{ball}}(t) \\
&< \ \text{dist}(\boldsymbol{\mu}, \partial R) \\
&\leqslant \|\boldsymbol{\mu} - \boldsymbol{x}\|_2\,,
\end{aligned}$$

which is a contradiction. Thus, for all $t \geqslant \rho$, TestF-CB(t) returns False. The result follows.

□

D.3 Proof of Upper Bound for TF-LUCB-C

First, we prove a more general version of Theorem D.4 that allows for any polyhedron.

Theorem D.3. *Let $\delta > 0$, $P = \{\boldsymbol{x} \in \mathbb{R}^D : A\boldsymbol{x} \leqslant \boldsymbol{b}\}$, and $(\nu, P, \boldsymbol{r}, m) \in \mathcal{M}$. For all $i \in [K]$ such that $\boldsymbol{\mu}_i \notin P$, let $\tilde{\Delta}_i = \max_{s \in [M]} \boldsymbol{a}_s^\top \boldsymbol{\mu}_i - b_s$. With probability at least $1-\delta$, TF-LUCB-C returns TOP_τ such that $TOP_\tau = OPT$ and*

$$\tau \leqslant \min_{(S,I) \in \textit{Valid-Partitions}} c\sigma^2 \Big[\sum_{i \in S} F(\min\nolimits_{j \in OPT} \boldsymbol{r}^\top (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i), \frac{K}{\delta}) + \sum_{i \in I} F(\tilde{\Delta}_i, \frac{KM}{\delta}) \tag{77}$$

$$+ \sum_{i \in OPT} \max(F(\min_{j \in S} \boldsymbol{r}^\top (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j), \frac{KM}{\delta})), F(\operatorname{dist}(\boldsymbol{\mu}_i, \partial P), \frac{KM}{\delta}))\Big]. \tag{78}$$

where c is a universal positive constant.

Proof. By Theorem 3, it suffices to show that for any $(\xi, R) \in \mathcal{N}_{poly}$ where ξ is σ-sub-Gaussian and has mean $\boldsymbol{\mu} \in \mathbb{R}^D$, with probability at least $1 - \frac{\delta}{2K}$, if $\boldsymbol{\mu} \in R$, then TestF-C only returns either ? or True and for all

$$t \geqslant c\sigma^2 \operatorname{dist}(\boldsymbol{\mu}, \partial R)^{-2} \log(\log(\operatorname{dist}(\boldsymbol{\mu}, \partial R)^{-2}) \frac{K}{\delta}$$

where c is a universal positive constant, TestF-C(t) returns True, and if $\boldsymbol{\mu} \notin R$, then TestF-C only returns either ? or False and for all

$$t \geqslant c\sigma^2 \tilde{\Delta}^{-2} \log(\log(\tilde{\Delta}_i \frac{KM}{\delta})$$

where $\tilde{\Delta} = \max_{s \in [M]} \boldsymbol{a}_s^\top \boldsymbol{\mu} - b_s$ and c is a universal positive constant, TestF-C(t) returns False.

Step 1: Define the event. Let $\widehat{\boldsymbol{\mu}}_t$ denote the empirical mean of ξ after t samples. Define the event $B = \{\forall t \in \mathbb{N}, \forall s \in [M] : |\boldsymbol{a}_s^\top (\widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu})| \leqslant U_{\text{con}}(t, \delta)\}$.

Observe that since for any $s \in [M]$, $\|\boldsymbol{a}_s\|_2 = 1$ and ν_i is σ-sub-Gaussian, if $\boldsymbol{X} \sim \nu_i$, then

$$\|\boldsymbol{a}_s^\top \boldsymbol{X}\|_{\psi_2} \leqslant \|\boldsymbol{X}\|_{\psi_2} \leqslant \sigma$$

so that $\boldsymbol{a}_s^\top \boldsymbol{X}$ is σ-sub-Gaussian.

Then, by Lemma F.10,

$$\begin{aligned}
\Pr(B^c) &= \Pr(\exists t \in \mathbb{N}, \exists s \in [M] : |\boldsymbol{a}_s^\top (\widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu})| > U_{\text{con}}(t, \delta)) \\
&= M \Pr(\exists t \in \mathbb{N} : |\boldsymbol{a}_s^\top (\widehat{\boldsymbol{\mu}}_t - \boldsymbol{\mu})| > U_{\text{con}}(t, \delta)) \\
&\leqslant M \frac{\delta}{2KM} \\
&= \frac{\delta}{2K}.
\end{aligned}$$

So, $\Pr(B) \geqslant 1 - \frac{\delta}{2K}$. For the remainder of the proof, we suppose that B occurs.

Step 2: $\boldsymbol{\mu} \in R$. Suppose $\boldsymbol{\mu} \in R$. Towards a contradiction, suppose that TestF-C(t) = False. Then, there exists $s \in [M]$ such that $\boldsymbol{a}_s^\top \widehat{\boldsymbol{\mu}}_t - U_{\text{con}}(t, \delta) > b_s$. Then, by the event B,

$$b_s < \boldsymbol{a}_s^\top \widehat{\boldsymbol{\mu}}_t - U_{\text{con}}(t, \delta) \leqslant \boldsymbol{a}_s^\top \boldsymbol{\mu}$$

which contradicts the assumption that $\boldsymbol{\mu} \in R$. Thus, TestF-C only returns either ? or True.

Next, we show that TestF-C(t) = True for all

$$t \geqslant c\sigma^2 \ \operatorname{dist}(\boldsymbol{\mu}, \partial R)^{-2} \log(\log(\operatorname{dist}(\boldsymbol{\mu}, \partial R)^{-2})\frac{K}{\delta}$$

where c is a universal positive constant. Let ρ denote the smallest integer such that

$$U_{\text{con}}(\rho, \delta) < \frac{\operatorname{dist}(\boldsymbol{\mu}, \partial R)}{2} = \frac{\min_{s\in[M]} b_s - \boldsymbol{a}_s^\top \boldsymbol{\mu}}{2}.$$

where the equality follows by Lemma F.3. By Lemma F.11,

$$\rho \leqslant c\sigma^2 \ \operatorname{dist}(\boldsymbol{\mu}, \partial R)^{-2} \log(\frac{\log(\operatorname{dist}(\boldsymbol{\mu}, \partial R)^{-2})KM}{\delta})$$

for some universal positive constant c. Let $t \geqslant \rho$. Fix $r \in [M]$. Then, by the event B,

$$\begin{aligned}
\boldsymbol{a}_r^\top \widehat{\boldsymbol{\mu}}_t + U_{\text{con}}(t, \delta) &\leqslant \boldsymbol{a}_r^\top \boldsymbol{\mu} + 2\, U_{\text{con}}(t, \delta) \\
&\leqslant \boldsymbol{a}_r^\top \boldsymbol{\mu} + b_r - \boldsymbol{a}_r^\top \boldsymbol{\mu} \\
&= b_r.
\end{aligned}$$

Thus, TestF-C(t) = True.

Step 3: $\boldsymbol{\mu} \in R^c$. Next, suppose $\boldsymbol{\mu} \in R^c$. Let $s \in [M]$ such that $\tilde{\Delta} = \boldsymbol{a}_s^\top \boldsymbol{\mu} - b_s$. Towards a contradiction, suppose that TestF-C(t) returns True. Then, $\boldsymbol{a}_s^\top \widehat{\boldsymbol{\mu}}_t + U_{\text{con}}(t, \delta) \leqslant b_s$. Then, by the event B,

$$b_s \geqslant \boldsymbol{a}_s^\top \widehat{\boldsymbol{\mu}}_t + U_{\text{con}}(t, \delta) \geqslant \boldsymbol{a}_s^\top \boldsymbol{\mu}$$

which contradicts the assumption that $\boldsymbol{\mu} \notin R$ and our choice of $s \in [M]$. Thus, TestF-C(t) only returns ? or False.

