Best Arm Identification in Linear Bandits with Linear Dimension Dependency

Chao Tao [1] **Saúl A. Blanco** [1] **Yuan Zhou** [1 2 3]

Abstract

We study the best arm identification problem in linear bandits, where the mean reward of each arm depends linearly on an unknown d-dimensional parameter vector θ, and the goal is to identify the arm with the largest expected reward. We first design and analyze a novel randomized θ estimator based on the solution to the convex relaxation of an optimal G-allocation experiment design problem. Using this estimator, we describe an algorithm whose sample complexity depends linearly on the dimension d, as well as an algorithm with sample complexity dependent on the reward gaps of the best d arms, matching the lower bound arising from the ordinary top-arm identification problem. We finally compare the empirical performance of our algorithms with other state-of-the-art algorithms in terms of both sample complexity and computational time.

1. Introduction

The stochastic multi-armed bandit (MAB) problems are an important framework that not only addresses the fundamental trade-off between exploration and exploitation in sequential experiments, but also crystallizes the challenge of efficiently gathering information before committing to a final decision. Many authors have investigated the *pure exploration* form of the problem with N independent arms, e.g., to identify the best arm (top arm) ((Carpentier & Locatelli, 2016; Chen et al., 2017b; Even-Dar et al., 2006; Jamieson et al., 2014; Karnin et al., 2013; Kaufmann et al., 2016)) as well as identifying the set of best (top) K arms under different metrics ((Bubeck et al., 2013; Chen et al., 2017a; 2014; Kalyanakrishnan & Stone, 2010; Zhou et al., 2014)).

In this paper, we study the top arm identification problem in the stochastic *linear bandit* setting, introduced and studied in (Auer, 2002; Soare et al., 2014) [1]. In the linear bandit setting, instead of having N independent arms, each of the N arms is associated with a d-dimensional *attribute vector*, and the expected reward is the linear combination of its own attribute vector and an unknown global vector θ. Each pull of an arm reports a stochastic reward centered at its expectation, and the goal is to pull as few times as possible to identify the arm with the highest expected reward. Due to the linear structure of the problem, pulling each arm reveals information about the global vector θ and therefore, indirectly, about the expected reward of other arms.

Linear bandit problems find many applications in practice. For example, in online advertising, suppose the goal is to select an advertisement from a pool to maximize the likelihood for a group of unknown audience to click. The likelihood can usually be approximated by a linear function of logical combinations of various attributes associated with the audience and the ads (such as age, sex, the domain, keywords, ad genres, etc.). Now the linear top-arm identification directly addresses this problem if each ad is abstracted to be an arm with the known attribute vector. While the regret analysis of the exploration-and-exploitation linear bandit problem has been extensively studied (Abbasi-Yadkori et al., 2011; Li et al., 2010; 2017), this paper studies the pure-exploration scenario and is devoted to showing a fixed-confidence algorithm with the optimal sample complexity.

Problem formulation. We are given a set of N arms $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$. Each arm is associated with a d-dimensional attribute vector and we overload the notation by using $x_i \in \mathbb{R}^d$ to denote the vector for the arm x_i. We assume $\|x_i\|_2 \leq 1$ for all arms. We also assume w.l.o.g. that the rank of $\mathcal{X}$ is exactly d (otherwise we can project the set of vectors to a smaller-dimensional space).

There is an unknown global vector $\theta \in \mathbb{R}^d$ with $\|\theta\|_2 \leq L$. Each pull of arm x_i reports a stochastic reward $x_i^{\mathrm{T}}\theta + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon}$ is an independent κ-subgaussian noise (see Sec-

[1]Department of Computer Science, Indiana University at Bloomington, Indiana, USA [2]Institute for Theoretical Computer Science, Shanghai University of Finance and Economics, Shanghai, China [3]Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, Illinois, USA. Correspondence to: Yuan Zhou <yzhoucs@iu.edu>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

[1]Very recently, the problem was also studied independently in (Xu et al., 2018).

tion 2 for the necessary definitions).

Given a confidence parameter $\delta \in (0,1)$, the goal is to design an algorithm with as small query complexity Q as possible that, with probability at least $(1-\delta)$, sequentially makes at most Q pulls of the arms, and then identifies the arm with the highest mean reward.

Previous results. To make the best arm unambiguously defined, we assume $\mathcal{X}_{[1]}^{\mathrm{T}}\theta > \mathcal{X}_{[i]}^{\mathrm{T}}\theta$ for $i > 1$ where $\mathcal{X}_{[i]}$ denotes the arm in $\mathcal{X}$ whose mean reward (i.e., $\mathcal{X}_{[i]}^{\mathrm{T}}\theta$) is the i-th largest. We let $\Delta_i = \mathcal{X}_{[1]}^{\mathrm{T}}\theta - \mathcal{X}_{[i]}^{\mathrm{T}}\theta$ be the *reward gap* between arm $\mathcal{X}_{[1]}$ and $\mathcal{X}_{[i]}$ for $i \geq 2$. In (Soare et al., 2014), the authors showed a few algorithms to identify the top arm and their best query complexity is [2]

$$O\left(\frac{d}{\Delta_2^2}(\ln\delta^{-1} + \ln N + \ln\Delta_2^{-1}) + d^2\right). \qquad (1)$$

These algorithms were combined with the Explore-Verify framework proposed in (Karnin, 2016). For very small δ (i.e. $\delta = (\Delta_2/N)^{\omega(1)}$), the sample complexity was improved (in (Karnin, 2016)) to

$$O\left(\frac{d}{\Delta_2^2}(\ln N + \ln\Delta_2^{-1}) + \rho^*\ln\delta^{-1} + d^2\right), \qquad (2)$$

where $\rho^* \geq \Delta_2^{-2}$ is an information-theoretic lower bound of the sample complexity on the input instance.

Our contribution and techniques. In this paper, we show linear top-arm identification algorithms with confidence $(1-\delta)$ and sample complexity

$$O\left(\frac{d}{\Delta_2^2}(\ln\delta^{-1} + \ln N + \ln\ln\Delta_2^{-1})\right). \qquad (3)$$

The main improvement compared to (1) is the removal of the additive d^2 term, so that our sample complexity is truly linear with the dimension. We achieve this goal by proposing and analyzing a novel estimator of θ described in Section 3. Both our algorithms and (Soare et al., 2014) carefully choose the pulling strategy in order to get a better estimate for θ. In (Soare et al., 2014), this is done by (deterministically) approximating the optimal G-allocation experiment design (which is NP-Hard (Çivril & Magdon-Ismail, 2009)), and leads to the additional d^2 term in the sample complexity. In our algorithms, however, we solve the convex relaxation of the optimal G-allocation problem, and then, based on the optimal solution, construct a novel randomized estimator of θ. While our solution does not guarantee to solve the optimal G-allocation problem, it saves the extra d^2 term.

Our algorithms also make significant improvement in terms of time complexity. The previous algorithms in (Soare et al., 2014) have to spend $\Omega(N^2d^2)$ time to compare the inverse of several matrices before *each* pull. In contrast, the overhead of our algorithm is to solve the convex relaxation and compute the inverse of a $d \times d$ matrix (which costs roughly $O(Nd^2)$ time), and the remaining time cost before each pull is negligible.

We also develop fully data-dependent algorithms. For confidence parameter δ, the sample complexity of our algorithm is

$$O\left(\sum_{i=2}^{d}\Delta_i^{-2}(\ln\delta^{-1} + \ln N + \ln\ln\Delta_i^{-1})\right). \qquad (4)$$

We finally empirically evaluate our algorithms with extensive synthetic and real-world data-sets, and compare the sample complexity and run time with other state-of-the-art algorithms in Section 5.

