Robust and private stochastic linear bandits

Vasileios Charisopoulos [1] **Hossein Esfandiari** [2] **Vahab Mirrokni** [2]

Abstract

In this paper, we study the stochastic linear bandit problem under the additional requirements of *differential privacy*, *robustness* and *batched observations*. In particular, we assume an adversary randomly chooses a constant fraction of the observed rewards in each batch, replacing them with arbitrary numbers. We present differentially private and robust variants of the arm elimination algorithm using logarithmic batch queries under two privacy models and provide regret bounds in both settings. In the first model, every reward in each round is reported by a potentially different client, which reduces to standard local differential privacy (LDP). In the second model, every action is "owned" by a different client, who may aggregate the rewards over multiple queries and privatize the aggregate response instead. To the best of our knowledge, our algorithms are the first simultaneously providing differential privacy and adversarial robustness in the stochastic linear bandits problem.

1. Introduction

Bandits model is a popular formulation for online learning, wherein a learner interacts with her environment by choosing a sequence of actions, each of which presents a *reward* to the learner, from an available (potentially infinite) set of actions. The goal of the learner is to minimize her *regret*, defined as the difference between the rewards obtained by the chosen sequence of actions and the best possible action in hindsight. To achieve this, the learner must balance between *exploration* (choosing actions that reveal information about the action set) and *exploitation* (repeating actions that offered the highest rewards in previous rounds).

In theory, deciding the next action sequentially is easiest. However, there are several obstacles to overcome when it comes to practice. The first obstacle is that the rewards in bandit algorithms are often the result of interactions with physical entities (Bouneffouf et al., 2020) (e.g., recommendation systems, clinical trials, advertising, etc.), raising concerns about the privacy of participating entities. For example, responses of an individual to medical treatments can inadvertently reveal privacy-sensitive health information. Therefore, it is essential to design learning algorithms that preserve the privacy of reward sequences.

Furthermore, observations collected from multiple users or external resources are prone to failures or corruptions. These corruptions are modeled by *adversaries*, which can tamper with a fraction of the observed rewards. Adversarial corruptions can be strategic, (e.g., simultaneously hijacking the devices of multiple users), or random (such as misclicks in the context of an ad campaign). Regardless of their nature, they highlight the need for developing *robust* learning algorithms that succeed in the presence of such corruptions. Developing robust private policies has drawn considerable attention in the past couple of years ((Esfandiari et al., 2022; Liu et al., 2021; Kothari et al., 2022; Ghazi et al., 2021; Dimitrakakis et al., 2014; Li et al., 2022b)). However, despite the importance of the bandits model, we are not aware of any provably robust and private policy for this model.

Lastly, in practice, it is often desirable or even necessary for the learner to perform actions in parallel. For example, ad campaigns present an assortment of advertisements to multiple users at the same time and are only periodically recalibrated (Bertsimas & Mersereau, 2007). Consequently, batch policies must optimally balance between parallelization, which can offer significant time savings, and information exchange, which must happen frequently enough to allow for exploration of the action space (Esfandiari et al., 2021).

In this paper, we develop a learning policy that addresses both privacy and robustness challenges, while enjoying the benefits of parallelization. Specifically, our policy protects the privacy of reward sequences by respecting the standard differential privacy measure, while withstanding an adversary that changes a constant fraction of the observed rewards in each batch. In the remainder of this section, we formally

[1]Operations Research & Information Engineering, Cornell University. Part of this work was completed while the author was with Google. [2]Google Research. Correspondence to: Vasileios Charisopoulos <vc333@cornell.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

introduce the problem and survey related work in the bandit literature.

1.1. Problem formulation and provable guarantees

We study the **stochastic linear bandit** problem: given an action space $\mathcal{A} \subset \mathbb{R}^d$ with K elements satisfying $\max_{a\in\mathcal{A}} \|a\|_2 \le 1$, a learner "plays" actions $a \in \mathcal{A}$ and receives rewards

$$r_a := \langle a, \theta^\star \rangle + \eta, \quad \eta \sim \mathrm{SubG}(1), \tag{1}$$

where $\theta^\star$ is an unknown vector in $\mathbb{R}^d$ and $\mathrm{SubG}(1)$ denotes a zero-mean subgaussian random variable. Our assumption that $|\mathcal{A}| \le K$ is without loss of generality, since our results extend to the infinite case by a standard covering argument (Lattimore & Szepesvári, 2020, Chapter 20). For simplicity, we also assume that $\|\theta^\star\| \le 1$. Given a budget of T total actions, the goal of the learner is to minimize her *expected regret*:

$$\mathbb{E}\left[R_T\right] := \max_{a\in\mathcal{A}} \sum_{t=1}^{T} \langle a - a_t, \theta^\star \rangle \tag{2}$$

Batched observations. In bandits problems with batch policies, the learner commits to a sequence (i.e., a *batch*) of actions and observes the rewards of the actions *only after the entire batch of actions has been played*. The learner may play multiple batches of actions, whose sizes may be chosen adaptively, subject to the requirement that the total number of batches does not exceed B (in addition to the total number of actions played not exceeding the budget T). We assume that B is also known to the learner.

Robustness. We require that our algorithm is robust under possibly adversarial corruptions suffered. In particular, we assume that an adversary replaces every observation by an arbitrary number with some small probability α. Thus, during each batch, the *observed rewards* will satisfy

$$r_i = \begin{cases} \langle a_i, \theta^\star \rangle + \eta_i, & \text{w.p. } 1-\alpha, \\ *, & \text{w.p. } \alpha \end{cases}, \quad i = 1, \dots, n, \tag{3}$$

where $*$ is an arbitrary value, $\alpha \in [0, 1/4)$ is the corruption probability, and n is the size of the batch.

Diffential privacy. Our other requirement is that the algorithm is *differentially private* (DP).

Definition 1.1 (Differential Privacy for Bandits (Basu et al., 2019))**.** A randomized mechanism $\mathcal{M}$ for stochastic linear bandits is called $(\varepsilon_{\mathsf{priv}}, \delta_{\mathsf{priv}})$-differentially private if, for any two neighboring sequences of rewards $\mathcal{R} = (r_1, \dots, r_T)$ and $\mathcal{R}' = (r'_1, \dots, r'_T)$ where $r_i \neq r'_i$ for at most one index i, and any subset of outputs $O \in \mathcal{M}^T$, it satisfies

$$\mathbb{P}\left(\mathcal{M}(\mathcal{R}) \in O\right) \le e^{\varepsilon_{\mathsf{priv}}} \cdot \mathbb{P}\left(\mathcal{M}(\mathcal{R}') \in O\right) + \delta_{\mathsf{priv}}. \tag{4}$$

The main contribution of our paper is a batched arm elimination algorithm that satisfies both desiderata, presented in detail in Section 2. We assume a *distributed* setting where a central server takes on the role of the learner, connected with several clients that report back rewards. The clients do not trust the central server and therefore choose to privatize their reward sequences; this model is better known as *local differential privacy* (LDP) (Kasiviswanathan et al., 2011). Our algorithm addresses the following client response models:

- **(M1)** Each reward $\bar{r}_i$ may be solicited from a different client i.
- **(M2)** Each client "owns" an action $a \in \mathcal{A}$ and may report multiple rewards in each batch.