Next, we show that TestF-C(t) = False for all $t \geqslant c\sigma^2 \tilde{\Delta}^{-2} \log(\frac{\log(\tilde{\Delta}^{-2})KM}{\delta})$ where c is a universal positive constant. Let ρ denote the smallest integer such that

$$U_{\text{con}}(\tau, \delta) < \frac{\tilde{\Delta}}{2}.$$

By Lemma F.11, $\rho \leqslant c\sigma^2 \tilde{\Delta}^{-2} \log(\frac{\log(\tilde{\Delta}^{-2})KM}{\delta})$ for some universal positive constant c. Let $t \geqslant \rho$. Then, by the event B

$$\begin{aligned}
\boldsymbol{a}_s^\top \widehat{\boldsymbol{\mu}}_t - U_{\text{con}}(t, \delta) &\geqslant \boldsymbol{a}_s^\top \boldsymbol{\mu} - 2\, U_{\text{con}}(t, \delta) \\
&\geqslant \boldsymbol{a}_s^\top \boldsymbol{\mu} - (\boldsymbol{a}_s^\top \boldsymbol{\mu} - b_s) \\
&= b_s.
\end{aligned}$$

Thus, TestF-C(t) = False.

□

In general, $\tilde{\Delta}_i$ can be arbitrarily smaller than $\operatorname{dist}(\boldsymbol{\mu}_i, P)$, as indicated by the following Proposition.

Proposition D.1. *For all $M > 0$ and for all $\epsilon > 0$, there exists a polyhedron $P = \{\boldsymbol{x} \in \mathbb{R}^D : A\boldsymbol{x} \leqslant b\}$ and $\boldsymbol{x}_0 \in \mathbb{R}^D$ such that $\operatorname{dist}(\boldsymbol{x}_0, P) \geqslant M$ and $\max_{i=1,\ldots,M} \operatorname{dist}(\boldsymbol{x}_0, \{\boldsymbol{x} \in \mathbb{R}^D : \boldsymbol{a}_i^\top \boldsymbol{x} \leqslant b_i\}) \leqslant \epsilon$.*

Proof of Proposition D.1. Consider the case $D = 2$. Fix $M > 0$ and $\epsilon > 0$. Consider

$$P_\alpha = \{\boldsymbol{x} \in \mathbb{R}^2 : \boldsymbol{e}_2^\top \boldsymbol{x} \geqslant 0, (\alpha \boldsymbol{e}_1 + (1-\alpha)\boldsymbol{e}_2)^\top \boldsymbol{x} \geqslant 0\}$$

where $\alpha \in (0,1)$. Let $\boldsymbol{x}_0 = -M\boldsymbol{e}_1$. Then, for sufficiently small $\alpha \in (0,1)$, we have that $\operatorname{dist}(\boldsymbol{x}_0, P) \geqslant M$ and $\operatorname{dist}(\boldsymbol{x}_0, \{\boldsymbol{x} \in \mathbb{R}^D : (\alpha \boldsymbol{e}_1 + (1-\alpha)\boldsymbol{e}_2)^\top \boldsymbol{x} \geqslant 0\}) \leqslant \epsilon$ □

However, the Theorem D.4 shows that it has good performance in the setting where $\boldsymbol{a}_i^\top \boldsymbol{a}_j = 0$ for all $i \neq j \in [K]$.

Theorem D.4. *Let $\delta > 0$, $P = \{\boldsymbol{x} \in \mathbb{R}^D : A\boldsymbol{x} \leqslant \boldsymbol{b}\}$ such that for any $l \neq k \in [K]$, $\boldsymbol{a}_l^\top \boldsymbol{a}_k = 0$, and $(\nu, P, \boldsymbol{r}, m) \in \mathcal{M}$. For each $i \in [K]$ such that $\boldsymbol{\mu}_i \notin P$, define $v_i = |\{j : \boldsymbol{a}_j^\top \boldsymbol{\mu}_i > b_j\}|$. With probability at least $1-\delta$, TF-LUCB-C returns $(\widehat{O}, \widehat{S}, \widehat{I})$ such that $\widehat{O} =$ OPT, $\widehat{S} \subset$ SUBOPT, $\widehat{I} \subset$ INFEAS, and $\tau \leqslant$*

$$\min_{(S,I)\in \textit{Valid-Partitions}} c\sigma^2 \Big[\sum_{i \in S} F(\min_{j \in OPT} \boldsymbol{r}^\top (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i), \frac{K}{\delta}) + \sum_{i \in I} v_i F(\operatorname{dist}(\boldsymbol{\mu}_i, P), \frac{KM}{\delta}) \tag{79}$$

$$+ \sum_{i \in OPT} \max(F(\min_{j \in S} \boldsymbol{r}^\top (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j), \frac{KM}{\delta}), F(\operatorname{dist}(\boldsymbol{\mu}_i, \partial P), \frac{KM}{\delta})\Big]. \tag{80}$$

Proof of Theorem D.4. Let $l \in [K]$ such that $\boldsymbol{\mu}_l \notin P$. Without loss of generality, by relabeling $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M$, let

$$[r] = \{j \in [K] : \boldsymbol{a}_j^\top \boldsymbol{\mu}_i > b_j\}.$$

Define

$$S_i = \{\boldsymbol{x} \in \mathbb{R}^D : \boldsymbol{a}_i^\top \boldsymbol{x} = b_i\}$$
$$S = \cap_{i=1,\ldots,r} S_i$$

We will show that

$$\operatorname{dist}(\boldsymbol{\mu}_l, P)^2 \leqslant \operatorname{dist}(\boldsymbol{\mu}_l, S)^2 \leqslant r \max_{i=1,\ldots,r} \operatorname{dist}(\boldsymbol{\mu}_l, S_i)^2;$$

Then, the result will following by plugging the above inequality into the upper bound (78) in the statement of Theorem D.3. By relabeling the subspaces, we may assume without loss of generality that

$$\max_{i=1,\ldots,r} \operatorname{dist}(\boldsymbol{\mu}_l, S_i) = \operatorname{dist}(\boldsymbol{\mu}_l, S_1).$$

Define

$$\boldsymbol{x}_0 = \boldsymbol{\mu}_l,$$
$$\boldsymbol{x}_1 = \text{Proj}_{S_1}(\boldsymbol{x}_0),$$
$$\boldsymbol{x}_{i+1} = \text{Proj}_{S_{i+1}}(\boldsymbol{x}_i).$$

We claim that for all $i \in [r]$, $\boldsymbol{x}_i = \boldsymbol{x}_0 + \sum_{j=1}^{i} (b_j - \boldsymbol{a}_j^\top \boldsymbol{x}_0)\boldsymbol{a}_j$. We prove this inductively. By the closed form solution of the distance from a point to a hyperplane and $\|\boldsymbol{a}_j\|_2 = 1$ for all $j \in [M]$ [1],