1.1. Related Work

Relation to optimal experiment designs. The problem is closely related to the optimal G-allocation problem in experiment design. The greedy approach proposed in (Soare et al., 2014) was analyzed for different optimality criteria such as A-optimality ((Bian et al., 2017; Chamon & Ribeiro, 2017a)), E-optimality ((Chamon & Ribeiro, 2017a)), and V-optimality ((Chamon & Ribeiro, 2017b)). The A, D, E, V, and G-allocation design problems were also studied using other approximation frameworks (e.g. (Allen-Zhu et al., 2017)). We also know that exactly solving the G-allocation design problem is NP-Hard (Çivril & Magdon-Ismail, 2009).

Relation to the ordinary top-arm identification problem. As pointed out in (Soare et al., 2014), the ordinary top-arm identification problem for d independent arms is a special case of the linear bandit setting when each arm x_i is associated with the i-th unit basis vector in $\mathbb{R}^d$. Therefore, the sample complexity lower bound (e.g. from (Chen et al., 2017b)) $\Omega(\sum_{i=2}^{d}\Delta_i^{-2}\ln\delta^{-1})$ directly applies to the linear bandit setting.

[2] The termination condition of their algorithm is $Q \leq cd(1+\beta)(\ln Q + \ln N + \ln\delta^{-1})/\Delta_2^2$, where c is a constant and $\beta = d^2/Q$ results from the efficient approximation of the NP-Hard optimal G-allocation experiment design problem. Solving the inequality for Q gives the sample complexity bound in (1). The extra d^2 term appears in (2) for the same reason.

We also note that there is a gap between the proofs and the algorithms in (Soare et al., 2014). More specifically, both G-Allocation and $\mathcal{XY}$-Allocation algorithms adopt a greedy approach to construct the set of arms to pull. However, the theoretical analysis (Lemmas 5 and 7 in Appendix C of (Soare et al., 2014)) are for a different convex-relaxation-and-rounding procedure which was proposed in (Pukelsheim, 2006).

Regret minimization for linear bandits. Regret minimization, another important goal of multi-armed bandit problems, for linear (and generalized linear) bandits has been extensively studied (e.g., (Abbasi-Yadkori et al., 2011; Auer, 2002; Chu et al., 2011; Dani et al., 2008; Filippi et al., 2010; Li et al., 2017; Rusmevichientong & Tsitsiklis, 2010)). However, its pure exploration counterpart has been less investigated. While regret and sample complexity bounds are not directly comparable, the main technical difference between our algorithms and most regret minimization algorithms is that our sample strategy is computed by solving the convex relaxation of an optimal G-allocation problem beforehand, which is important for achieving linear dimensional dependence and avoiding computationally expensive matrix inverse updates after each sample.

2. Preliminaries

Throughout the paper, we use bold letters to denote random variables. Moreover, for any positive semi-definite matrix A, $\|x\|_A$ is defined to be $\sqrt{x^{\mathrm{T}}Ax}$. Furthermore, recall that if A is a $d \times d$ positive definite matrix, then A^{-1} exists and is also positive definite. So by the Spectral Theorem, there exists an orthonormal matrix O and a diagonal matrix D such that $A^{-1} = ODO^{\mathrm{T}}$ with the elements at the diagonal of D being the eigenvalues of A^{-1}. Now we define $A^{-1/2} = OD^{1/2}O^{\mathrm{T}}$ where $D^{1/2}$ is also a diagonal matrix with $D_{ii}^{1/2} = \sqrt{D_{ii}}$ for all $i \in \{1, 2, 3, \ldots, d\}$.

We now define *subgaussian random variables* and spell out some of the mathematical tools that will be used in the proofs.

Definition 1 (Subgaussian Random Variable)**.** *For any $b > 0$, a real-valued random variable $\boldsymbol{X}$ is said to be b-subgaussian if it has the property that for every $t \in \mathbb{R}$ one has $\mathrm{E}[e^{t\boldsymbol{X}}] \le e^{b^2t^2/2}$.*

Proposition 2 (Special case of Lemma 2.3.18, (Stroock, 2011))**.** *If $\boldsymbol{X}$ is b-subgaussian, then for any $\alpha \in \mathbb{R}$, the random variable $\alpha\boldsymbol{X}$ is $|\alpha|b$-subgaussian. Moreover, if $\boldsymbol{X_1}$, $\boldsymbol{X_2}$ are independent random variables such that $\boldsymbol{X_i}$ is b_i-subgaussian, then $\boldsymbol{X_1} + \boldsymbol{X_2}$ is $\sqrt{b_1^2 + b_2^2}$-subgaussian.*

Proposition 3 (Subgaussian Tail Estimate, special case of Lemma 2.3.18, (Stroock, 2011))**.** *If $\boldsymbol{X}$ is b-subgaussian, then for any $\epsilon > 0$ we have that $\Pr[|\boldsymbol{X}| \ge \epsilon] \le 2\exp\left(-\frac{\epsilon^2}{2b^2}\right)$.*

We will find the following concentration inequalities useful.

Proposition 4 (Multiplicative Chernoff Bound)**.** *Suppose $\boldsymbol{X_i}$'s ($1 \le i \le n$) are independent random variables taking values in $[0, 1]$. Let $\boldsymbol{X} = \sum_{i=1}^n \boldsymbol{X_i}$ and $\mu = \mathrm{E}[\boldsymbol{X}]$. Then for any $\delta \in [0, 1]$, it holds that $\Pr[\boldsymbol{X} \ge (1+\delta)\mu] \le e^{-\frac{\delta^2\mu}{3}}$.*

Proposition 5 (Bernstein Inequality)**.** *Let $\boldsymbol{X_i}$ ($1 \le i \le n$) be independent zero-mean random variables. Suppose that $|\boldsymbol{X_i}| \le M$ almost surely, for any i. Then, for all positive t, it holds that $\Pr\left(\sum_{i=1}^n \boldsymbol{X_i} > t\right) \le \exp\left(-\frac{t^2/2}{\sum_{i=1}^n \mathrm{E}[\boldsymbol{X_i}^2] + \frac{1}{3}Mt}\right)$.*

Finally, we introduce the following theorem which will be crucially used in the design of our estimator for θ.

Proposition 6 (Restatement of the Equivalence-Theorem in (Kiefer & Wolfowitz, 1960))**.** *Given a set of d-dimensional vectors $\mathcal{X} \subseteq \mathbb{R}^d$, for any distribution λ supported on $\mathcal{X}$ such that $M(\lambda) = \mathrm{E}_{\boldsymbol{z}\sim\lambda}[\boldsymbol{z}\boldsymbol{z}^{\mathbf{T}}]$ is non-singular, we define $f(x;\lambda) = x^{\mathrm{T}}M(\lambda)^{-1}x$. The following extremum problems are equivalent:*

(a) Choosing λ so that
$$\lambda \text{ maximizes } \det M(\lambda). \tag{5}$$

(b) Choosing λ so that
$$\lambda \text{ minimizes } \max_{x\in\mathcal{X}} f(x;\lambda). \tag{6}$$

Moreover, since $\mathrm{E}_{\boldsymbol{x}\sim\lambda}[f(\boldsymbol{x};\lambda)] = d$, it follows that $\max_{x\in\mathcal{X}} f(x;\lambda) \ge d$. Therefore, a sufficient condition for λ to satisfy (6) *is*
$$\max_{x\in\mathcal{X}} f(x;\lambda) = d. \tag{7}$$

3. Main Theorems on the Estimator

3.1. The New Estimator

Let λ^* be the solution to the three problems defined in Proposition 6. Let $\boldsymbol{y_1}, \ldots, \boldsymbol{y_n}$ be n *i.i.d.* samples following from the distribution λ^*, corresponding to the n pulled arms. Let the corresponding rewards be $\boldsymbol{r_1}, \ldots, \boldsymbol{r_n}$ respectively. In the setting of linear bandits, we have $\boldsymbol{r_i} = \boldsymbol{y_i}^{\mathrm{T}}\theta + \boldsymbol{\epsilon_i}$ where $\boldsymbol{\epsilon_i}$ is κ-subgaussian. Let $\boldsymbol{b_i} = \boldsymbol{r_i}\boldsymbol{y_i}$ and $\boldsymbol{b} = \sum_{i=1}^n \boldsymbol{b_i}$. Now we define our estimator $\widehat{\boldsymbol{\theta}}$ to be
$$\widehat{\boldsymbol{\theta}} = A^{-1}\boldsymbol{b}, \tag{8}$$
where $A = nM(\lambda^*)$.