Remark 1.2. In Model **(M1)**, we may assume without loss of generality that every reward $\bar{r}_i$ is solicited from a different client, and thus every client returns at most 1 response.

Below, we provide informal statements for the expected regret that our algorithms achieves under each model. While the regret under **(M1)** has better dependence on the dimension d, **(M2)** leads to a better dependence on the privacy parameter $\varepsilon_{\mathsf{priv}}$. The improved dependence on $\varepsilon_{\mathsf{priv}}$ in the latter should not come as a complete surprise, since model **(M2)** can be viewed as interpolating between the *local* and *central* models of differential privacy. For simplicity, we focus on the case where B scales logarithmically in T, although our analysis can be easily modified for general B. Figure 1 illustrates the qualitative behavior of our regret bounds.

Theorem 1.3 (Informal)**.** *Under Model **(M1)**, there is an $\varepsilon_{\mathsf{priv}}$-locally differentially private algorithm that is robust to adversarial corruptions with expected regret satisfying*

$$\mathbb{E}\left[R_T\right] = \tilde{O}\left(\left[\sqrt{dT} + T\max\left\{\sqrt{\alpha d}, \alpha d\right\}\right]\left(1 + \tfrac{1}{\varepsilon_{\mathsf{priv}}}\right)\right)$$

It is worth noting that in the non-private setting, the regret bound above scales as $\tilde{O}(T\sqrt{\alpha d} + \sqrt{dT})$ when $\alpha < 1/d$ and $\tilde{O}(T\alpha d + \sqrt{dT})$ when $\alpha \ge 1/d$. Note that the total amount of corruption injected by the adversary is upper bounded by $C = \alpha T$. Interestingly, our result shaves off a factor of at least $\sqrt{d}$ compared to the regret bound of the best previous work on robust stochastic linear bandits (Bogunovic et al., 2021), which scales as $\tilde{O}(\sqrt{dT} + Cd^{3/2})$.

Theorem 1.4 (Informal)**.** *Under Model **(M2)**, there is an $\varepsilon_{\mathsf{priv}}$-differentially private algorithm that is robust to adversarial corruptions with expected regret satisfying*

$$\mathbb{E}\left[R_T\right] = \tilde{O}\left(d\sqrt{T} + \frac{d}{\varepsilon_{\mathsf{priv}}}\right) + \tilde{O}\left(d^{3/2}\sqrt{\alpha}\left(T + d\sqrt{T} + \frac{d}{\varepsilon_{\mathsf{priv}}}\right)\right) + \alpha T.$$

Compared to Theorem 1.3, Theorem 1.4 yields an improved dependence on T in the "private" part of the regret at the expense of an additional $\sqrt{d}$ factor in the non-private part.

1.2. Related work

In this section, we survey related work in the bandit literature that addresses differential privacy and/or robustness to corruptions. We note that, to the best of our knowledge, our work is the first to simultaneously provide robustness and differential privacy guarantees for the stochastic linear bandit setting. While preparing the camera-ready version of this manuscript we were made aware of the work of (Wu et al., 2023), which studies private and robust *multi-armed* bandits.

Differential privacy in linear bandits. Differential privacy has been well-studied in the context of bandit learning. In the central DP model, which is the focus of this paper, (Shariff & Sheffet, 2018) proved a lower bound of $\Omega(\sqrt{T} + \frac{\log(T)}{\varepsilon_{\mathsf{priv}}})$ on the expected regret and proposed a private variant of the `LinUCB` algorithm with additive noise that achieves expected regret of $\tilde{O}(\sqrt{T} + \sqrt{T}/\varepsilon_{\mathsf{priv}})$. In recent work, (Li et al., 2022a; Hanna et al., 2022) proposed a private variant of the arm elimination algorithm that obtains a regret bound of $O(\sqrt{T \log T} + \frac{\log^2(T)}{\varepsilon_{\mathsf{priv}}})$ which is tight up to logarithmic factors; in particular, the work of (Li et al., 2022a) achieves (ϵ, δ)-differential privacy using the Gaussian mechanism while (Hanna et al., 2022) achieve ϵ-differential privacy (also known as *pure* differential privacy) via the Laplace mechanism. While conceptually similar to that of (Li et al., 2022a; Hanna et al., 2022), our algorithm guarantees differential privacy and robustness to corrupted observations simultaneously and maintains an order-optimal regret bound.

In the local DP model, (Zheng et al., 2020) used a reduction to private bandit convex optimization to achieve expected regret $\tilde{O}(T^{3/4}/\varepsilon_{\mathsf{priv}})$. Under additional distributional assumptions on the action set, this was improved to $\tilde{O}(T^{1/2}/\varepsilon_{\mathsf{priv}})$ by (Han et al., 2021). The same rate was obtained by (Hanna et al., 2022), who removed the requirement that actions are generated from a distribution. Finally, a recent line of work focused on so-called *shuffle differential privacy* (Bittau et al., 2017; Cheu, 2021), wherein a trusted *shuffler* can preprocess client responses before transmitting them to the central server. A sequence of works (Tenenbaum et al., 2021; Chowdhury & Zhou, 2022; Garcelon et al., 2022; Hanna et al., 2022) proposed shuffle-DP algorithms for linear bandits, with (Li et al., 2022a; Hanna et al., 2022) achieving essentially the same regret bound as in the central DP setting.

Robustness to adversarial attacks. Recent work proposed various adversarial attacks in the bandit setting, as well as algorithms to protect against them. (Lykouris et al., 2018) (and (Gupta et al., 2019) in a follow-up work) study multi-armed bandits with *adversarial scaling*, wherein an adversary can shrink the means of the arm distributions in each round, and propose robust algorithms for this setting. The corruption in this work differs from our setting, where the adversary can replace a random fraction of rewards arbitrarily. The works of (Jun et al., 2018; Liu & Shroff, 2019; Garcelon et al., 2020) study multi-armed and contextual bandit algorithms from the attacker perspective, demonstrating how an adversary can induce linear regret with logarithmic effort.

(Li et al., 2019) and (Bogunovic et al., 2021) study additive adversarial corruptions in contextual bandits. In particular, they assume that the observed reward in round i suffers an additive perturbation by $c_i(a_i)$, where a_i is the i^{th} context and $c_i : \mathcal{A} \to [-1, 1]$ is a context-dependent corruption function. Crucially, the adversary is subject to a budget constraint given some budget C unknown to the learner:

$$\sum_{i=1}^{T} \max_{a \in \mathcal{A}} |c_i(a)| \leq C. \tag{5}$$

In (Li et al., 2019), the authors present a robust exploration algorithm for contextual bandits using the Löewner ellipsoid. Letting Δ denote the gap between the highest and lowest expected rewards, their algorithm achieves a regret of $O\big(\frac{d^{5/2} C \log T}{\Delta} + \frac{d^6 \log^2 T}{\Delta^2}\big)$, under the key assumption that the action space $\mathcal{A}$ is a full-dimensional polytope, and requires no knowledge of the corruption budget C.