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + (b_1 - \boldsymbol{a}_1^\top \boldsymbol{x}_0)\boldsymbol{a}_1,$$

which shows that base case. Next, we show the inductive step; suppose $\boldsymbol{x}_i = \boldsymbol{x}_0 + \sum_{j=1}^{i}(b_j - \boldsymbol{a}_j^\top \boldsymbol{x}_0)\boldsymbol{a}_j$. Then,

$$\begin{aligned}\boldsymbol{x}_{i+1} &= \boldsymbol{x}_i + (b_{i+1} - \boldsymbol{a}_{i+1}^\top \boldsymbol{x}_i)\boldsymbol{a}_{i+1} \\ &= \boldsymbol{x}_0 + \sum_{j=1}^{i}(b_j - \boldsymbol{a}_j^\top \boldsymbol{x}_0)\boldsymbol{a}_j + (b_{i+1} - \boldsymbol{a}_{i+1}^\top[\boldsymbol{x}_0 + \sum_{j=1}^{i}(b_j - \boldsymbol{a}_j^\top \boldsymbol{x}_0)\boldsymbol{a}_j])\boldsymbol{a}_{i+1} \\ &= \boldsymbol{x}_0 + \sum_{j=1}^{i+1}(b_j - \boldsymbol{a}_j^\top \boldsymbol{x}_0)\boldsymbol{a}_j\end{aligned}$$

where we used the assumption that $\boldsymbol{a}_{i+1}^\top \boldsymbol{a}_j = 0$ for all $j \neq i+1$. Thus, the claim follows. Note that this implies that $\boldsymbol{x}_r \in S$.

Next, we note that for $i \neq j$,

$$(\boldsymbol{x}_i - \boldsymbol{x}_{i+1})^\top(\boldsymbol{x}_j - \boldsymbol{x}_{j+1}) = [-(b_{i+1} - \boldsymbol{a}_{i+1}^\top \boldsymbol{x}_0)\boldsymbol{a}_{i+1}]^\top[-(b_{j+1} - \boldsymbol{a}_{j+1}^\top \boldsymbol{x}_0)\boldsymbol{a}_{j+1}] = 0.$$

Then, by the pythagorean theorem,

$$\begin{aligned}\text{dist}(\boldsymbol{x}_0, S)^2 &\leqslant \|\boldsymbol{x}_0 - \boldsymbol{x}_r\|_2^2 \\ &= \|(\boldsymbol{x}_0 - \boldsymbol{x}_1) + (\boldsymbol{x}_1 - \boldsymbol{x}_2) + \ldots + (\boldsymbol{x}_{r-1} - \boldsymbol{x}_r)\|_2^2 \\ &= \sum_{i=1}^{r} \|\boldsymbol{x}_{i-1} - \boldsymbol{x}_r\|_2^2 \\ &\leqslant r \ \text{dist}(\boldsymbol{x}_0, S_1) \\ &= r \max_{i=1,\ldots,r} \ \text{dist}(\boldsymbol{x}_0, S_i).\end{aligned}$$

Next, we show that $\text{dist}(\boldsymbol{\mu}_l, P) \leqslant \text{dist}(\boldsymbol{\mu}_l, S)$. It suffices to show that $\boldsymbol{x}_r \in P$. For $s \in [r]$, $\boldsymbol{a}_s^\top \boldsymbol{x}_r = b_s$ by construction, so let $s \in [M]\backslash[r]$. Then, since $\boldsymbol{a}_s^\top \boldsymbol{a}_k = 0$ for all $k \in [r]$, it follows that

$$\boldsymbol{a}_s^\top \boldsymbol{x}_r = \boldsymbol{a}_s^\top \boldsymbol{x}_0 + \sum_{j=1}^{r}(b_j - \boldsymbol{a}_j^\top \boldsymbol{x}_0)\boldsymbol{a}_s^\top \boldsymbol{a}_i \leqslant b_s + 0.$$

Thus, it follows that $\text{dist}(\boldsymbol{\mu}_l, P) \leqslant \text{dist}(\boldsymbol{\mu}_l, S)$. □

E Alternative Lower Bound

To begin, we discuss our conjecture that there is a small gap between δ- PAC and δ- PAC-EXPLANATORY algorithms. Essentially a δ- PAC algorithm that is not δ- PAC-EXPLANATORY is allowed to rule out suboptimal feasible arms by incorrectly concluding that they are infeasible and to make the analogous mistake for infeasible arms with reward greater than $\max_{j \in \text{FEAS}}^{(m)} \boldsymbol{r}^\top \boldsymbol{\mu}_j$. We do not believe that this affords significant savings in sample complexity since δ- PAC algorithms typically use confidence bounds and to satisfy the δ- PAC criterion, these confidence bounds must be strong enough to determine that arms in OPT are feasible and have optimal rewards and to rule out every arm in OPT^c as either suboptimal or infeasible–all without prior knowledge of the number of infeasible or suboptimal arms. Nevertheless, we leave this as an open question.

Next, we discuss the differences between Theorems 1 and 2. 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 in fact the case. The main difference between the bounds occurs in the terms corresponding to $i \in$ OPT. The term $\min_{j \in \text{OPT}^c \cap \text{FEAS}} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$ in Theorem 2 is replaced with $\min_{j \in S} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$ where $S \supseteq \text{OPT}^c \cap \text{FEAS}$. 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 FEAS $\cap$ OPT^c. We conjecture that Theorem 2 is loose in this respect since intuitively if an algorithm rules out an arm by concluding that it is suboptimal, then regardless of whether the arm is feasible, the algorithm must determine that the arms in OPT have reward greater than it. To see the difference between theorems 1 and 2, consider the case where $K = 3$, $m = 1$, $\boldsymbol{r}^\top\boldsymbol{\mu}_1 > \boldsymbol{r}^\top\boldsymbol{\mu}_2 > \boldsymbol{r}^\top\boldsymbol{\mu}_3$, arms 1 and 3 are feasible and arm 2 is feasible. If arm 2 is very close to the boundary, then it may be much easier to show that arm 2 is suboptimal than to show that it is infeasible. In this case, the term reflecting the difficulty of showing that arm 1 is optimal will differ in the two theorems. Specifically, in this case, OPT$^c \cap$ FEAS $= \{3\}$ and $S = \{2, 3\}$, so

$$\min_{j \in \text{OPT}^c \cap \text{FEAS}} \boldsymbol{r}^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_j) = \boldsymbol{r}^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_3) > \boldsymbol{r}^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \min_{j \in S} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j).$$

Next, we prove Theorem 2. The proof has many similarities with the proof of Theorem 1. Recall the notation that for a given problem $(\nu, P, \boldsymbol{r}, m)$, we define

$$\begin{aligned}
&\text{FEAS}(\nu, P, \boldsymbol{r}, m) = \{i \in [K] : \boldsymbol{\mu}_i \in P\}, \ \text{INFEAS}(\nu, P, \boldsymbol{r}, m) = \text{FEAS}(\nu, P, \boldsymbol{r}, m)^c, \\
&\text{OPT}(\nu, P, \boldsymbol{r}, m) = \{i \in \text{FEAS}(\nu, P, \boldsymbol{r}, m) : \boldsymbol{r}^\top\boldsymbol{\mu}_i \geqslant \max\nolimits^{(m)}_{j \in \text{FEAS}(\nu, P, \boldsymbol{r}, m)} \boldsymbol{r}^\top\boldsymbol{\mu}_j\}, \\
&\text{SUBOPT}(\nu, P, \boldsymbol{r}, m) = \{i \in [K] : \boldsymbol{r}^\top\boldsymbol{\mu}_i < \max\nolimits^{(m)}_{j \in \text{FEAS}(\nu, P, \boldsymbol{r}, m)} \boldsymbol{r}^\top\boldsymbol{\mu}_j\}.
\end{aligned}$$

Proof of Theorem 2. Fix $\delta > 0$. Let $(\nu, P, \boldsymbol{r}, m)$ satisfy the hypotheses of the Theorem statement; note that these properties imply that $(\nu, P, \boldsymbol{r}, m) \in \mathcal{M}$. Let $\mathcal{A}$ denote a δ- PAC algorithm with stopping time τ.