Remark 1. *When $y_1, \ldots, y_n$ are the n sampled arms, the usual maximum-likelihood estimator used in literature (e.g. (Li et al., 2017; Soare et al., 2014)) is $(\sum_{i=1}^n y_i y_i^{\mathrm{T}})^{-1}b$. However, it is technically difficult to analyze the (spectral) concentration properties of $(\sum_{i=1}^n y_i y_i^{\mathrm{T}})^{-1}$, and we novelly use $\left(\mathrm{E}\left[\sum_{i=1}^n \boldsymbol{y_i}\boldsymbol{y_i}^{\mathrm{T}}\right]\right)^{-1} = A^{-1}$ instead.*

3.2. The Key Lemma

For any $\theta, x \in \mathbb{R}^d$, and $\ell \ge 1$, we define a θ-dependent error and a θ-free error, $\mathrm{Err}_{\lambda^*}(x, \ell, \theta)$ and $\mathrm{err}_{\lambda^*}(x, \ell)$, respectively, as follows:

$$\begin{aligned}\mathrm{Err}_{\lambda^*}(x,\ell,\theta) = &\sqrt{\frac{2\,\mathrm{E}_{\boldsymbol{z}\sim\lambda^*}[(x^{\mathrm{T}}M(\lambda^*)^{-1}\boldsymbol{z}\cdot\boldsymbol{z}^{\mathrm{T}}\theta)^2]}{\ell}} \\ &+ \frac{2(|x^{\mathrm{T}}\theta| + L\|x\|_{M(\lambda^*)^{-1}}\sqrt{d})}{3\ell} \\ &+ \sqrt{2}\kappa\sqrt{\frac{\|x\|^2_{M(\lambda^*)^{-1}}}{\ell}\sqrt{\frac{3d}{\ell}} + \frac{\|x\|^2_{M(\lambda^*)^{-1}}}{\ell}},\end{aligned}$$

$$\text{and } \mathrm{err}_{\lambda^*}(x,\ell) = \frac{\sqrt{2}L\|x\|_{M(\lambda^*)^{-1}}}{\sqrt{\ell}} + \frac{2L(\|x\|_2 + \|x\|_{M(\lambda^*)^{-1}}\sqrt{d})}{3\ell} + \sqrt{2}\kappa\sqrt{\frac{\|x\|^2_{M(\lambda^*)^{-1}}}{\ell}\sqrt{\frac{3d}{\ell}} + \frac{\|x\|^2_{M(\lambda^*)^{-1}}}{\ell}}.$$

We show the following key lemma.

Lemma 7. *When $n \geq \ell\ln(5/\delta)$ where $\ell \geq 3d$, for any fixed vector (not necessarily an arm), with probability at least $1-\delta$, it holds that $|x^{\mathrm{T}}(\theta - \widehat{\boldsymbol{\theta}})| \leq \mathrm{Err}_{\lambda^*}(x,\ell,\theta) \leq \mathrm{err}_{\lambda^*}(x,\ell)$.*

Proof. Note that $\boldsymbol{b} = \sum_{i=1}^n \boldsymbol{r_i}\boldsymbol{y_i} = \sum_{i=1}^n (\boldsymbol{y_i}\boldsymbol{y_i}^{\mathrm{T}}\theta + \boldsymbol{y_i}\boldsymbol{\epsilon_i})$. For any given vector x, the absolute error between the real reward and the estimated reward is

$$|x^{\mathrm{T}}(\theta - \widehat{\boldsymbol{\theta}})| = |x^{\mathrm{T}}(\theta - A^{-1}\boldsymbol{b})| = \left|\sum_{i=1}^n (x^{\mathrm{T}}\theta/n - x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}\boldsymbol{y_i}^{\mathrm{T}}\theta) - \sum_{i=1}^n x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}\boldsymbol{\epsilon_i}\right|. \quad (9)$$

For any $i \in \{1,2,\ldots,n\}$, set $\boldsymbol{Y_i} = x^{\mathrm{T}}\theta/n - x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}\boldsymbol{y_i}^{\mathrm{T}}\theta$ and $\boldsymbol{Z_i} = x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}\boldsymbol{\epsilon_i}$, then (9) can be written as

$$\left|\sum_{i=1}^n \boldsymbol{Y_i} - \sum_{i=1}^n \boldsymbol{Z_i}\right| \leq |\boldsymbol{Y}| + |\boldsymbol{Z}|,$$

where $\boldsymbol{Y} = \sum_{i=1}^n \boldsymbol{Y_i}$ and $\boldsymbol{Z} = \sum_{i=1}^n \boldsymbol{Z_i}$.

The following two claims bound $\boldsymbol{Y}$ and $\boldsymbol{Z}$ respectively. We provide the proofs of both claims after the proof of this lemma.

Claim 8. *For $n \geq \ell\ln\delta^{-1}$, with probability $(1-\delta)$, we have $|\boldsymbol{Y}| \leq \sqrt{\frac{2\,\mathrm{E}_{\boldsymbol{z}\sim\lambda^*}[(x^{\mathrm{T}}M(\lambda^*)^{-1}\boldsymbol{z}\cdot\boldsymbol{z}^{\mathrm{T}}\theta)^2]}{\ell}} + \frac{2(|x^{\mathrm{T}}\theta| + L\|x\|_{M(\lambda^*)^{-1}}\sqrt{d})}{3\ell}$.*

Claim 9. *For $n \geq \ell\ln(4/\delta)$ where $\ell \geq 3d$, with probability $(1-\delta)$, we have $|\boldsymbol{Z}| \leq \sqrt{2}\kappa\sqrt{\frac{\|x\|^2_{M(\lambda^*)^{-1}}}{\ell}\sqrt{\frac{3d}{\ell}} + \frac{\|x\|^2_{M(\lambda^*)^{-1}}}{\ell}}$.*

The first inequality of Lemma 7 follows immediately by combining Claim 8 and Claim 9, and a union bound. The second inequality of Lemma 7 holds because of $\|\boldsymbol{z}\|_2 \leq 1$, $\|\theta\|_2 \leq L$, and the definition of $M(\lambda^*)$. □

Proof of Claim 8. Let t denote $\|x\|^2_{M(\lambda^*)^{-1}}$ and s denote $\mathrm{E}_{\boldsymbol{z}\sim\lambda^*}[(x^{\mathrm{T}}M(\lambda^*)^{-1}\boldsymbol{z}\cdot\boldsymbol{z}^{\mathrm{T}}\theta)^2]$. Recall that $\boldsymbol{Y_i} = x^{\mathrm{T}}\theta/n - x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}\boldsymbol{y_i}^{\mathrm{T}}\theta$.