On the other hand, the work of (Bogunovic et al., 2021) introduces a robust variant of the phased arm elimination algorithm for stochastic linear bandits that achieves an expected regret of $\tilde{O}(\sqrt{dT} + Cd^{3/2})$, assuming the budget C is known to the learner; for unknown budgets, an additional C^2 factor appears in the regret bound. Our work deviates from that of (Bogunovic et al., 2021) in the sense that we measure corruption using the probability α of an adversary interfering with each observation; moreover, assuming that C scales as αT, our work shaves off a $\sqrt{d}$ factor from the result of (Bogunovic et al., 2021) in certain regimes, while it also ensures differential privacy.

1.3. Notation

We let $\langle x, y \rangle := x^{\mathsf{T}} y$ denote the Euclidean inner product with induced norm $\|x\| = \sqrt{\langle x, x \rangle}$ and write $\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d \mid \|x\|_2 = 1\}$ for the unit sphere in d dimensions. When M is a positive-definite matrix, we write $\|x\|_M := \sqrt{\langle x, Mx \rangle}$ for the norm induced by M. Finally, we write

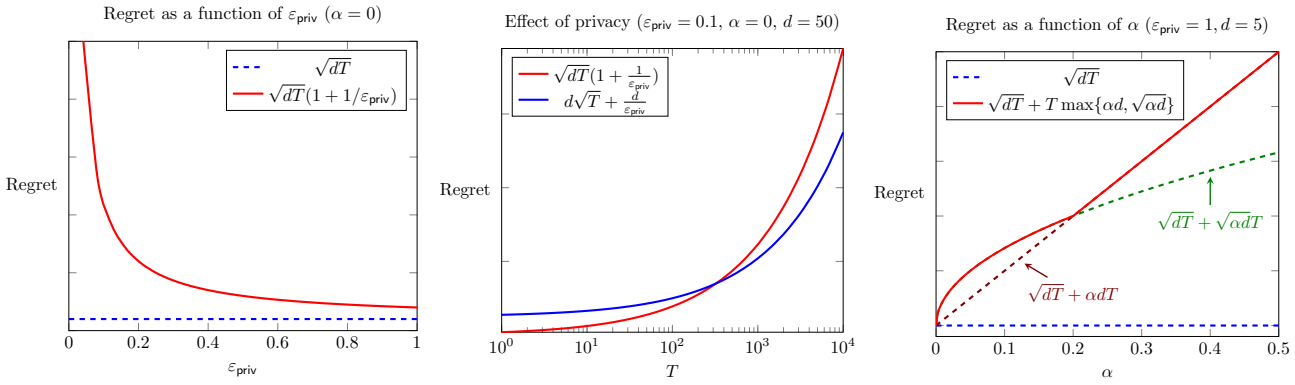

Figure 1. Demonstration of regret bounds. **Left**: private vs. non-private regret bounds under (M1). **Center**: scaling of private regret bound under (M1) and (M2). **Right**: effect of corruption parameter α under model (3).

$\|A\|_{\text{op}} := \sup_{x \in \mathbb{S}^{n-1}} \|Ax\|$ for the $\ell_2 \to \ell_2$ operator norm of a matrix $A \in \mathbb{R}^{m \times n}$.

1.4. Coresets and G-optimal designs

Our algorithms make use of *coresets*, which in turn are formed with the help of a concept called *G-optimal design*. We formally define this concept below.

Definition 1.5 (G-optimal design)**.** Let $\mathcal{A} \subset \mathbb{R}^d$ be a finite set of vectors and let $\pi : \mathcal{A} \to [0, 1]$ be a probability distribution on $\mathcal{A}$ satisfying $\sum_{a \in \mathcal{A}} \pi(a) = 1$. Then π is called a *G-optimal design* for $\mathcal{A}$ if it solves the following optimization problem:

$$\text{minimize} \quad \left\{ \max_{a \in \mathcal{A}} \|a\|^2_{M^{-1}(\pi)} \right\}, \tag{6}$$

where $M^{-1}(\pi) := \left(\sum_{a \in \mathcal{A}} \pi(a) a a^\mathsf{T} \right)^{-1}$.

A standard result in experiment design (Lattimore & Szepesvári, 2020, Theorem 21.1) shows that the optimal value of (6) is equal to d. Moreover, it is possible to find a probability distribution π satisfying the following:

Definition 1.6 (Approximate G-optimal design)**.** Let $\mathcal{A} \subset \mathbb{R}^d$ be a finite set of vectors and let $\pi : \mathcal{A} \to [0, 1]$ be a probability distribution on $\mathcal{A}$. We call π an *approximate G-optimal design* for $\mathcal{A}$ if it satisfies

$$\max_{a \in \mathcal{A}} \|a\|^2_{M^{-1}(\pi)} \le 2d, \quad |\text{supp}(\pi)| \le Cd \log\log d \tag{7}$$

for a universal constant $C > 0$.

In particular, an approximate G-optimal design π in the sense of Definition 1.6 can be found in time $O(d \log\log d)$.

Given a (approximate) G-optimal design π in the sense of Definition 1.5 or Definition 1.6, a *coreset* $\mathcal{S}_\mathcal{A}$ of total size n is a multiset $\{a_1, \ldots, a_n\}$ where each action $a \in \text{supp}(\pi)$ appears a total of $n_a := \lceil \pi(a) \cdot n \rceil$ times.

2. Algorithm and main results

To minimize the regret of the learner, we use a variation of the standard arm elimination algorithm (Lattimore & Szepesvári, 2020). In this algorithm, the learner uses batches of actions to construct confidence intervals for the optimal rewards and eliminates a set of suboptimal arms in each round based on their performance on the current batch. While the vanilla arm elimination algorithm *is neither robust nor differentially private*, we develop a variant that simultaneously ensures both these properties. An additional attractive property of our algorithm is that its implementation only requires a simple modification.

2.1. Our approach

To motivate our approach, we first sketch a naive attempt at modifying the arm elimination algorithm and briefly explain why it is unable to achieve good regret guarantees.

Recall that in the standard arm elimination algorithm, the learner first forms a so-called *coreset* of the action space $\mathcal{A}$, which is a multiset of vectors $a_1, \ldots, a_n \in \mathcal{A}$, and plays all the actions a_j receiving rewards r_j. To prune the action space, the learner first computes the least-squares estimate:

$$\widehat{\theta} := \left(\sum_{j=1}^n a_j a_j^\mathsf{T} \right)^{-1} \sum_{j=1}^n a_j r_j, \tag{8}$$

and chooses a suitable threshold γ to eliminate arms with

$$\left\langle a, \widehat{\theta} \right\rangle < \max_{j=1,\ldots,n} \left\langle a_j, \widehat{\theta} \right\rangle - 2\gamma.$$

Clearly, the arm elimination algorithm interacts with the rewards directly only when forming the least squares estimate $\widehat{\theta}$. Therefore, estimating $\widehat{\theta}$ with a differentially private algorithm is sufficient to protect the privacy of rewards. Likewise, computing $\widehat{\theta}$ robustly will ensure robustness of the overall algorithm.