In each of the next steps, we will define a new problem to obtain a lower bound. To avoid notational clutter, we will redefine the symbols $\boldsymbol{\mu}'_i$, ν'_i, and $\nu^{(i)}$ in each step. The context should make their meaning clear.

Step 1.a: reward bound for $i \in$ OPT. Fix $i \in$ OPT. First, we show that

$$\mathbb{E}_\nu[N_i(\tau)] \geqslant 2\ln(\frac{1}{2.4\delta})[\min_{j \in \text{FEAS} \cap \text{OPT}^c} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \epsilon]^{-2}$$

for a sufficiently small $\epsilon > 0$. If FEAS $\cap$ OPT$^c = \varnothing$, $\min_{j \in \text{FEAS} \cap \text{OPT}^c} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) = -\infty$ by definition and there is nothing to show. So, suppose that FEAS $\cap$ OPT$^c \neq \varnothing$. Define

$$j_0 = \arg\max_{j \in \text{FEAS} \cap \text{OPT}^c} \boldsymbol{r}^\top\boldsymbol{\mu}_j.$$

Define for all $j \in [K]$

$$\begin{aligned}
\boldsymbol{\mu}'_j &= \begin{cases} \begin{pmatrix} \mu_{i,1} - \mu_{j_0,1} - \epsilon \\ \boldsymbol{\mu}_{i,2:D} \end{pmatrix} & \text{if } j = i \\ \boldsymbol{\mu}_j & \text{if } j \neq i \end{cases} \\
\nu'_j &= N(\boldsymbol{\mu}'_j, I_D).
\end{aligned}$$

where $\epsilon > 0$ is chosen sufficiently small such that for any $\delta \in [0, \epsilon)$ $\boldsymbol{r}^\top \boldsymbol{\mu}_i' + \delta \neq \boldsymbol{r}^\top \boldsymbol{\mu}_j'$ for all $j \neq i$ (which is possible since $\boldsymbol{r}^\top \boldsymbol{\mu}_l \neq \boldsymbol{r}^\top \boldsymbol{\mu}_k$ for all $l \neq k \in [K]$). Define $\nu^{(i)} = (\nu_1', \ldots, \nu_K')$ and consider the problem $(\nu^{(i)}, P, \boldsymbol{r}, m)$. We claim that $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$. Since $\boldsymbol{\mu}_i \notin \partial P$ and $\partial P = \partial(\mathbb{R} \times P') = \mathbb{R} \times \partial P'$ for some $P' \subset \mathbb{R}^{D-1}$, $\boldsymbol{\mu}_i' \notin \partial P$. Further, by construction, $\boldsymbol{r}^\top \boldsymbol{\mu}_i' \neq \boldsymbol{r}^\top \boldsymbol{\mu}_j'$ for all $j \neq i$. Thus, none of the arms have means on the boundary of P and all of the rewards of the arms are distinct, so $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$.

Consider the event $B = \{i \in \widehat{\text{O}}\}$. Define $\text{OPT}_i = \text{OPT}(\nu^{(i)}, P, \boldsymbol{r}, m)$ and $\text{FEAS}_i = \text{FEAS}(\nu^{(i)}, P, \boldsymbol{r}, m)$. Observe that $i \notin \text{OPT}_i$ since $j_0 \in \text{FEAS}_i$ and $\boldsymbol{r}^\top \boldsymbol{\mu}_i' < \boldsymbol{r}^\top \boldsymbol{\mu}_{j_0}' = \max_{j \in \text{FEAS} \cap \text{OPT}^c} \boldsymbol{r}^\top \boldsymbol{\mu}_j$., so that there are m feasible arms with reward greater than $\boldsymbol{r}^\top \boldsymbol{\mu}_i'$.

Then, since $\mathcal{A}$ is δ- PAC wrt to $\mathcal{M}$, $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$, and arm $i \notin \text{OPT}_i$, we have that

$$\Pr_{\nu^{(i)}}(B) \leqslant \Pr_{\nu^{(i)}}(\text{OPT} \neq \widehat{\text{O}}) \leqslant \delta. \tag{81}$$

Further, since $\mathcal{A}$ is δ- PAC wrt $\mathcal{M}$,

$$\Pr_{\nu}(i \in \widehat{\text{O}}) \geqslant \Pr_{\nu}(\text{OPT} = \widehat{\text{O}}) \geqslant 1 - \delta. \tag{82}$$

Then,

$$\begin{aligned}
\frac{1}{2}[\boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{j_0}) + \epsilon]^2 \mathbb{E}_\nu[N_i(\tau)] &= \text{KL}(\nu_i, \nu_i')\mathbb{E}_\nu[N_i(\tau)] &(83)\\
&\geqslant d(\Pr_\nu(B), \Pr_{\nu^{(i)}}(B)) &(84)\\
&\geqslant d(\Pr_\nu(B), \delta) &(85)\\
&\geqslant d(\frac{1-\delta}{2}, \delta) &(86)\\
&\geqslant \ln(\frac{1}{2.4\delta}). &(87)
\end{aligned}$$

Line (83) follows by the formula for the KL-divergence of two multivariate normal distributions, (84) follows by Lemma F.1, (85) follows since $x \mapsto d(x, y)$ is increasing when $x > y$, (81), (82), and $\delta < .1$, (86) follows since $y \mapsto d(x, y)$ is decreasing when $x > y$, (81), (82), and $\delta < .1$, and (87) follows by Lemma F.8. The claim follows by rearranging the inequality.

Step 1.b: feasibility bound for $i \in$ OPT. A similar argument to step 2.b from the proof of Theorem 1 yields

$$\frac{1}{2}(\text{ dist}(\boldsymbol{\mu}_i, \partial P) + \epsilon)^2 \mathbb{E}_\nu[N_i(\tau)] \geqslant \ln(\frac{1}{2.4\delta}). \tag{88}$$

Step 2: $i \in$ FEAS $\cap$ OPT^c.

This step is very similar to step 3 of the proof of Theorem 1 and yields

$$\ln(\frac{1}{2.4\delta}) \leqslant \frac{1}{2}[\boldsymbol{r}^\top(\boldsymbol{\mu}_{j_0} - \boldsymbol{\mu}_i) + \epsilon]^2 \mathbb{E}_\nu[N_i(\tau)]. \tag{89}$$

Step 3: $i \in$ INFEAS $\cap$ SUBOPT^c. Since $P \neq \mathbb{R}^D$ and P is nonempty, by Lemma F.6 ∂P is nonempty. Since in addition ∂P is closed, by Lemma F.2, there exists $\boldsymbol{\tau}_i \in \text{Proj}_{\partial P}(\boldsymbol{\mu}_i)$. By definition of $\mathcal{M}$, since $\boldsymbol{\tau}_i \in \partial P$, for every $\epsilon > 0$, $B_\epsilon(\boldsymbol{\tau}_i) \cap P^\circ \neq \varnothing$. Thus, for sufficiently small $\epsilon > 0$, there exists a direction $\boldsymbol{v} \in \mathbb{R}^D$ with $\|\boldsymbol{v}\|_2 = 1$ such that $\boldsymbol{\tau}_i + \epsilon \boldsymbol{v} \in P^\circ$. Since by definition of $\mathcal{M}$, $P = \mathbb{R} \times P'$ for some $P' \subset \mathbb{R}^{D-1}$, we can choose $\boldsymbol{v}$ such that $v_1 = 0$. Define for all $j \in [K]$

$$\boldsymbol{\mu}'_j = \begin{cases} \boldsymbol{\tau}_i + \epsilon \boldsymbol{v} & \text{if } j = i \\ \boldsymbol{\mu}_j & \text{if } j \neq i \end{cases}$$
$$\nu'_j = N(\boldsymbol{\mu}'_j, I_D).$$

Define $\nu^{(i)} = (\nu'_1, \ldots, \nu'_K)$ and consider the problem $(\nu^{(i)}, P, \boldsymbol{r}, m)$. It follows that $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$ by a similar argument that showed in step 2.b of the proof of Theorem 1 that when $i \in$ OPT, $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$.