It is easy to verify that $\mathrm{E}[\boldsymbol{Y_i}] = 0$. We also have

$$|\boldsymbol{Y_i}| \leq |x^{\mathrm{T}}\theta/n| + |x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}\boldsymbol{y_i}^{\mathrm{T}}\theta| \leq |x^{\mathrm{T}}\theta|/n + L|x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}|,$$

where the last inequality is due to $\|\theta\|_2 \leq L$ and $\|\boldsymbol{y_i}\|_2 \leq 1$. By Cauchy-Schwartz inequality, we have

$$|x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}| = |\langle A^{-1/2}x, A^{-1/2}\boldsymbol{y_i}\rangle| \leq \|A^{-1/2}x\|_2 \cdot \|A^{-1/2}\boldsymbol{y_i}\|_2 = \sqrt{x^{\mathrm{T}}A^{-1}x}\sqrt{\boldsymbol{y_i}^{\mathrm{T}}A^{-1}\boldsymbol{y_i}} \leq \sqrt{td/n^2}, \quad (10)$$

where the last inequality is due to the definition of A and Proposition 6. Hence it holds that

$$|\boldsymbol{Y_i}| \leq |x^{\mathrm{T}}\theta|/n + L\sqrt{td/n^2}. \quad (11)$$

Next, we bound $\sum_{i=1}^n \mathrm{E}[\boldsymbol{Y_i}^2]$

$$\begin{aligned}
&= \sum_{i=1}^n \mathrm{E}[(x^{\mathrm{T}}\theta)^2/n^2 - 2x^{\mathrm{T}}\theta/n \cdot x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}\boldsymbol{y_i}^{\mathrm{T}}\theta \\
&\quad + (x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}\cdot\boldsymbol{y_i}^{\mathrm{T}}\theta)^2] \\
&\leq (x^{\mathrm{T}}\theta)^2/n - \sum_{i=1}^n 2x^{\mathrm{T}}\theta/n \cdot x^{\mathrm{T}}A^{-1}(\mathrm{E}\,\boldsymbol{y_i}\boldsymbol{y_i}^{\mathrm{T}})\theta \\
&\quad + \sum_{i=1}^n \mathrm{E}[(x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}\cdot\boldsymbol{y_i}^{\mathrm{T}}\theta)^2] \\
&= -(x^{\mathrm{T}}\theta)^2/n + \mathop{\mathrm{E}}_{\boldsymbol{z}\sim\lambda^*}[(x^{\mathrm{T}}M(\lambda^*)^{-1}\boldsymbol{z}\cdot\boldsymbol{z}^{\mathrm{T}}\theta)^2]/n \\
&\leq s/n, \qquad (12)
\end{aligned}$$

where the second inequality is due to $\|\theta\|_2 \leq L$ and $\|\boldsymbol{y_i}\|_2 \leq 1$, and the third equality is due to the definition of A.

Since $\mathrm{E}[\boldsymbol{Y_i}] = 0$, the $\boldsymbol{Y_i}$'s are independent, and also we have (11) and (12), applying Proposition 5, it holds that

$$\Pr[|\boldsymbol{Y}| > \epsilon] \leq \exp\left(-\frac{\epsilon^2/2}{s/n + \frac{1}{3}(|x^{\mathrm{T}}\theta|/n + L\sqrt{td/n^2})\epsilon}\right) = \exp\left(-\frac{n\epsilon^2}{2s + \frac{2}{3}(|x^{\mathrm{T}}\theta| + L\sqrt{td})\epsilon}\right).$$

By letting $\epsilon = \sqrt{\frac{2s}{\ell}} + \frac{2(|x^{\mathrm{T}}\theta| + L\sqrt{td})}{3\ell}$, it holds that

$$\Pr\left[|\boldsymbol{Y}| > \sqrt{\frac{2s}{\ell}} + \frac{2(|x^{\mathrm{T}}\theta| + L\sqrt{td})}{3\ell}\right] \leq \delta$$

for $n \geq \ell\ln\delta^{-1}$, which concludes the proof. □

Proof of Claim 9. Let t denote $\|x\|^2_{M(\lambda^*)^{-1}}$. Recall that $\boldsymbol{Z_i} = x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}\boldsymbol{\epsilon_i}$.

By Proposition 2, $\boldsymbol{Z_i}$ can be seen as a $(\kappa\boldsymbol{\sigma_i})$-subgaussian random variable where $\boldsymbol{\sigma_i} = \sqrt{x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}\boldsymbol{y_i}^{\mathrm{T}}A^{-1}x}$. Again by Proposition 2, $\boldsymbol{Z}$ is a $(\kappa\boldsymbol{\sigma})$-subgaussian random variable where $\boldsymbol{\sigma} = \sqrt{\sum_{i=1}^n \boldsymbol{\sigma_i^2}}$.

For a given ϵ and a realization of $\boldsymbol{\sigma}^2$, by Proposition 3 we know $\Pr[|\boldsymbol{Z}| > \epsilon] \le 2\exp(-\epsilon^2/(2\kappa^2\sigma^2))$. When $\boldsymbol{\sigma}^2 \le \frac{\epsilon^2}{2\kappa^2\ln(4/\delta)}$, we have $\Pr[|\boldsymbol{Z}| > \epsilon] \le \delta/2$, which means

$$\Pr\left[|\boldsymbol{Z}| > \epsilon \,\middle|\, \boldsymbol{\sigma}^2 \le \frac{\epsilon^2}{2\kappa^2\ln(4/\delta)}\right] \le \delta/2. \tag{13}$$

Next, we give a bound on $\Pr\left[\boldsymbol{\sigma^2} \le \frac{\epsilon^2}{2\kappa^2\ln(4/\delta)}\right]$. Note that

$$\begin{aligned}\mathrm{E}[\boldsymbol{\sigma_i^2}] &= \mathrm{E}[x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}\boldsymbol{y_i}^{\mathrm{T}}A^{-1}x] \\ &= x^{\mathrm{T}}A^{-1}\,\mathrm{E}[\boldsymbol{y_i}\boldsymbol{y_i}^{\mathrm{T}}]A^{-1}x = t/n^2 \end{aligned}\tag{14}$$

and by (10) we have

$$\boldsymbol{\sigma_i^2} = |x^{\mathrm{T}}A^{-1}\boldsymbol{y_i}|^2 \le td/n^2. \tag{15}$$

Let $\boldsymbol{\lambda_i} = \boldsymbol{\sigma_i^2}n^2/(td)$. By (14) we have $\mathrm{E}[\boldsymbol{\lambda_i}] = 1/d$ and by (15) we have $\boldsymbol{\lambda_i} \in [0,1]$. Now applying Proposition 4, for $\tau \in [t/n, 2t/n]$, we have

$$\begin{aligned}\Pr\left[\boldsymbol{\sigma^2} > \tau\right] &= \Pr\left[\sum_{i=1}^n \boldsymbol{\lambda_i} > \tau n^2/(td)\right] \\ &\le \exp\left(-\frac{(\tau n/t - 1)^2 \cdot n/d}{3}\right).\end{aligned}$$

Hence, by letting $\tau = \frac{t}{n}\sqrt{\frac{3d\ln(2/\delta)}{n}} + \frac{t}{n}$, it holds that

$$\Pr\left[\boldsymbol{\sigma^2} \le \frac{t}{n}\sqrt{\frac{3d\ln(2/\delta)}{n}} + \frac{t}{n}\right] \ge 1 - \delta/2. \tag{16}$$

Therefore, combining (13) and (16), we have $\Pr[|\boldsymbol{Z}| \le \epsilon]$

$$\begin{aligned}&\ge \Pr\left[|\boldsymbol{Z}| \le \epsilon \wedge \boldsymbol{\sigma}^2 \le \frac{\epsilon^2}{2\kappa^2\ln(4/\delta)}\right] \\ &= \Pr\left[|\boldsymbol{Z}| \le \epsilon \,\middle|\, \boldsymbol{\sigma^2} \le \frac{\epsilon^2}{2\kappa^2\ln(4/\delta)}\right]\Pr\left[\boldsymbol{\sigma^2} \le \frac{\epsilon^2}{2\kappa^2\ln(4/\delta)}\right] \\ &\ge (1-\delta/2)(1-\delta/2) \ge 1-\delta,\end{aligned}$$

whenever $\frac{\epsilon^2}{2\kappa^2\ln(4/\delta)} \ge \frac{t}{n}\sqrt{\frac{3d\ln(2/\delta)}{n}} + \frac{t}{n}$.