The main idea behind our arm elimination variant is the following. First, let us dispense with the differential privacy requirement. Notice that in the absence of corruptions, $\widehat{\theta}$ is the empirical mean of the sequence of variables $\{Z_1, \ldots, Z_n\}$:

$$Z_j := \left(\sum_{i=1}^n a_i a_i^\mathsf{T}\right)^{-1} r_j a_j.$$

To compute $\widehat{\theta}$ robustly, one may attempt to run an algorithm such as the geometric median. However, the approximation guarantee of the geometric median method scales proportionally to $\max_{j\in[n]} \|Z_j - \widehat{\theta}\|$, for which worst-case bounds are overly pessimistic. Indeed, letting M denote the Gram matrix of the coreset used in the current arm elimination round, a tedious but straightforward calculation shows that these bounds scale as $\kappa_2(M)$, the condition number of M. In turn, the latter quantity depends on the geometry of the maintained action set and is difficult to control in general. For example, even if the original action set is "well-conditioned", that property will not necessarily hold throughout the algorithm.

To work around this issue, we take advantage of the probabilistic nature of the adversary. The main idea is that, *in expectation*, the least-squares estimate computed over the subset of non-corrupted rewards, $\mathcal{I}_{\text{good}}$, and given by

$$\widehat{\theta}_{\mathcal{I}_{\text{good}}} = \left(\sum_{i=1}^n a_i a_i^\mathsf{T}\right)^{-1} \sum_{j\in\mathcal{I}_{\text{good}}} a_j r_j, \tag{9}$$

is close to the true least-squares estimate in the absence of any corruptions. While the set $\mathcal{I}_{\text{good}}$ is not known a-priori to the learner, we may still estimate $\widehat{\theta}_{\mathcal{I}_{\text{good}}}$ from Eq. (9) using a well-known spectral filtering algorithm from the robust statistics literature. In doing so, we reduce the problem of robust linear regression (with a *fixed* design matrix) to that of robust mean estimation (over an appropriately weighted set of inputs). We mention in passing that the work of(Chen et al., 2022) also develop a distribution-free algorithm for robust linear regression which applies to a more general class of problems. However, their algorithm requires repeatedly solving a semidefinite program, while our spectral filtering-based method is simpler to implement.

In what follows, we describe our robust linear regression primitive and state its theoretical approximation guarantees, and finally sketch how to take advantage of it to design a robust and diffentially private algorithm for batched bandits.

2.2. Robust linear regression with fixed designs

In this section, we describe an efficient algorithm for Huber-robust linear regression with a fixed design matrix. In particular, we let the (clean) set of observations satisfy

$$y_i = \langle a_i, \theta^\star\rangle + \eta_i, \quad i = 1, \ldots, n. \tag{10}$$

where η_i are independent noise realizations and $a_1, \ldots, a_n$ are design vectors. The least-squares estimate of $\theta^\star$ is given by

$$\widehat{\theta} := M_n^{-1} \sum_{i=1}^n y_i a_i, \qquad M_n := \sum_{i=1}^n a_i a_i^\mathsf{T}.$$

Now, suppose that an adversary corrupts each y_i independently with probability $\alpha \in (0, 1/2)$, so the learner observes

$$\hat{y}_i = \begin{cases} y_i, & \text{if } Z_i = 1, \\ *, & \text{otherwise} \end{cases}, \quad Z_i \sim \mathrm{Ber}(1-\alpha). \tag{11}$$

The goal is to estimate the least-squares solution $\widehat{\theta}$ robustly. Our strategy will be to first estimate the least-squares solution over the subset of "good" indices G_0:

$$\theta_{G_0} = \sum_{i\in G_0} M_n^{-1} a_i y_i, \quad G_0 = \{i \mid Z_i = 1\}. \tag{12}$$

To estimate θ_{G_0}, we will apply the well-known (randomized) spectral filtering algorithm for robust mean estimation (see, e.g., (Diakonikolas & Kane, 2019; Prasad et al., 2019)), provided in Algorithm 2 for completeness, to the components of the least-squares solution after an appropriate reweighting. In particular, we will estimate

$$\begin{aligned} \gamma_{\{a_i\}_{i=1}^n} &:= \frac{\max_{a\in\mathcal{A}} \|a\|^2_{M_n^{-1}} \sum_{i=1}^n y_i^2}{n}; \\ \widetilde{w} &:= \texttt{Filter}\left(\left\{M_n^{-1/2} a_i y_i\right\}_{i=1}^n, \gamma_{\{a_i\}_{i=1}^n}\right); \\ \widetilde{\theta} &:= n M_n^{-1/2} \widetilde{w} \end{aligned} \tag{13}$$

We prove the following guarantee for this method. The proof of this proposition is deferred to Appendix A. We use this proposition in the next section to design robust and differentially private algorithms for stochastic linear bandits.

Proposition 2.1. *Fix a $\delta \in (0,1)$, $a \in \mathcal{A}$ and let $e_i = y_i - \langle a_i, \theta^\star\rangle$. Then with probability at least $1 - 2\delta$, we have*

$$\begin{aligned} \left|\langle a, \widetilde{\theta} - \theta^\star\rangle\right| \lesssim\ & \max_{a\in\mathcal{A}} \|a\|^2_{M_n^{-1}} \sqrt{n \sum_{i=1}^n y_i^2} \left(\alpha + \frac{\log(1/\delta)}{n}\right)^{1/2} \\ & + \max_{a\in\mathcal{A}} \|a\|^2_{M_n^{-1}} \sqrt{\sum_{i=1}^n y_i^2} + \sqrt{\alpha \log(1/\delta)} \\ & + \sum_{i=1}^n e_i \left\langle a, M_n^{-1} a_i\right\rangle + \alpha, \end{aligned} \tag{14}$$

where $M_n := \sum_{i=1}^n a_i a_i^\mathsf{T}$.

Algorithm 1 Robust arm elimination

Input: action space $\mathcal{A}$, T, B, failure prob. δ, corruption prob. $\alpha \in (0, 1/4)$, truncation parameter $\nu > 0$.
Set $\mathcal{A}_0 := \mathcal{A}$, $q = T^{1/B}$
for $i = 1, \ldots, B-1$ **do**
  Compute approximate G-optimal design π with $|\mathrm{supp}(\pi)| \lesssim d \log\log d$.
  Form a coreset $\mathcal{S}_{\mathcal{A}_{i-1}}$ by playing each distinct $a \in \mathrm{supp}(\pi)$ a total of

$$n_a = \begin{cases} \lceil q^i \pi(a) \rceil, & \text{under Model (M1);} \\ \lceil q^i \max\{\pi(a), \nu\} \rceil, & \text{under Model (M2).} \end{cases}.$$