Define the event $B = \{i \notin \widehat{\text{O}}\}$. Define OPT$_i$ = OPT$(\nu^{(i)}, P, \boldsymbol{r}, m)$ and SUBOPT$_i$ = SUBOPT$(\nu^{(i)}, P, \boldsymbol{r}, m)$. Then, $i \in$ OPT$_i$ since $\boldsymbol{\mu}'_i \in P$ and $i \in$ SUBOPT$_i^c$ implies that $\boldsymbol{r}^\top \boldsymbol{\mu}'_i \geqslant \max^{(m)}_{l \in \text{FEAS}} \boldsymbol{r}^\top \boldsymbol{\mu}'_l$. Thus, since $\mathcal{A}$ is δ- PAC wrt $\mathcal{M}$,

$$\Pr_{\nu^{(i)}}(B) \leqslant \Pr_{\nu^{(i)}}(\widehat{\text{O}} \neq \text{OPT}) \leqslant \delta, \quad \text{and} \quad \Pr_{\nu}(B) \geqslant 1 - \delta.$$

Therefore, by a series of inequalities similar to those in (22)-(25) in ste 2.b of the proof of Theorem 1,

$$\ln(\frac{1}{2.4\delta}) \leqslant \frac{1}{2}(\text{ dist}(\boldsymbol{\mu}_i, P) + \epsilon)^2 \mathbb{E}_\nu[N_i(\tau)]. \tag{90}$$

Step 4: $i \in$ **INFEAS** $\cap$ **SUBOPT.** If INFEAS $\cap$ SUBOPT $= \varnothing$, there is nothing to show. Thus, we may suppose without loss of generality that INFEAS $\cap$ SUBOPT $\neq \varnothing$. Then, since in particular SUBOPT $\neq \varnothing$, there are m feasible arms and we may define

$$j_0 = \arg\max^{(m)}_{l \in \text{FEAS}} \boldsymbol{r}^\top \boldsymbol{\mu}_l.$$

By the same argument at the beginning of Step 3, there exists $\boldsymbol{\tau}_i \in \text{Proj}_{\partial P}(\boldsymbol{\mu}_i)$ and for sufficiently small $\epsilon > 0$, there exists a direction $\boldsymbol{v} \in \mathbb{R}^D$ with $\|\boldsymbol{v}\|_2 = 1$ and $v_1 = 0$ such that $\boldsymbol{\tau}_i + \epsilon \boldsymbol{v} \in P^\circ$. Define for all $j \in [K]$

$$\boldsymbol{\mu}'_j = \begin{cases} \begin{pmatrix} \mu_{i,1} + \mu_{j_0,1} + \epsilon \\ \boldsymbol{\tau}_{i,2:D} + \epsilon \boldsymbol{v}_{2,D} \end{pmatrix} & \text{if } j = i \\ \boldsymbol{\mu}_j & \text{if } j \neq i \end{cases}$$
$$\nu'_j = N(\boldsymbol{\mu}'_j, I_D).$$

where we choose $\epsilon > 0$ sufficiently small so that for any $\delta \in [0, \epsilon)$, $\boldsymbol{r}^\top \boldsymbol{\mu}'_i - \delta \neq \boldsymbol{r}^\top \boldsymbol{\mu}'_j$ for all $j \neq i$ (which is possible since $\boldsymbol{r}^\top \boldsymbol{\mu}_l \neq \boldsymbol{r}^\top \boldsymbol{\mu}_k$ for all $l \neq k \in [K]$). Then, define $\nu^{(i)} = (\nu'_1, \ldots, \nu'_K)$ and consider the problem $(\nu^{(i)}, P, \boldsymbol{r}, m)$. Using arguments similar to those in step 1, it follows that $(\nu^{(i)}, P, \boldsymbol{r}, m) \in \mathcal{M}$.

Consider the event $B = \{i \notin \widehat{\text{O}}\}$. Then, $\Pr_\nu(B) \geqslant 1 - \delta$. Define for the sake of brevity OPT$_i$ = OPT$(\nu^{(i)}, P, \boldsymbol{r}, m)$. Observe that $\boldsymbol{\mu}'_i \in P$ and $\boldsymbol{r}^\top \boldsymbol{\mu}'_i > \boldsymbol{r}^\top \boldsymbol{\mu}'_{j_0}$, so that $i \in$ OPT$_i$. Then, since $\mathcal{A}$ is δ- PAC wrt $\mathcal{M}$, $\Pr_{\nu^{(i)}}(B) \leqslant \delta$. Then,

$$\ln(\frac{1}{2.4\delta}) \leqslant \text{KL}(\nu_i, \nu^{(\epsilon)}_{i,P,\boldsymbol{r}}) \mathbb{E}_\nu[N_i(\tau)] \tag{91}$$
$$= \frac{1}{2}(\text{ dist}(\boldsymbol{\mu}_i, P) + \epsilon)^2 + [\boldsymbol{r}^t(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{j_0}) + \epsilon]^2] \mathbb{E}_\nu[N_i(\tau)] \tag{92}$$
$$\leqslant \max((\text{ dist}(\boldsymbol{\mu}_i, P) + \epsilon)^2, [\boldsymbol{r}^t(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{j_0}) + \epsilon]^2) \mathbb{E}_\nu[N_i(\tau)] \tag{93}$$

where line (91) follows by a series of inequalities similar to (83)-(87), and line (92) follows by the definition of KL divergence of multivariate normal distributions.

Step 5: Putting it together. Using $\mathbb{E}_\nu[\tau] = \sum_{i=1}^K \mathbb{E}_\nu[N_i(\tau)]$ and inequalities (87). (88), (89), and (90), we establish for all sufficiently small $\epsilon > 0$,

$$\begin{aligned}\mathbb{E}_\nu[\tau] \geqslant 2\ln(\frac{1}{2.4\delta})\Big[&\sum_{i \in \text{OPT}} \max([\min_{j \in \text{OPT}^c \cap \text{FEAS}} \boldsymbol{r}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \epsilon]^{-2}, [\,\text{dist}(\boldsymbol{\mu}_i, \partial P) + \epsilon]^{-2}) \\ &+ \sum_{i \in \text{OPT}^c \cap \text{FEAS}} [\min_{j \in \text{OPT}} \boldsymbol{r}^\top(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i) + \epsilon]^{-2} + \sum_{i \in \text{INFEAS} \cap \text{SUBOPT}^c} [\,\text{dist}(\boldsymbol{\mu}_i, P) + \epsilon]^{-2} \\ &+ \sum_{i \in \text{INFEAS} \cap \text{SUBOPT}} \frac{1}{2}\min([\min_{j \in \text{OPT}} \boldsymbol{r}^\top(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i) + \epsilon]^{-2}, [\,\text{dist}(\boldsymbol{\mu}_i, P) + \epsilon]^{-2})\Big].\end{aligned}$$

Since this bound holds for all $\epsilon > 0$ sufficiently small, letting $\epsilon \longrightarrow 0$ on the RHS of the above inequality establishes the result.