By letting $\epsilon = \sqrt{2}\kappa\sqrt{\frac{t}{\ell}\sqrt{\frac{3d}{\ell}} + \frac{t}{\ell}}$, we get $\Pr\left[|\boldsymbol{Z}| \le \sqrt{2}\kappa\sqrt{\frac{t}{\ell}\sqrt{\frac{3d}{\ell}} + \frac{t}{\ell}}\right] \ge 1-\delta$ for $n \ge \ell\ln(4/\delta)$, which concludes the proof of this claim. □

3.3. Error Bounds for Every Arm

We first simplify $\mathrm{err}_{\lambda^*}(x,\ell)$ and get the following lemma.

Lemma 10. *Let* $c_0 = \max\{4L^2, 3\kappa^2, 3\}$, *when* $n \ge c_0\ell\ln(5/\delta)$ *where* $\ell \ge d$, *with probability* $(1-\delta)$, *we have* $|x^{\mathrm{T}}(\theta - \widehat{\boldsymbol{\theta}})| \le \sqrt{\frac{2\|x\|_2^2 + 2\sqrt{d}\|x\|_2\|x\|_{M(\lambda^*)^{-1}} + (4+2\sqrt{d/\ell})\|x\|^2_{M(\lambda^*)^{-1}}}{\ell}}$.

Please refer to Appendix A for the proof of Lemma 10.

If we further make assumptions on $\|x\|_{M(\lambda^*)^{-1}}$, we have

Lemma 11. *When* $n \ge c_0\ell\ln(5/\delta)$ *where* $\ell \ge d$, *and* $\|x\|_{M(\lambda^*)^{-1}} \le c_1\sqrt{d}$ *where* $\|x\|_2 \le c_1$, *we have* $\Pr\left[|x^{\mathrm{T}}(\theta - \widehat{\boldsymbol{\theta}})| \le c_1\sqrt{\frac{2+6d+2d\sqrt{d/\ell}}{\ell}}\right] \ge 1-\delta$.

Note that $\|x\|_{M(\lambda^*)^{-1}} \le \sqrt{d}$ and $\|x\|_2 \le 1$ for every $x \in \mathcal{X}$ (by Proposition 6). Applying Lemma 11 for every arm in $\mathcal{X}$ and via a union bound, we have

Theorem 12. *When* $n \ge c_0 \cdot \frac{2+(6+\epsilon)d}{\epsilon^2}\ln(5N/\delta)$ *where* $\epsilon \le 3$, *we have* $\Pr\left[|x^{\mathrm{T}}\theta - x^{\mathrm{T}}\widehat{\boldsymbol{\theta}}| \le \epsilon, \forall x \in \mathcal{X}\right] \ge 1-\delta$.

Please refer to Appendix B for the proof of Theorem 12.

Next, we apply Lemma 7 for every $y \in \mathcal{Y}$ where $\mathcal{Y}$ is the set $\{x - x' | x, x' \in \mathcal{X}\}$. Via a union bound, we have

Theorem 13. *When* $n \ge \ell\ln(5N^2/(2\delta))$ *where* $\ell \ge 3d$, *with probability* $(1-\delta)$, *we have* $|y^{\mathrm{T}}\theta - y^{\mathrm{T}}\widehat{\boldsymbol{\theta}}| \le \mathrm{Err}_{\lambda^*}(y,\ell,\theta) \le \mathrm{err}_{\lambda^*}(y,\ell), \forall y \in \mathcal{Y}$.

For every $y = x - x' \in \mathcal{Y}$ (where $x, x' \in \mathcal{X}$), by Proposition 6, we have $\|y\|_{M(\lambda^*)^{-1}} = \|x - x'\|_{M(\lambda^*)^{-1}} \le \|x\|_{M(\lambda^*)^{-1}} + \|x'\|_{M(\lambda^*)^{-1}} \le 2\sqrt{d}$. Note that $\|y\|_2 \le 2$. Similar to Theorem 12, we apply Lemma 11 for every $y \in \mathcal{Y}$ and via a union bound, we have

Theorem 14. *When* $n \ge 4c_0\,\frac{2+(6+\epsilon)d}{\epsilon^2}\ln(5N^2/(2\delta))$ *where* $\epsilon \le 7$, *with probability* $(1-\delta)$, *we have* $|y^{\mathrm{T}}\theta - y^{\mathrm{T}}\widehat{\boldsymbol{\theta}}| \le \mathrm{err}_{\lambda^*}(y,\ell) \le \epsilon, \forall y \in \mathcal{Y}$, *where* $\ell = 4c_0\,\frac{2+(6+\epsilon)d}{\epsilon^2}$.

3.4. Bounds for a Significant Number of Arms

We also prove the following theorem and corollary which work better when we have a significant number of (or infinitely many) arms.

Theorem 15. *When* $n \ge c_0\ell\ln(5d/\delta)$ *where* $\ell \ge d$, *with probability* $1-\delta$, *we have* $\left|x^{\mathrm{T}}(\theta - \widehat{\boldsymbol{\theta}})\right| \le \sqrt{\frac{(2+d)d + (5d+2d\sqrt{d/\ell})\|x\|^2_{M(\lambda^*)^{-1}}}{\ell}}, \forall x \in \mathbb{R}^d : \|x\|_2 \le 1$.

Please refer to Appendix C for the proof of Theorem 15.

Corollary 16. *When* $n \ge c_0\,\frac{(4d+(6+\epsilon)d^2)}{\epsilon^2}\ln(5d/\delta)$ *where* $\epsilon \le 3\sqrt{d}$, *we have* $\Pr\left[\left|x^{\mathrm{T}}(\theta - \widehat{\boldsymbol{\theta}})\right| \le \epsilon, \forall x \in \mathcal{X}\right] \ge 1-\delta$.

Proof. Note that $\|x\|^2_{M(\lambda^*)^{-1}} \le d$ for all $x \in \mathcal{X}$. We prove the corollary by applying Theorem 15 with $\ell = (4d + (6+\epsilon)d^2)/\epsilon^2$. □

4. Algorithms

Given a set S of arms, we use λ^*_S to represent the solution to the three problems defined in Proposition 6, which can be computed by solving the optimization problem stated in (5).

4.1. The Estimator for θ

First we define a procedure, VECTOREST(S, n), that aims to estimate the underlying unknown θ given a set S of arms and n samples. VECTOREST(S, n) utilizes the estimator in (8), and its descbrived in Algorithm 1.

Algorithm 1 VECTOREST(S, n)

Input: A set S of arms, and n samples.
Let $y_1, \dots, y_n$ be the n samples acquired from S according to the distribution λ^*_S.
Pull arms $y_1, \dots, y_n$, and suppose their corresponding rewards are $r_1, \dots, r_n$ respectively.
Compute $A \leftarrow n \cdot \sum_{x \in S} \lambda^*_S(x) x x^{\mathrm{T}}$ and $b \leftarrow \sum_{i=1}^n r_i y_i$.
$\widehat{\theta} \leftarrow A^{-1} b$.
Output: The estimate $\widehat{\theta}$.