  Play actions $a_j \in \mathcal{S}_{\mathcal{A}_{i-1}}$ and collect rewards r_j according to (3).
  Compute $\widetilde{w}_i := \texttt{Filter}\left(\left\{M_n^{-1/2} a_i r_i\right\}_{i=1}^n, \frac{\max_{a\in\mathcal{A}} \|a\|^2_{M_n^{-1}} \sum_{i=1}^n r_i^2}{n}\right)$, where $M_n := \sum_{i=1}^n a_i a_i^\mathsf{T}$. {Alg. 2}
  Compute $\widetilde{\theta}_i := n M_n^{-1/2} \widetilde{w}$.
  Set the elimination threshold

$$\gamma_i := \begin{cases} \sqrt{d}\left(\sqrt{\log(q^i/\delta)} + \frac{\log(q^i/\delta)}{\varepsilon_{\mathsf{priv}}}\right)(\sqrt{\alpha} + \alpha\sqrt{d}) + \alpha + \sqrt{\frac{d\log(1/\delta)}{q^i}}\left(1 + \frac{\sqrt{\log(1/\delta)}}{\varepsilon_{\mathsf{priv}}}\right), & \text{under (M1);} \\ \sqrt{\frac{d\log(1/\delta)}{\nu m}}\left(1 + \frac{1}{\varepsilon_{\mathsf{priv}}}\sqrt{\frac{\log(1/\delta)}{\nu m}}\right) + 2d\left(1 + \sqrt{\frac{\log(k/\delta)}{\nu m}} + \frac{\log(k/\delta)}{\nu m \varepsilon_{\mathsf{priv}}}\right)\left(\sqrt{k\alpha} + \sqrt{\alpha\log(1/\delta)}\right) + \alpha, & \text{under (M2),} \end{cases}$$

  where $k := |\mathrm{supp}(\pi)|$ in the second option.
  Eliminate suboptimal arms:

$$\mathcal{A}_i := \left\{ a \in \mathcal{S}_{\mathcal{A}_{i-1}} \mid \langle a, \widetilde{\theta}_i \rangle \geq \max_{a' \in \mathcal{S}_{\mathcal{A}_{i-1}}} \langle a', \widetilde{\theta}_i \rangle - 2\gamma_i. \right\},$$

end for
Play the "best" action in $\mathcal{S}_{\mathcal{A}_{B-1}}$ in the last round.

Algorithm 2 $\texttt{Filter}(S := \{X_i\}_{i=1}^m, \lambda)$

Compute empirical mean and covariance:

$$\theta_S := \frac{1}{|S|}\sum_{i\in S} X_i, \ \Sigma_S := \frac{1}{|S|}\sum_{i\in S}(X_i - \theta_S)(X_i - \theta_S)^\mathsf{T}.$$

Compute leading eigenpair (μ, v) of Σ_S.
if $\mu < 4\lambda$ **then**
  return θ_S
else
  Compute outlier scores $\tau_i := \langle v, X_i - \theta_S \rangle^2$ for all i.
  Sample an element Y with $\mathbb{P}(Y = X_i) \propto \tau_i$
  return $\texttt{Filter}(S \setminus \{Y\}, \lambda)$
end if

3. Robust differentially private bandits

In this section, we consider the requirement of *differential privacy*. In particular, we assume that the learner is an *untrusted* server; every client must therefore privatize their rewards before reporting them to the learner. Recall that we consider two different models for generating client responses:

- (**M1**) Every reward is obtained from a distinct client.
- (**M2**) All rewards associated with a distinct action a are obtained from the same client.

Algorithm 1 documents the parameter choices under each of the models above. For our regret analysis, we rely on the following facts for each round i.

Fact 1: The optimal arm is not eliminated. Let $a^\star$ denote the "optimal" action in the sence of maximizing the inner products $\langle a, \theta \rangle$. Then, with high probability,

$$\begin{aligned} \langle a, \widetilde{\theta}\rangle - \langle a^\star, \widetilde{\theta}\rangle &= \langle a, \theta^\star\rangle + \langle a, \widetilde{\theta} - \theta^\star\rangle \\ &\quad - \langle a^\star, \theta^\star\rangle - \langle a^\star, \widetilde{\theta} - \theta^\star\rangle \\ &\leq \langle a - a^\star, \theta^\star\rangle + 2\gamma_i \\ &\leq 2\gamma_i, \end{aligned}$$

using the bound on the difference in the penultimate inequality and the fact that $\langle a, \theta^\star\rangle \leq \langle a^\star, \theta^\star\rangle$ in the last inequality.

Thus, $a^\star$ always satisfies the condition of the algorithm and is not eliminated.

Fact 2: Surviving arms have bounded gap. Fix an arm a and let $\Delta := \langle a^\star - a, \theta^\star \rangle$ be its gap. We have

$$\begin{aligned}\langle a^\star - a, \widetilde{\theta} \rangle &\geq \langle a^\star, \theta^\star \rangle - \gamma_i - (\langle a, \theta^\star \rangle + \gamma_i) \\ &\geq \Delta - 2\gamma_i.\end{aligned}$$

Now, let i be the smallest positive integer such that $\gamma_i < \Delta/4$. Then the above implies that

$$\langle a^\star - a, \widetilde{\theta} \rangle \geq 2\gamma_i.$$

Consequently, any arm a with gap $\Delta_a > 4\gamma_i$ for some index i will be eliminated at the end of that round. Therefore, all arms that are active at the beginning of round i will necessarily satisfy $\Delta_a \leq 4\gamma_{i-1}$.

3.1. Local differential privacy under (M1)

In this setting, we can achieve pure LDP using the Laplace mechanism (Dwork & Roth, 2014). In particular, we define

$$\mathcal{M}(r) = r + \xi, \quad \xi \sim \mathsf{Lap}\left(\frac{2}{\varepsilon_{\mathsf{priv}}}\right),$$

where $\varepsilon_{\mathsf{priv}}$ is a desired privacy parameter. Then, when queried for a response, client i reports the privatized reward:

$$\hat{r}_i = \mathcal{M}(r_i) = \langle a_i, \theta^\star \rangle + \eta_i + \xi_i, \quad \xi_i \sim \mathsf{Lap}\left(\frac{2}{\varepsilon_{\mathsf{priv}}}\right). \tag{15}$$

The three forthcoming lemmata control different terms appearing in the confidence interval from Eq. (14). Lemma 3.1 below controls the contribution of the additive noise.

Lemma 3.1. *Under the model (M1), with probability at least $1 - 2\delta$ we have*

$$\begin{aligned}&\sum_{i=1}^{n} e_i \left\langle a, M_n^{-1} a_i \right\rangle \leq \\ &\qquad \|a\|_{M_n^{-1}} \sqrt{\log(1/\delta)} \left(c_1 + \frac{c_2 \sqrt{\log(1/\delta)}}{\varepsilon_{\mathsf{priv}}} \right).\end{aligned} \tag{16}$$

Two of the three terms in Eq. (14) depend on the maximal weighted norm $\|a\|^2_{M_n^{-1}}$ over the action set; Lemma 3.2 bounds that norm for an arbitrary round of the arm elimination algorithm.

Lemma 3.2. *Under Model (M1), we have the bound:*

$$\max_{a \in \mathcal{A}} \|a\|^2_{M_n^{-1}} \leq \frac{2d}{n}. \tag{17}$$

Finally, Lemma 3.3 below controls the contribution of $\sqrt{\sum_{i=1}^n y_i^2}$ to the robust confidence interval.

Lemma 3.3. *With probability at least $1 - \delta$, we have*

$$\sqrt{\sum_{i=1}^{n} y_i^2} \lesssim \sqrt{n} \left(1 + \sqrt{\log(n/\delta)} + \frac{\log(n/\delta)}{\varepsilon_{\mathsf{priv}}} \right). \tag{18}$$

With control over the confidence interval (14) at hand, we arrive at the regret bound in Theorem 3.4 below. The proof follows standard arguments (see, e.g., (Esfandiari et al., 2021, Theorem 5.1)) and can be found in Appendix B.1.4.