□

F Technical Lemmas

We use the following lemma from Kaufmann et al. [4]. Although they prove it for the case where arms are associated with scalar distributions, the proof generalizes to multi-dimensional distributions by simply replacing the scalar-valued distributions in the proof with vector-valued distributions. Let $I_t \in [K]$ denote the arm chosen by an agent at time t and $\boldsymbol{X}_t \sim \nu_{I_t}$. Let $\mathcal{F}_t = \sigma(I_1, \boldsymbol{X}_1, \ldots, I_t, \boldsymbol{X}_t)$, i.e., the sigma-algebra generated by $I_1, \boldsymbol{X}_1, \ldots, I_t, \boldsymbol{X}_t$.

Lemma F.1. *Let ν and ν' be two bandit models with K arms such that for all a, the distributions ν_a and ν'_a are mutually absolutely continuous. Let τ denote a stopping time wrt $(\mathcal{F}_t)$. Then,*

$$\sum_{i=1}^K \mathbb{E}_\nu[N_i(\tau)]\, KL(\nu_a, \nu'_a) \geqslant \sup_{E \in \mathcal{F}_\tau} d(\Pr_\nu(E), \Pr_{\nu'}(E))$$

Lemma F.2. *Let $\boldsymbol{x} \in \mathbb{R}^D$ and $A \subset \mathbb{R}^D$ be a closed nonempty set. Then, $\text{Proj}_A(\boldsymbol{x})$ is nonempty.*

Proof. Let $r > 0$ large enough such that $\bar{B}_r(\boldsymbol{x}) \cap A \neq \emptyset$. Then, observe that there exists $\boldsymbol{y} \in \text{Proj}_{A \cap \bar{B}_r(\boldsymbol{x})}(\boldsymbol{x})$ since $A \cap \bar{B}_r(\boldsymbol{x})$ is a compact set and $\|\cdot\|_2$ is continuous. Towards a contradiction, suppose there exists $\boldsymbol{z} \in A$ such that

$$\|\boldsymbol{z} - \boldsymbol{x}\|_2 < \|\boldsymbol{y} - \boldsymbol{x}\|_2\,.$$

Then, $\boldsymbol{z} \in A \cap \bar{B}_r(\boldsymbol{x})$, which implies that $\boldsymbol{y} \notin \text{Proj}_{A \cap \bar{B}_r(\boldsymbol{x})}(\boldsymbol{x})$, a contradiction. Thus, for all $\boldsymbol{z} \in A$,

$$\|\boldsymbol{y} - \boldsymbol{x}\|_2 \leqslant \|\boldsymbol{x} - \boldsymbol{y}\|_2\,.$$

Thus, $\boldsymbol{y} \in \text{Proj}_A(\boldsymbol{x})$. □

Lemmas F.3 and F.4 appear in Katz-Samuels and Scott [3]. For the sake of completeness, we restate the proof.

Lemma F.3. *Let* $P = \{\boldsymbol{x} \in \mathbb{R}^D : A\boldsymbol{x} \leqslant \boldsymbol{b}\}$ *with* $A \in \mathbb{R}^{M \times D}$. *Let* $\boldsymbol{\mu} \in P$. *Then,*

$$\operatorname{dis}(\boldsymbol{\mu}, \partial P) = \min_{i=1,\ldots,M} \operatorname{dis}(\boldsymbol{\mu}, \{\boldsymbol{x} : \boldsymbol{a}_i^\top \boldsymbol{x} = b_i\}).$$

Proof. It is not hard to establish that $\partial P = P \cap (\cup_{i=1}^M \{\boldsymbol{x} : \boldsymbol{a}_i^\top \boldsymbol{x} = b_i\})$. We claim that

$$\operatorname{dis}(\boldsymbol{\mu}, \cup_{i=1}^M \{\boldsymbol{x} : \boldsymbol{a}_i^\top \boldsymbol{x} = b_i\}) = \operatorname{dis}(\boldsymbol{\mu}, P \cap (\cup_{i=1}^M \{\boldsymbol{x} : \boldsymbol{a}_i^\top \boldsymbol{x} = b_i\})).$$

Since $\cup_{i=1}^M \{\boldsymbol{x} : \boldsymbol{a}_i^\top bx = b_i\}$ is closed, there exists $\boldsymbol{y} \in \cup_{i=1}^M \{\boldsymbol{x} : \boldsymbol{a}_i^\top bx = b_i\}$ such that

$$\|\boldsymbol{\mu} - \boldsymbol{y}\|_2 = \operatorname{dis}(\boldsymbol{\mu}, \cup_{i=1}^M \{\boldsymbol{x} : \boldsymbol{a}_i^\top bx = b_i\}).$$

We claim that $\boldsymbol{y} \in P$. Suppose not (towards a contradiction). Then, there exists $\theta \in (0, 1)$ such that $\boldsymbol{z} = (1 - \theta)\boldsymbol{\mu} + \theta \boldsymbol{y} \in \partial P$. Then,

$$\operatorname{dis}(\boldsymbol{\mu}, (\cup_{i=1}^M \{\boldsymbol{x} : \boldsymbol{a}_i^\top \boldsymbol{x} = b_i\})) \leqslant \|\boldsymbol{z} - \boldsymbol{\mu}\|_2 < \|\boldsymbol{y} - \boldsymbol{\mu}\|_2 = \operatorname{dis}(\boldsymbol{\mu}, \cup_{i=1}^M \{\boldsymbol{x} : \boldsymbol{a}_i^\top bx = b_i\}),$$

which is a contradiction, establishing the claim. Then,

$$\begin{aligned}
\min_{i=1,\ldots,M} \operatorname{dis}(\boldsymbol{\mu}, \{\boldsymbol{x} : \boldsymbol{a}_i^\top \boldsymbol{x} = b_i\}) &= \operatorname{dis}(\boldsymbol{\mu}, \cup_{i=1}^M \{\boldsymbol{x} : \boldsymbol{a}_i^\top \boldsymbol{x} = b_i\}) \\
&= \operatorname{dis}(\boldsymbol{\mu}, P \cap (\cup_{i=1}^M \{\boldsymbol{x} : \boldsymbol{a}_i^\top \boldsymbol{x} = b_i\})) \\
&= \operatorname{dis}(\boldsymbol{\mu}, \partial P).
\end{aligned}$$

□

Lemma F.4. *Let* $\epsilon > 0$ *and* $\mathcal{N}_\epsilon$ *be an* ϵ*-net of* $\mathcal{S}^{D-1}$. *For any* $\boldsymbol{y} \in \mathbb{R}^D$,

$$\|\boldsymbol{y}\|_2 \leqslant \frac{1}{1-\epsilon} \sup_{\boldsymbol{z} \in \mathcal{N}_\epsilon} \boldsymbol{y}^\top \boldsymbol{z}.$$

Proof. Let $\boldsymbol{z}_0 \in \mathcal{N}_\epsilon$ such that $\left\| \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2} - \boldsymbol{z}_0 \right\|_2 \leqslant \epsilon$. Then, by Cauchy-Schwarz,

$$\|\boldsymbol{y}\|_2 = \frac{\boldsymbol{y}^\top \boldsymbol{y}}{\|\boldsymbol{y}\|_2} = \boldsymbol{y}^\top (\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2} - \boldsymbol{z}_0) + \boldsymbol{y}^\top \boldsymbol{z}_0 \leqslant \|\boldsymbol{y}\|_2 \left\| \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2} - \boldsymbol{z}_0 \right\|_2 + \boldsymbol{y}^\top \boldsymbol{z}_0 \leqslant \epsilon \|\boldsymbol{y}\|_2 + \boldsymbol{y}^\top \boldsymbol{z}_0.$$

Rearranging the inequality, we obtain

$$\|\boldsymbol{y}\|_2 \leqslant \frac{1}{1-\epsilon} \boldsymbol{y}^\top \boldsymbol{z}_0 \leqslant \frac{1}{1-\epsilon} \sup_{\boldsymbol{z} \in \mathcal{N}_\epsilon} \boldsymbol{y}^\top \boldsymbol{z}.$$

□

The following Lemma appears in Vershynin et al. [5] (see Corollary 4.2.13).