4.2. $\mathcal{X}$-dependent Algorithm

We now describe an iterative algorithm that eliminates $(|S| - p)$ suboptimal arms from S with probability at least $(1 - \delta)$. We call this algorithm ELIMTIL$_p$ and present the details in Algorithm 2. In essence, during the r-th iteration, ELIMTIL$_p$ uses VECTOREST to get an $\frac{\epsilon_r}{2}$-close estimate of $x^{\mathrm{T}}\theta$ for all $x \in S_r$ with probability at least $1 - \delta_r$, and then discards all arms whose estimated mean rewards are ϵ_r worse than that of the highest estimated mean. ELIMTIL$_p$ continues until at most p arms remain. By letting ϵ_r and δ_r decrease exponentially, we are able to keep the best arm in the set of output arms with probability at least $(1 - \delta)$.

Algorithm 2 ELIMTIL$_p(S, \delta)$

Input: Arms set S and confidence level δ.
Initialize $S_1 \leftarrow S$, $r \leftarrow 1$.
while $|S_r| > p$ **do**
  Set $\epsilon_r \leftarrow 1/2^r$, $\delta_r \leftarrow 6/\pi^2 \cdot \delta / r^2$.
  $\widehat{\theta}_r \leftarrow$ VECTOREST $\left(S, c_0 \frac{2+(6+\epsilon_r/2)d}{(\epsilon_r/2)^2} \ln \frac{5|S|}{\delta_r}\right)$.
  Select arm $a_r \leftarrow \operatorname{argmax}_{x \in S_r} x^{\mathrm{T}} \widehat{\theta}_r$.
  $S_{r+1} \leftarrow S_r \backslash \{x \in S_r | x^{\mathrm{T}} \widehat{\theta}_r < x^{\mathrm{T}}_{a_r} \widehat{\theta}_r - \epsilon_r\}$.
  $r \leftarrow r + 1$.
Output: Set S_r.

Lemma 17. *With probability at least $(1 - \delta)$, ELIMTIL$_p(S, \delta)$ satisfies the following properties: (i) the procedure outputs a set of at most p arms with the best arm included; and (ii) the sample complexity is $O\left(\frac{c_0 d}{\Delta^2_{p+1}}(\ln \delta^{-1} + \ln |S| + \ln \ln \Delta^{-1}_{p+1})\right)$.*

Please refer to Appendix D for the proof of Lemma 17.

As a consequence of Lemma 17, we get the following corollary by setting $S = \mathcal{X}$.

Corollary 18. *With probability at least $(1 - \delta)$, ELIMTIL$_1(\mathcal{X}, \delta)$ outputs the best arm, and the sample complexity is $O\left(\frac{c_0 d}{\Delta^2_2}(\ln \delta^{-1} + \ln |\mathcal{X}| + \ln \ln \Delta_2^{-1})\right)$.*

If we use the bound in Theorem 15 and Corollary 16 instead, we have the following statement.

Corollary 19. *There exists an algorithm that, with probability at least $(1 - \delta)$, outputs the best arm, and uses at most $O\left(\frac{c_0 d^2}{\Delta^2_2}(\ln \delta^{-1} + \ln d + \ln \ln \Delta_2^{-1})\right)$ samples.*

4.3. $\mathcal{Y}$-dependent Algorithm

We now present the algorithm $\mathcal{Y}$-ELIMTIL$_p$, which has the same asymptotic sample complexity guarantee as ELIMTIL$_p$, but performs more efficiently in practice. This improvement is achieved thanks to the following two ideas.

First, when comparing the mean rewards of two arms (namely x and x'), instead of checking the confidence intervals of the two arms, we turn to the confidence interval of $y^{\mathrm{T}}\theta$ (where $y = x - x'$) and check the corresponding confidence interval.

Second, by Lemma 7, at the r-th iteration, the confidence interval of $y^{\mathrm{T}}\theta$ has length $\mathrm{Err}_{\lambda^*_T}(y, \ell_r, \theta) \le \mathrm{err}_{\lambda^*_T}(y, \ell_r)$ where the latter quantity is an upper estimate of the first one and does not need the knowledge about θ in advance. At the r-th iteration, we use $\widehat{\theta}_r$ as an estimate of θ and define

$$\widehat{\mathrm{Err}}_{\lambda^*_T}(y, \ell_r) = \sqrt{\frac{2\,\mathrm{E}_{\boldsymbol{z} \sim \lambda^*_T}[(y^{\mathrm{T}} M(\lambda^*_T)^{-1} \boldsymbol{z} \cdot (|\boldsymbol{z}^{\mathrm{T}} \widehat{\theta}_r| + \epsilon_r/2))^2]}{\ell}} + \frac{2((|y^{\mathrm{T}} \widehat{\theta}_r| + \epsilon_r/2) + L\|y\|_{M(\lambda^*_T)^{-1}} \sqrt{d})}{3\ell} + \sqrt{2}\kappa \sqrt{\frac{\|y\|^2_{M(\lambda^*_T)^{-1}} \sqrt{3d/\ell}}{\ell} + \frac{\|y\|^2_{M(\lambda^*_T)^{-1}}}{\ell}}.$$

We will be able to show that $\mathrm{Err}_{\lambda^*_T}(y, \ell_r, \theta) \le \widehat{\mathrm{Err}}_{\lambda^*_T}(y, \ell_r)$ holds with high probability and therefore we also use $\widehat{\mathrm{Err}}_{\lambda^*_T}(y, \ell_r)$ as an upper estimate of the length of the confidence interval which in many cases proves to be tighter than $\mathrm{err}_{\lambda^*_T}(y, \ell_r)$.

We present the details of $\mathcal{Y}$-ELIMTIL$_p$ in Algorithm 3. Note that our theoretical analysis does not require $\widehat{\theta}_r$ for different values of r to be independent. Therefore, the samples may be reused across different invocations of VECTOREST.

Algorithm 3 $\mathcal{Y}$-ELIMTIL$_p(S, T, \delta)$

Input: Set S of active arms (the ones not yet eliminated), set $T \supseteq S$ of arms those can be pulled, and confidence level δ.
Initialize $S_1 \leftarrow S$, $r \leftarrow 1$.
while $|S_r| > p$ **do**
  Set $\delta_r \leftarrow 6/\pi^2 \cdot \delta/r^2$, $\ell_r \leftarrow 4c_0(2 + (6 + 4/1.1^r)d)(1.1^r/4)^2$.
  $\widehat{\theta}_r \leftarrow$ VECTOREST$(T, \ell_r \ln(5|S|^2/(2\delta_r)))$.
  Select arm $a_r \leftarrow \operatorname{argmax}_{x \in S_r} x^{\mathrm{T}} \widehat{\theta}_r$.
  $S_{r+1} \leftarrow S_r \backslash \{x \in S_r | (x_{a_r} - x)^{\mathrm{T}} \widehat{\theta}_r > \min\{\mathrm{err}_{\lambda_T^*}(x_{a_r} - x, \ell_r), \widehat{\mathrm{Err}}_{\lambda_T^*}(x_{a_r} - x, \ell_r)\}$.
  $r \leftarrow r + 1$.
Output: Set S_r.

Lemma 20. *With probability at least* $(1 - \delta)$, $\mathcal{Y}$-ELIMTIL$_p(S, T, \delta)$ *satisfies the following properties: (i) the procedure outputs a set of at most* p *arms in* S *with the best arm included; and (ii) the sample complexity is* $O\left(\frac{c_0 d}{\Delta_{p+1}^2}(\ln \delta^{-1} + \ln |S| + \ln \ln \Delta_{p+1}^{-1})\right)$.

Please refer to Appendix E for the proof of Lemma 20.

Lemma 20 gives the following corollary by setting S and T to $\mathcal{X}$.

Corollary 21. *With probability at least* $(1 - \delta)$, $\mathcal{Y}$-ELIMTIL$_1(\mathcal{X}, \mathcal{X}, \delta)$ *outputs the best arm, and the sample complexity is* $O\left(\frac{c_0 d}{\Delta_2^2}(\ln \delta^{-1} + \ln |\mathcal{X}| + \ln \ln \Delta_2^{-1})\right)$.