Theorem 3.4. *Under Model (M1), the expected regret of Algorithm 1 is at most*

$$\begin{aligned}&\sqrt{Td\log(T/\delta)} \left(1 + \frac{\sqrt{\log(T/\delta)}}{\varepsilon_{\mathsf{priv}}} \right) \\ &+ T\sqrt{\log(T/\delta)} \max\left\{ \sqrt{\alpha d}, \alpha d \right\} \left(1 + \frac{\log(T/\delta)}{\varepsilon_{\mathsf{priv}}} \right),\end{aligned} \tag{19}$$

up to a dimension-independent multiplicative constant.

3.2. Local differential privacy under (M2)

In this setting, every client achieves differential privacy by aggregating their responses before transmitting them to the server. In particular, let n_a denote the number of times action a is played during the current round. The parameter n_a can be considered public, since it is known to the untrusted server. Then, client a may report

$$\begin{aligned}\hat{r}_a &= \mathcal{M}\left(\frac{1}{n_a} \sum_{i=1}^{n_a} \langle a, \theta^\star \rangle + \eta_i \right) \\ &= \frac{1}{n_a} \sum_{i=1}^{n_a} \langle a, \theta^\star \rangle + \eta_i + \xi_a,\end{aligned} \tag{20}$$

where $\eta_i \sim \mathrm{SubG}(1)$ and ξ_a is Laplace noise. The amount of noise needed to achieve privacy scales inversely with n_a.

Lemma 3.5. *With $\xi_a \sim \mathsf{Lap}\left(\frac{2}{n_a \varepsilon_{\mathsf{priv}}}\right)$, the mechanism $\mathcal{M}$ in (20) is $\varepsilon_{\mathsf{priv}}$-differentially private for client a.*

Recall that in this model, the arm elimination algorithm follows the modifications below:

1. We receive $|\mathrm{supp}(\pi)|$ distinct responses in each round, where π is an approximately G-optimal design.
2. Every action $a \in \mathrm{supp}(\pi)$ is played a total of $n_a = \lceil m \max\{\pi(a), \nu\} \rceil$ times for fixed m and $\nu > 0$.

Our proof for this setting is analogous to the proof under Model (M1). We have the following analogue of Lemma 3.1:

Lemma 3.6. *Under the model (M2), with probability at least* $1 - 2\delta$ *we have*

$$\sum_{v \in \mathrm{supp}(\pi)} e_v \left\langle a, M^{-1} v \right\rangle \leq \frac{\|a\|_{M^{-1}}}{\sqrt{\nu m}} \sqrt{\log(1/\delta)} \left(c_1 + \frac{c_2}{\varepsilon_{\mathsf{priv}}} \sqrt{\frac{\log(1/\delta)}{\nu m}} \right), \tag{21}$$

where $e_v = \mathcal{M}(r_v) - \langle v, \theta^\star \rangle$ *and* $M = \sum_{a \in \mathrm{supp}(\pi)} a a^\mathsf{T}$.

Lemma 3.7. *Under Model (M2), we have the bound:*

$$\max_{a \in \mathcal{A}} \|a\|^2_{M^{-1}} \leq 2d. \tag{22}$$

We also have the following analogue of Lemma 3.3.

Lemma 3.8. *With probability at least* $1 - \delta$*, we have*

$$\sqrt{\sum_{v \in \mathrm{supp}(\pi)} y_v^2} \lesssim 1 + \sqrt{\frac{\log(|\mathrm{supp}(\pi)|/\delta)}{\nu m}} + \frac{\log(|\mathrm{supp}(\pi)|/\delta)}{\nu m \varepsilon_{\mathsf{priv}}}. \tag{23}$$

Putting everything together, we arrive at Theorem 3.9 below, whose proof can be found in Appendix B.2.4.

Theorem 3.9. *Under Model (M2), the expected regret of Algorithm 1 is at most*

$$\begin{aligned} &\nu d \log\log d \left(\sqrt{\frac{dT \log(1/\delta)}{\nu}} + \frac{\log(1/\delta)\log(T)\sqrt{d}}{\varepsilon_{\mathsf{priv}}\sqrt{\nu}} \right) \\ &+ 2d \left(\sqrt{\alpha d \log\log d} + \sqrt{\alpha \log(1/\delta)} \right) \times \\ &\left(T + \sqrt{\frac{Td \log\log d/\delta}{\nu}} + \frac{\log(\frac{d \log\log d}{\delta}) \log T}{\nu \varepsilon_{\mathsf{priv}}} \right) \\ &+ \alpha T, \end{aligned} \tag{24}$$

up to a dimension-dependent multiplicative constant.

4. Conclusion

In this paper we presented a robust and $\varepsilon_{\mathsf{priv}}$-LDP policy for batched stochastic linear bandits with an expected regret

$$\mathbb{E}[R_T] = \tilde{O}\left([\sqrt{dT} + T \max\{\sqrt{\alpha d}, \alpha d\}](1 + 1/\varepsilon_{\mathsf{priv}}) \right),$$

where α is the probability of corruption of each reward, which only requires a logarithmic number of batch queries. In the absence of corruption ($\alpha = 0$), our regret matches that of the best-known non-robust differentially private algorithm (Hanna et al., 2022). On the other hand, when no differential privacy is required, our regret bounds shaves off a factor of $\sqrt{d}$ compares to previous work on robust linear bandits (Bogunovic et al., 2021). In addition, a variant of our policy is immediately applicable to a differential privacy model that interpolates between the local and central settings and achieves improved dependence on the privacy parameter $\varepsilon_{\mathsf{priv}}$.

While simple to implement, our algorithms require the learner to provide an upper bound on the corruption probability α, which may be difficult to estimate in practice. We leave the task of designing an adaptive policy as exciting future work. At the same time, it is unclear if our regret bounds for the privacy model (M2) are tight (in terms of the dependence on $\varepsilon_{\mathsf{priv}}$ and d). A natural question left open by our work is constructing tight lower bounds in this setting.

References

Basu, D., Dimitrakakis, C., and Tossou, A. Differential Privacy for Multi-armed Bandits: What Is It and What Is Its Cost? *arXiv e-prints*, art. arXiv:1905.12298, May 2019.

Bertsimas, D. and Mersereau, A. J. A learning approach for interactive marketing to a customer segment. *Operations Research*, 55(6):1120–1135, 2007. doi: 10.1287/opre.1070.0427.

Bittau, A., Erlingsson, U., Maniatis, P., Mironov, I., Raghunathan, A., Lie, D., Rudominer, M., Kode, U., Tinnes, J., and Seefeld, B. Prochlo: Strong Privacy for Analytics in the Crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles*, SOSP '17, pp. 441–459, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450350853. doi: 10.1145/3132747.3132769.

Bogunovic, I., Losalka, A., Krause, A., and Scarlett, J. Stochastic Linear Bandits Robust to Adversarial Attacks. In Banerjee, A. and Fukumizu, K. (eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 991–999. PMLR, 13–15 Apr 2021.

Bouneffouf, D., Rish, I., and Aggarwal, C. Survey on applications of multi-armed and contextual bandits. In *2020 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–8. IEEE, 2020.