Lemma F.5. *Let* $\epsilon > 0$ *and* $\mathcal{N}_\epsilon$ *be a minimal* ϵ*-net of* $\mathcal{S}^{D-1}$. *Then,* $|\mathcal{N}_\epsilon| \leqslant (\frac{2}{\epsilon} + 1)^D$.

Lemma F.6. *Suppose* $A \subset \mathbb{R}^D$ *is nonempty and* $A \neq \mathbb{R}^D$. *Then,* A *has nonempty boundary.*

Proof. Suppose that A has empty boundary. Then, for every $\boldsymbol{x} \in A$, there exists a sufficiently small ball B containing $\boldsymbol{x}$ such that $B \subset A$ and for every $\boldsymbol{y} \in A^c$, there exists a sufficiently small ball B' containing $\boldsymbol{y}$ such that $B' \subset A^c$. Then, A and A^c are both open sets, which contradicts the assumption that A is nonempty and $A \neq \mathbb{R}^D$. ☐

Recall that $d(x, y) \coloneqq x \log(\frac{x}{y}) + (1-x)\log(\frac{1-x}{1-y})$.

Lemma F.7. *For $x \leqslant .1$,*

$$d(\frac{1-x}{2}, x) = \frac{1-x}{2}\ln(\frac{1-x}{2x}) + \frac{1+x}{2}\ln(\frac{1+x}{2(1-x)}) \geqslant \frac{1}{15}\ln(\frac{1}{2x}).$$

Proof. We note that the term

$$\frac{1+x}{2}\ln(\frac{1+x}{2(1-x)}) = \frac{1+x}{2}(\ln(1+x) - \ln(2(1-x)))$$

is increasing in $x \in (0, 1)$. Thus, for all $x \in (0, 1)$,

$$\frac{1+x}{2}(\ln(1+x) - \ln(2(1-x))) \geqslant \frac{1}{2}\ln(1/2) \geqslant -.35.$$

Next, for $x \leqslant .1$,

$$\begin{aligned}
\frac{1-x}{2}\ln(\frac{1-x}{2x}) &= \frac{1-x}{2}[\ln(1-x) + \ln(\frac{1}{2x})] \\
&\geqslant \frac{1-x}{2}[\ln(0.9) + \ln(\frac{1}{2x})] \\
&\geqslant \frac{1}{2}\cdot(-0.106) + \frac{1-x}{2}\ln(\frac{1}{2x}) \\
&\geqslant \frac{1}{2}\cdot(-0.106) + \frac{1}{3}\ln(\frac{1}{2x}).
\end{aligned}$$

Then, putting it together, for $x \leqslant .1$,

$$\begin{aligned}
\frac{1+x}{2}\ln(\frac{1+x}{2(1-x)}) + \frac{1-x}{2}\ln(\frac{1-x}{2x}) &\geqslant \frac{1}{3}\ln(\frac{1}{2x}) - 0.35 - \frac{1}{2}\cdot(0.106) \\
&\geqslant \frac{1}{3}\ln(\frac{1}{2x}) - \frac{4}{15}\ln(\frac{1}{2x}) \\
&= \frac{1}{15}\ln(\frac{1}{2x}).
\end{aligned}$$

where we used the fact that $\frac{4}{15}\ln(\frac{1}{2x}) \geqslant 0.403$ for all $x \leqslant 0.1$. ☐

The following Lemma is from Kaufmann et al. [4].

Lemma F.8. *For any $x \in [0, 1]$, $d(x, 1-x) \geqslant \ln(\frac{1}{2.4x})$.*

Lemma F.9. *Let $P \subset \mathbb{R}^D$ and let $\boldsymbol{x} \in P$ and $\boldsymbol{y} \in P^c$. Then, there exists $\theta \in [0, 1]$ such that $\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y} \in \partial P$.*

Proof. Since $\boldsymbol{x} \in P$ and $\boldsymbol{y} \in P^c$, by Lemma F.6, $\partial P \neq \emptyset$. Consider the following sequence, which resembles binary search.

$$\begin{aligned}
\boldsymbol{x}_0 &= \boldsymbol{x} \\
\boldsymbol{x}_1 &= \boldsymbol{y} \\
\boldsymbol{x}_2 &= \frac{1}{2}(\boldsymbol{x} + \boldsymbol{y}) \\
\boldsymbol{x}_n &= \begin{cases} \frac{1}{2}\boldsymbol{x}_{n-1} + \frac{1}{2}\boldsymbol{x}_{\min(k:\forall l \in \{k+1,\ldots,n-1\}, \boldsymbol{x}_l \in P)} & : \boldsymbol{x}_{n-1} \in P \\ \frac{1}{2}\boldsymbol{x}_{n-1} + \frac{1}{2}\boldsymbol{x}_{\min(k:\forall l \in \{k+1,\ldots,n-1\}, \boldsymbol{x}_l \in P^c)} & : \boldsymbol{x}_{n-1} \in P^c \end{cases}.
\end{aligned}$$

$\{\boldsymbol{x}_n\}$ is clearly a Cauchy sequence so that it has a a limit $\bar{\boldsymbol{x}} = \theta \boldsymbol{x} + (1-\theta)\boldsymbol{y} \in \partial P$ for some $\theta \in [0,1]$. If for every $N \in \mathbb{N}$, there exist $n, m \geqslant N$ such that $\boldsymbol{x}_n \in P$ and $\boldsymbol{x}_m \in P^c$, then it is clear that $\bar{\boldsymbol{x}} \in \partial P$. Suppose that there exists N such $\boldsymbol{x}_N \in P$ and for every $n > N$, $\boldsymbol{x}_n \notin P$ (the other case is similar). Then, it is clear that $\bar{\boldsymbol{x}} = \boldsymbol{x}_N$ and that every open ball containing $\bar{\boldsymbol{x}}$ contains some point not in P, so that $\bar{\boldsymbol{x}} \in \partial P$.

□

We use the anytime confidence interval from Kaufmann et al. [4].

Lemma F.10. *Let $X_1, X_2, \ldots$ be i.i.d. zero-mean sub-Gaussian random variables with scale $\sigma > 0$ and $\delta \in (0,1)$. Then,*

$$\Pr(\exists t : |\frac{1}{t}\sum_{s=1}^{\top} X_s| \geqslant \sigma\sqrt{\frac{2\log(1/\delta) + 6\log\log(1/\delta) + 3\log\log(et)}{t}}) \leqslant \delta.$$

Recall that $U(t,\delta) = \sigma\sqrt{\frac{2\log(1/\delta)+6\log\log(1/\delta)+3\log\log(et)}{t}}$. We use the following fact from Jamieson and Jain [2].