4.4. Fully Data-dependent Algorithm

Using the $\mathcal{Y}$-ELIMTIL algorithm as a subroutine, we introduce an algorithm that outputs the best arm with probability at least $(1 - \delta)$ and whose sample complexity depends on $\Delta_1, \ldots, \Delta_d$. To achieve this goal, the algorithm runs in rounds. During each round r, it invokes $\mathcal{Y}$-ELIMTIL to identify the top-$(d/2^r)$ arms and discards the remaining arms. We call this algorithm ALBA and include the details in Algorithm 4.

Algorithm 4 **A**daptive **L**inear **B**est **A**rm, ALBA$(\mathcal{X}, \delta)$

Input: Arms set $\mathcal{X}$ and confidence level δ.
Initialize $S_0 \leftarrow \mathcal{X}$.
for $r \leftarrow 0$ **to** $\lfloor \log_2 d \rfloor$ **do**
  Set $\delta_r \leftarrow 6/\pi^2 \cdot \delta/((r+1)^2)$.
  $S_{r+1} \leftarrow \mathcal{Y}$-ELIMTIL$_{\lfloor d/2^r \rfloor}(S_r, \mathcal{X} \cap \mathrm{span}(S_r), \delta_r)$.[3]
  $r \leftarrow r + 1$.
Output: Best arm in $\mathcal{X}$.

Theorem 22. *With probability at least* $(1 - \delta)$, *the following are true: (i)* ALBA$(\mathcal{X}, \delta)$ *outputs the best arm; and (ii) the sample complexity is* $O\left(\sum_{i=2}^{d} \frac{c_0}{\Delta_i^2}(\ln \delta^{-1} + \ln |\mathcal{X}| + \ln \ln \Delta_i^{-1})\right)$.

Please refer to Appendix F for the proof of Theorem 22.

5. Experiments

We test our algorithms $\mathcal{Y}$-ELIMTIL$_1(\mathcal{X}, \mathcal{X}, \delta)$ and ALBA$(\mathcal{X}, \delta)$, and compare them with the staet-of-the-art algorithms $\mathcal{XY}$-Allocation (Figure 2 of (Soare et al., 2014)) and $\mathcal{XY}$-Adaptive (Figure 3 of (Soare et al., 2014)). We do not include the Explore-Verify algorithm in (Karnin, 2016) as its main contribution is considered to be theoretical.

In the implementation of our algorithms, we use the *entropic mirror descent* method introduced in (Beck & Teboulle, 2003) to compute the optimal solution $\lambda_{\mathcal{X}}^*$ defined in Proposition 6(a) (see Appendix G for details). In the $\mathcal{XY}$-Adaptive algorithm, we set $\alpha = 1/10$ following the choice made in (Soare et al., 2014). All algorithms are implemented in Python 3, and are tested without parallelization.

We test the algorithms using both synthetic (similar to that in (Soare et al., 2014), and random data) and real-world data. We fix the confidence parameter $\delta = 0.05$, and report the total number of samples and time used by each algorithm and their empirical failure probabilities (i.e. to fail to identify the best arm). For each setting, the reported numbers are averaged over 100 runs.

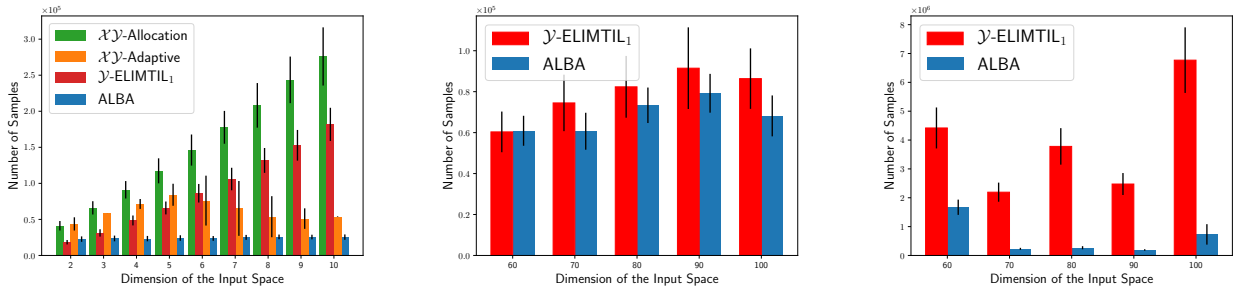
5.1. Synthetic Data Set 1

In this experiment, we consider a set $\mathcal{X} \subset \mathbb{R}^d$ of arms, with $|\mathcal{X}| = d + 1$. $\mathcal{X}$ includes the canonical basis $\{e_1, \ldots, e_d\}$ of $\mathbb{R}^d$, and an additional arm $x_{d+1} = [\cos(\omega), \sin(\omega), 0, \ldots, 0]^{\mathrm{T}}$. We choose $\theta = [2, 0, \ldots, 0]^{\mathrm{T}}$, and fix $\omega = 0.1$, so that x_1 is the best arm and $\Delta_2 = (x_1 - x_{d+1})^{\mathrm{T}}\theta$ is the minimum reward gap. Also, we set the noise $\epsilon \sim \mathcal{N}(0, 1)$ independently for each pull. We test the algorithms for $d = 2, \ldots, 10$, and report the number of samples used in Figure 1(a). In Table 1, we summarize the runtime needed by the algorithms, from which we can see the improvement made by our algorithms. The empirical failure probability for each of the algorithms is 0.

5.2. Synthetic Data Set 2: Random Vectors

In this data set, the feature vectors of the $|\mathcal{X}| = 100$ arms are independently uniformly random sampled from $\mathbb{S}^{d-1}$, the unit sphere centered at the origin. We pick the two closest arms $x, y \in \mathcal{X}$ and set $\theta = x + \alpha \cdot (y - x)$ with

[3] Changing this step to $S_{r+1} \leftarrow$ ELIMTIL$_{\lfloor d/2^r \rfloor}(S_r, \delta_r)$ does not make a difference in the theoretical guarantee of the algorithm (i.e., Theorem 22). However, $\mathcal{Y}$-ELIMTIL leads to better empirical performance.



(a) Synthetic Data Set 1 (b) Synthetic Data Set 2 (c) Real-world Data Set

Figure 1: Number of samples needed to identify the best arm for each data set. Half of each vertical black line represents the standard deviation. Notice that the sample complexity may not be monotonically increasing on d. For (c), the Δ_2 corresponding to $d = 60$ is much smaller than the other cases (see Table 3), and so it may need more samples.

$\alpha = 0.01$ so that x becomes the best arm. We also set the noise $\epsilon \sim \mathcal{N}(0, 1)$ independently for each pull. We report the samples used by our algorithms in Figure 1(b) and more information (e.g. runtime and Δ_2 of each data point) in Table 2. The empirical failure probabilities for both algorithms are 0. No results are reported for $\mathcal{XY}$-Allocation and $\mathcal{XY}$-Adaptive as both failed to terminate within one hour.

5.3. Real-world Data Set

We now work with candidate arms generated from the advertisement placement dataset provided in (Lefortier et al., 2016). Given dimensional parameter d, we select d features with the highest variance and then randomly pick $|\mathcal{X}| = 200$ identical vectors from the data set and normalize every vector to unit vectors. We also pick the two closest arms $x, y \in \mathcal{X}$ and set $\theta = x + \alpha \cdot (y - x)$ with $\alpha = 0.01$ so that x becomes the best arm. We also set the noise $\epsilon \sim \mathcal{N}(0, 1)$ independently for each pull. We now report the samples used by our algorithms in Figure 1(c) and more information (e.g. runtime and Δ_2 of each data point) in Table 3. The empirical failure probabilities for both algorithms are 0. Again, no results are shown for $\mathcal{XY}$-Allocation and $\mathcal{XY}$-Adaptive as they failed to terminate within one hour.