Chen, S., Koehler, F., Moitra, A., and Yau, M. Online and distribution-free robustness: Regression and contextual bandits with huber contamination. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 684–695. IEEE, 2022.

Cheu, A. Differential Privacy in the Shuffle Model: A Survey of Separations. *arXiv e-prints*, art. arXiv:2107.11839, 2021.

Chowdhury, S. R. and Zhou, X. Shuffle private linear contextual bandits. *arXiv e-prints*, art. arXiv:2202.05567, 2022.

Diakonikolas, I. and Kane, D. M. Recent Advances in Algorithmic High-Dimensional Robust Statistics. *arXiv e-prints*, art. arXiv:1911.05911, November 2019.

Dimitrakakis, C., Nelson, B., Mitrokotsa, A., and Rubinstein, B. I. Robust and private bayesian inference. In *International Conference on Algorithmic Learning Theory*, pp. 291–305. Springer, 2014.

Dwork, C. and Roth, A. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Esfandiari, H., Karbasi, A., Mehrabian, A., and Mirrokni, V. Regret bounds for batched bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 7340–7348, 2021.

Esfandiari, H., Mirrokni, V., and Narayanan, S. Tight and robust private mean estimation with few users. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 16383–16412. PMLR, 17–23 Jul 2022.

Garcelon, E., Roziere, B., Meunier, L., Tarbouriech, J., Teytaud, O., Lazaric, A., and Pirotta, M. Adversarial attacks on linear contextual bandits. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 14362–14373. Curran Associates, Inc., 2020.

Garcelon, E., Chaudhuri, K., Perchet, V., and Pirotta, M. Privacy amplification via shuffling for linear contextual bandits. In Dasgupta, S. and Haghtalab, N. (eds.), *Proceedings of The 33rd International Conference on Algorithmic Learning Theory*, volume 167 of *Proceedings of Machine Learning Research*, pp. 381–407. PMLR, 29 Mar–01 Apr 2022.

Ghazi, B., Kumar, R., Manurangsi, P., and Nguyen, T. Robust and private learning of halfspaces. In *International Conference on Artificial Intelligence and Statistics*, pp. 1603–1611. PMLR, 2021.

Gross, D. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, 2011.

Gupta, A., Koren, T., and Talwar, K. Better algorithms for stochastic bandits with adversarial corruptions. In Beygelzimer, A. and Hsu, D. (eds.), *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 1562–1578. PMLR, 25–28 Jun 2019.

Han, Y., Liang, Z., Wang, Y., and Zhang, J. Generalized linear bandits with local differential privacy. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 26511–26522. Curran Associates, Inc., 2021.

Hanna, O. A., Girgis, A. M., Fragouli, C., and Diggavi, S. Differentially Private Stochastic Linear Bandits: (Almost) for Free. *arXiv e-prints*, art. arXiv:2207.03445, July 2022.

Jun, K.-S., Li, L., Ma, Y., and Zhu, J. Adversarial attacks on stochastic bandits. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Kasiviswanathan, S. P., Lee, H. K., Nissim, K., Raskhodnikova, S., and Smith, A. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

Kothari, P., Manurangsi, P., and Velingker, A. Private robust estimation by stabilizing convex relaxations. In *Conference on Learning Theory*, pp. 723–777. PMLR, 2022.

Lattimore, T. and Szepesvári, C. *Bandit algorithms*. Cambridge University Press, 2020.

Li, F., Zhou, X., and Ji, B. Differentially private linear bandits with partial distributed feedback. In *2022 20th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt)*, pp. 41–48. IEEE, 2022a.

Li, M., Berrett, T. B., and Yu, Y. On robustness and local differential privacy. *arXiv preprint arXiv:2201.00751*, 2022b.

Li, Y., Lou, E. Y., and Shan, L. Stochastic Linear Optimization with Adversarial Corruption. *arXiv e-prints*, art. arXiv:1909.02109, 2019.

Liu, F. and Shroff, N. Data poisoning attacks on stochastic bandits. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4042–4050. PMLR, 09–15 Jun 2019.

Liu, X., Kong, W., Kakade, S., and Oh, S. Robust and differentially private mean estimation. *Advances in Neural Information Processing Systems*, 34:3887–3901, 2021.

Lykouris, T., Mirrokni, V., and Paes Leme, R. Stochastic bandits robust to adversarial corruptions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pp. 114–122, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450355599. doi: 10.1145/3188745.3188918.

Prasad, A., Balakrishnan, S., and Ravikumar, P. A unified approach to robust mean estimation. *arXiv preprint arXiv:1907.00927*, 2019.

Shariff, R. and Sheffet, O. Differentially private contextual linear bandits. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Tenenbaum, J., Kaplan, H., Mansour, Y., and Stemmer, U. Differentially Private Multi-Armed Bandits in the Shuffle Model. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 24956–24967. Curran Associates, Inc., 2021.

Vershynin, R. *High-Dimensional Probability: An introduction with applications in data science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2018.

Wu, Y., Zhou, X., Tao, Y., and Wang, D. On Private and Robust Bandits. *arXiv e-prints*, art. arXiv:2302.02526, 2023. doi: 10.48550/arXiv.2302.02526.

Zheng, K., Cai, T., Huang, W., Li, Z., and Wang, L. Locally differentially private (contextual) bandits learning. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 12300–12310. Curran Associates, Inc., 2020.

A. Proofs from Section 2.2

We will work with the empirical second moment and covariance matrices defined below:

$$\widetilde{\Sigma}_{G_0} := \frac{1}{|G_0|}\sum_{i\in G_0} M_n^{-1/2} y_i^2 a_i a_i^{\mathsf{T}} M_n^{-1/2}, \qquad \Sigma_{G_0} := \widetilde{\Sigma}_{G_0} - \theta_{G_0}\theta_{G_0}^{\mathsf{T}}, \tag{25a}$$

$$\widetilde{\Sigma}_{n} := \frac{1}{n}\sum_{i=1}^{n} M_n^{-1/2} y_i^2 a_i a_i^{\mathsf{T}} M_n^{-1/2}, \qquad \Sigma_{n} := \widetilde{\Sigma}_{n} - \theta_{n}\theta_{n}^{\mathsf{T}}. \tag{25b}$$

In addition, we will use the vector notation below:

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\sigma} = \begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_n \end{pmatrix}. \tag{26}$$

Our guarantees will depend on the maximal weighted norm of the elements a_i, which we will denote by

$$\mu := \max_{i=1,\dots,n} \|a_i\|^2_{M_n^{-1}}. \tag{27}$$

Finally, we make the following assumption:

Assumption A.1. Fix δ to be a desired failure probability. The corruption probability α satisfies $\alpha \gtrsim \frac{\log(1/\delta)}{n}$.