Lemma F.11. *Let $\Delta \in (0,1)$ and $\delta \in (0,1)$. There is a universal constant $c > 0$ such that if*

$$N \geqslant c\Delta^{-2}\log(\frac{\log(\Delta^{-2})}{s})$$

then $U(N,s) \leqslant \Delta$.

G TF-LUCB with Tolerance

In this section, we present a variant of TF-LUCB that tolerates some violation of the constraints and some suboptimality: TF-LUCB-Tol. TF-LUCB-Tol also takes as input two scalars ϵ_P and $\epsilon_{\boldsymbol{r}}$, which quantify how much the algorithm tolerates a violation of the constraints and suboptimality, respectively. The main difference is that TestF-Tol also takes as input ϵ_P and the stopping condition associated with the rewards is now

$$\min_{i \in \text{TOP}_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(t)} - U_{\boldsymbol{r}}(N_i(t),\delta) + \epsilon_{\boldsymbol{r}} \geqslant \max_{j \in \text{TOP}_t^c \cap E_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{j,N_j(t)} + U_{\boldsymbol{r}}(N_j(t),\delta).$$

Next, we introduce variants of TestF-B and TestF-C that allow for a tolerance. TestF-B-Tol now returns True if $B_{U_{\text{ball}}(t,\delta)}(\widehat{\boldsymbol{\mu}}_{i,t})$ intersects P and P^c and $U_{\text{ball}}(t,\delta) \leqslant \frac{\epsilon_P}{2}$. Since the diameter of $B_{U_{\text{ball}}(t,\delta)}(\widehat{\boldsymbol{\mu}}_{i,t})$ is

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

Input: TestF, sub-Gaussian norm bound σ, confidence δ, ϵ_P, $\epsilon_{\boldsymbol{r}}$
for $t = 1, 2, \ldots$ **do**
  $F_t \longleftarrow \{i \in [K] : \text{TestF}(i, N_i(t), \epsilon_P) = \text{ True}\}$ # arms that are determined to be feasible whp
  $G_t \longleftarrow \{i \in [K] : \text{TestF}(i, N_i(t), \epsilon_P) = ?\}$# arms that have not be determined to be feasible or infeasible 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 $\min_{i \in \text{TOP}_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i, N_i(t)} - U_{\boldsymbol{r}}(N_i(t), \delta) + \epsilon_{\boldsymbol{r}} \geqslant \max_{j \in \text{TOP}_t^c \cap E_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{j, N_j(t)} + U_{\boldsymbol{r}}(N_j(t), \delta)$
    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

$2\, U_{\text{ball}}(t, \delta)$, this guarantees that on an event where the confidence bounds work appropriately, we only accept $\boldsymbol{\mu}_i$ such that $\text{dist}(\boldsymbol{\mu}_i, P) \leqslant \epsilon_P$. TestF-C-Tol tolerates violations on a constraint-basis instead. Now, it accepts arms if $U_{\text{con}}(t, \delta) \leqslant \frac{\epsilon_P}{2}$. Thus, assuming an event on which the confidence bounds work appropriately, it only tolerates mistakes on arms such that for every constraint $j \in [M]$, $\boldsymbol{a}_j^\top \boldsymbol{\mu}_i \leqslant b_j + \epsilon_P$.

Algorithm 2 TestF-B-Tol:

Input: arm index i, number of pulls t, ϵ_P
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
if $U_{\text{ball}}(t, \delta) \leqslant \frac{\epsilon_P}{2}$
  return True
else
  return ?

Algorithm 3 TestF-C-Tol:

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

Proving the upper bound for this algorithm would have a similar structure to what we have done in this paper. One subtlety is that finding the top m feasible arms depends on which arms we consider to be feasible so that accepting as feasible an arm that is in fact infeasible might make the problem more difficult. We conjecture that the upper bound would reflect this subtlety. We leave the proof of an upper bound of this to future work.

However, as a practical consideration, we also note that accepting as feasible an arm that is in fact infeasible might make the problem much easier. We conjecture that in most applications, there is no a priori reason to believe that doing this would make the problem easier or more difficult. Furthermore, this issue could be somewhat alleviated by allowing a tolerance for suboptimality.

H Pseudocode for algorithms TF-AE and FFAF

Algorithm 4 TF-AE: Top-m Feasible Action Elimination

Input: TestF, sub-Gaussian norm bound σ, confidence δ
$t \longleftarrow 1$
while True **do**

  $F_t \longleftarrow \{i \in [K] : \text{TestF}(i, N_i(t)) = \text{ True}\}$ # arms that are determined to be feasible whp
  $G_t \longleftarrow \{i \in [K] : \text{TestF}(i, N_i(t)) = ?\}$# arms that have not be determined to be feasible or infeasible whp
  $E_t \longleftarrow F_t \cup G_t$ # arms that are not ruled out as infeasible whp
  $H_t \longleftarrow \{i \in [K] : |\{j \in F_t : \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{j,N_j(t)} - U_{\boldsymbol{r}}(N_j(t), \delta) \geqslant \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(t)} + U_{\boldsymbol{r}}(N_i(t), \delta)\}| < m\}$
  $Q_t \longleftarrow E_t \cap H_t$
  for $i \in Q_t$ **do**
    pull arm i
    $t \longleftarrow t + 1$
  if $E_t = F_t$ and $|F_t| < m$
    return F_t
  if $Q_t \subset F_t$ and $|Q_t| = m$
    return Q_t

For FFAF, we require that it find the the feasible arms with probability at least $1 - \frac{\delta}{2}$ and, then, to find the best arms among those with probability at least $1 - \frac{\delta}{2}$. Thus, we require that TestF output the correct answer with probability at least $1 - \frac{\delta}{2K}$. We modify the confidence bound for the rewards in the second stage since in that stage there are only $|F_t|$ arms among which the m arms with the largest rewards must be identified.

References

[1] S. Boyd and L. Vandenberghe. Convex optimization. *Cambridge University press*, 2004.

[2] K. Jamieson and L. Jain. A bandit approach to multiple testing with false discovery control. *arXiv preprint arXiv:1809.02235*, 2018.

[3] 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.

[4] 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.

Algorithm 5 FFAF: Find Feasible Arms First

Input: TestF, sub-Gaussian norm bound σ, confidence δ
$t \longleftarrow 1$
for $i \in [K]$ **do**
  while TestF$(i, N_i(t))$ =? **do**
    Pull arm i
    $t \longleftarrow t + 1$
$F_t \longleftarrow \{i \in [K] :$ TestF$(i, N_i(t)) =$ True$\}$ # arms that are determined to be feasible whp
if $|F_t| \leqslant m$
  return F_t
$U_{\boldsymbol{r}}(s, \delta) \coloneqq U(s, \frac{\delta}{2|F_t|})$
while True **do**
  $\text{TOP}_t = \arg\max_{Z \subset F_t, |Z|=m} \sum_{i \in Z} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(t)}$
  if $\min_{i \in \text{TOP}_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(t)} - U_{\boldsymbol{r}}(N_i(t), \delta) \geqslant \max_{j \in \text{TOP}_t^c \cap F_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{j,N_j(t)} + U_{\boldsymbol{r}}(N_j(t), \delta)$
    return TOP_t
  $h_t = \arg\min_{i \in \text{TOP}_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{i,N_i(t)} - U_{\boldsymbol{r}}(N_i(t), \delta)$
  $l_t = \arg\max_{j \in \text{TOP}_t^c \cap F_t} \boldsymbol{r}^\top \widehat{\boldsymbol{\mu}}_{j,N_j(t)} + U_{\boldsymbol{r}}(N_j(t), \delta)$
  Pull arms h_t and l_t
  $t \longleftarrow t + 1$

[5] 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.