6. Conclusion

Via exploiting the global linear structure of the problem, we have shown that the sample complexity of identifying the best arm in linear bandit only depends on the reward gaps of the top d arms (where d is the dimension and up to a logarithmic factor). The experimental results also demonstrate the substantial improvement made by our algorithms in terms of both sample complexity and computational time.

However, it remains open to design algorithms with *instance-wise optimal* sample complexity. In (Soare et al., 2014), the authors proposed $\rho^* = \rho^*(Y^*) = \min_{\lambda \in \Delta_{\mathcal{X}}} \max_{y \in Y^*} \frac{y^\top \mathrm{E}_{\boldsymbol{z} \sim \lambda}[\boldsymbol{z}\boldsymbol{z}^\top]^{-1} y}{(y^\top \theta)^2}$ (where $Y^* = \{x - \mathcal{X}_{[1]} \mid x \neq \mathcal{X}_{[1]}\}$) as an information-theoretic lower bound on every input instance. It is a very interesting question to explore whether this lower bound is achievable by an algorithm, or some stronger lower bound exists.

Table 1: The average runtime (in seconds) needed for Synthetic Data Set 1.

d	$\mathcal{XY}$-ALLOC.	$\mathcal{XY}$-ADAPT.	$\mathcal{Y}$-ELIMTIL$_1$	ALBA
2	11.3	6.2	0.2	0.3
3	35.7	12.2	0.2	0.3
4	56.3	19.6	0.2	0.3
5	122.8	26.4	0.2	0.3
6	229.0	22.7	0.2	0.3
7	699.0	18.9	0.2	0.3
8	1015.3	27.5	0.2	0.3
9	1391.7	37.2	0.2	0.3
10	1869.0	44.4	0.2	0.3

Table 2: Experimental results for Synthetic Data Set 2.

d	Δ_2	#samples used / $d \cdot \Delta_2^{-2}$		Runtime (secs)	
		$\mathcal{Y}$-ELIMTIL$_1$	ALBA	$\mathcal{Y}$-ELIMTIL$_1$	ALBA
60	0.52	268.56	270.99	5.2	5.6
70	0.55	318.94	259.73	3.8	4.6
80	0.57	340.17	302.95	4.6	5.3
90	0.59	353.94	306.51	4.1	4.6
100	0.67	383.91	302.99	2.4	3.7

Table 3: Experimental results for the Real-world Data Set.

d	Δ_2	#samples used / $d \cdot \Delta_2^{-2}$		Runtime (secs)	
	($\times 10^{-3}$)	$\mathcal{Y}$-ELIMTIL$_1$	ALBA	$\mathcal{Y}$-ELIMTIL$_1$	ALBA
60	3.386	0.84	0.32	18.6	22.7
70	9.850	3.04	0.32	43.8	46.8
80	7.567	2.71	0.19	20.0	22.9
90	9.655	2.56	0.20	39.4	42.1
100	9.850	6.57	0.71	20.0	21.0

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pp. 2312–2320, 2011.

Allen-Zhu, Z., Li, Y., Singh, A., and Wang, Y. Near-optimal discrete optimization for experimental design: A regret minimization approach. *arXiv preprint arXiv:1711.05174*, 2017.

Auer, P. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Beck, A. and Teboulle, M. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

Bian, A. A., Buhmann, J. M., Krause, A., and Tschiatschek, S. Guarantees for greedy maximization of non-submodular functions with applications. In *Proceedings of International Conference of Machine Learning (ICML)*, 2017.

Bubeck, S., Wang, T., and Viswanathan, N. Multiple identifications in multi-armed bandits. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 258–265, 2013.

Carpentier, A. and Locatelli, A. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Conference on Learning Theory*, pp. 590–604, 2016.

Chamon, L. and Ribeiro, A. Approximate supermodularity bounds for experimental design. In *Advances in Neural Information Processing Systems*, pp. 5407–5416, 2017a.

Chamon, L. F. and Ribeiro, A. Greedy sampling of graph signals. *arXiv preprint arXiv:1704.01223*, 2017b.

Chen, J., Chen, X., Zhang, Q., and Zhou, Y. Adaptive multiple-arm identification. In *Proceedings of International Conference on Machine Learning (ICML)*, 2017a.

Chen, L., Li, J., and Qiao, M. Towards instance optimal bounds for best arm identification. In *Proceedings of the Conference on Learning Theory (COLT)*, 2017b.

Chen, S., Lin, T., King, I., Lyu, M. R., and Chen, W. Combinatorial pure exploration of multi-armed bandits. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, 2014.

Chu, W., Li, L., Reyzin, L., and Schapire, R. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 208–214, 2011.

Çivril, A. and Magdon-Ismail, M. On selecting a maximum volume sub-matrix of a matrix and related problems. *Theoretical Computer Science*, 410(47-49):4801–4811, 2009.

Dani, V., Hayes, T. P., and Kakade, S. M. Stochastic linear optimization under bandit feedback. In *Proceedings of the Conference on Learning Theory (COLT)*, 2008.

Even-Dar, E., Mannor, S., and Mansour, Y. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7:1079–1105, 2006.

Filippi, S., Cappe, O., Garivier, A., and Szepesvári, C. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems*, pp. 586–594, 2010.

Jamieson, K., Malloy, M., Nowak, R., and Bubeck, S. lil' UCB : An optimal exploration algorithm for multi-armed bandits. In *Proceedings of the Conference on Learning Theory (COLT)*, 2014.

Kalyanakrishnan, S. and Stone, P. Efficient selection of multiple bandit arms: Theory and practice. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pp. 511–518, 2010.

Karnin, Z., Koren, T., and Somekh, O. Almost optimal exploration in multi-armed bandits. In *Proceedings of International Conference on Machine Learning (ICML)*, 2013.

Karnin, Z. S. Verification based solution for structured MAB problems. In *Advances in Neural Information Processing Systems*, pp. 145–153, 2016.

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

Kiefer, J. and Wolfowitz, J. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12 (5):363–365, 1960.

Lefortier, D., Swaminathan, A., Gu, X., Joachims, T., and de Rijke, M. Large-scale validation of counterfactual learning methods: A test-bed. In *NIPS Workshop on Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems*, 2016.

Li, L., Chu, W., Langford, J., and Schapire, R. E. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web (WWW)*, pp. 661–670, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-799-8. doi: 10.1145/1772690.1772758.

Li, L., Lu, Y., and Zhou, D. Provable optimal algorithms for generalized linear contextual bandits. In *Proceedings of International Conference of Machine Learning (ICML)*, 2017.

Pukelsheim, F. *Optimal design of experiments*. SIAM, 2006.

Rusmevichientong, P. and Tsitsiklis, J. N. Linearly parameterized bandits. *Mathematics of Operations Research*, 35 (2):395–411, 2010.

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

Stroock, D. W. *Probability theory: an analytic view*. Cambridge University Press, Cambridge, second edition, 2011. ISBN 978-0-521-13250-3.

Xu, L., Honda, J., and Sugiyama, M. A fully adaptive algorithm for pure exploration in linear bandits. In Storkey, A. and Perez-Cruz, F. (eds.), *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pp. 843–851, Playa Blanca, Lanzarote, Canary Islands, 09–11 Apr 2018. PMLR.

Zhou, Y., Chen, X., and Li, J. Optimal PAC multiple arm identification with applications to crowdsourcing. In *Proceedings of International Conference on Machine Learning (ICML)*, pp. 217–225, 2014.