To approximate θ_{G_0}, we will first reduce the above problem to robust mean estimation and apply the spectral filtering algorithm from the robust statistics literature. In (Prasad et al., 2019), the authors provide the following guarantee:

Theorem A.2 (Prasad et al. (2019, Theorem 4))**.** *Suppose that $\lambda \geq \|\Sigma_{G_0}\|_{\mathrm{op}}$ and that the set of inliers, G_0, satisfies $\frac{n-|G_0|}{n} + \frac{\log(1/\delta)}{n} \leq c$, where c is a dimension-independent constant. Then with probability at least $1-\delta$, the spectral filtering algorithm for robust mean estimation terminates in at most $O((n-|G_0|)+\log(1/\delta))$ steps and returns an estimate $\widetilde{\theta}$ satisfying*

$$\left\|\widetilde{\theta} - \theta_{G_0}\right\| \leq C\sqrt{\lambda}\left(\frac{n-|G_0|}{n} + \frac{\log(1/\delta)}{n}\right)^{1/2}. \tag{28}$$

In light of Theorem A.2, we will control the quantities involved. Before we proceed, we state the following bound for the size of G_0 that we will repeatedly appeal to throughout:

Lemma A.3. *Let $n \gtrsim \frac{\log(1/\delta)}{\alpha}$. Then with probability at least $1-\delta$, we have*

$$\left|\frac{|G_0|}{n} - (1-\alpha)\right| \leq \sqrt{\frac{\alpha\log(1/\delta)}{n}} \tag{29}$$

Proof. Let $S_n = \sum_{i=1}^n \mathbf{1}\{i \notin G_0\}$, which is a sum of i.i.d. Bernoulli random variables with parameter α. From the Chernoff bound (Vershynin, 2018, Exercise 2.3.5), it follows that for $\delta \in (0,1]$, we have

$$\mathbb{P}\left(|S_n - n\alpha| \geq \sqrt{n\alpha\log(1/\delta)}\right) \leq \delta, \tag{30}$$

after setting $\delta = \sqrt{\frac{\log(1/\delta)}{\alpha n}} \leq 1$ in (30). The claim follows since

$$\frac{|G_0|}{n} = 1 - \frac{S_n}{n} \in (1-\alpha) \pm \sqrt{\frac{\alpha\log(1/\delta)}{n}}.$$

□

A.1. Controlling the empirical mean

We now control the deviation of θ_{G_0} from the mean of the dataset absent any corruptions.

Lemma A.4. *With probability at least $1-\delta$, we have*

$$\left\|\theta_{G_0} - \frac{n(1-\alpha)}{|G_0|}\theta_n\right\| \lesssim \frac{\|\boldsymbol{y}\|}{|G_0|}\sqrt{\mu\alpha(1-\alpha)\log(1/\delta)}. \tag{31}$$

Proof. Define the following collection of random variables:

$$Q_i := (Z_i - \mathbb{E}[Z_i])M_n^{-1/2} y_i a_i, \quad \text{with } Z_i = \mathbf{1}\{i \in G_0\}. \tag{32}$$

Clearly, we have $\mathbb{E}[Q_i] = 0$. At the same time,

$$\mathbb{E}\left[\|Q_i\|^2\right] = \mathrm{Var}(Z_i) y_i^2 \|a_i\|^2_{M_n^{-1}} \le \alpha(1-\alpha) y_i^2 \mu.$$

Applying the vector Bernstein inequality (Gross, 2011, Theorem 12), we obtain

$$\mathbb{P}\left(\left\|\sum_{i=1}^n Q_i\right\| \ge \|\boldsymbol{y}\|\sqrt{\mu\alpha(1-\alpha)} + t\right) \le \exp\left(-\frac{t^2}{\mu\alpha(1-\alpha)\|\boldsymbol{y}\|^2}\right).$$

Consequently, we may set $t = \|\boldsymbol{y}\|\sqrt{\mu\alpha(1-\alpha)\log(1/\delta)}$ to obtain the claimed probability. Finally, we note that

$$\sum_{i=1}^n Q_i = \sum_{i\in G_0} M_n^{-1/2} y_i a_i - (1-\alpha)\sum_{i=1}^n M_n^{-1/2} y_i a_i = |G_0|\,\theta_{G_0} - n(1-\alpha)\theta_n.$$

□

A.2. Putting everything together

We now combine the bounds from Appendix A.1 and Theorem A.2. We first note that

$$\begin{aligned}
\Sigma_{G_0} = \widetilde{\Sigma}_{G_0} - \theta_{G_0}\theta_{G_0}^{\mathsf{T}} &\preceq \widetilde{\Sigma}_{G_0} \\
&= \frac{1}{|G_0|}\sum_{i\in G_0} y_i^2 M_n^{-1/2} a_i a_i^{\mathsf{T}} M_n^{-1/2} \\
&\preceq \frac{1}{|G_0|}\sum_{i\in G_0} y_i^2 \left\|M_n^{-1/2} a_i a_i^{\mathsf{T}} M_n^{-1/2}\right\|_{\mathrm{op}} I_d \\
&\preceq \frac{1}{|G_0|}\sum_{i\in G_0} y_i^2 \left\|M_n^{-1/2} a_i\right\|^2 I_d \\
&\preceq \frac{1}{|G_0|}\sum_{i\in G_0} y_i^2 \|a_i\|^2_{M_n^{-1}} I_d,
\end{aligned}$$

which implies that the spectral norm of Σ_{G_0} is bounded from above by

$$\|\Sigma_{G_0}\|_{\mathrm{op}} \le \frac{\|\boldsymbol{y}_{G_0}\|^2}{|G_0|}\cdot \max_i \|a_i\|^2_{M_n^{-1}} \le \frac{\mu\|\boldsymbol{y}\|^2}{|G_0|}. \tag{33}$$

At the same time, we appeal to Lemma A.3 to deduce that

$$|G_0| \ge (1-\alpha)n - 3\sqrt{n\log(1/\delta)} \ge \frac{(1-\alpha)n}{2}, \quad \text{for } n \ge \frac{18\log(1/\delta)}{(1-\alpha)^2}.$$

Consequently, we can replace the previous upper bound with $\|\Sigma_{G_0}\|_{\mathrm{op}} \le \frac{2\mu\|\boldsymbol{y}\|^2}{n(1-\alpha)}$.

We now appeal to Theorem A.2. Note that Lemma A.3 yields

$$\frac{n-|G_0|}{n} = 1 - \frac{|G_0|}{n} \le 1 - (1-\alpha) + \sqrt{\frac{\alpha \log(1/\delta)}{n}} = \alpha + \sqrt{\frac{\alpha \log(1/\delta)}{n}}.$$

Therefore, the estimate $\widetilde{\theta}$ computed by the spectral filtering algorithm satisfies

$$\left\|\widetilde{\theta} - \theta_{G_0}\right\| \lesssim \|\boldsymbol{y}\| \sqrt{\frac{2\mu}{n(1-\alpha)}} \left(\alpha + \sqrt{\frac{\alpha \log(1/\delta)}{n}} + \frac{\log(1/\delta)}{n}\right)^{1/2}. \tag{34}$$

Taking a union bound over Lemmas A.3 and A.4, we deduce that (34) holds with probability at least $1 - 2\delta$.

A.3. Application to phased elimination

Let $\theta_{\mathsf{LS}} := M_n^{-1} \sum_{i=1}^n y_i a_i$ denote the least squares solution from an approximate G-optimal design, and define

$$\bar{\theta}_{G_0} = M^{-1} \sum_{i \in G_0} y_i a_i = M^{-1/2} |G_0| \, \theta_{G_0}, \tag{35a}$$

$$\bar{\theta} = n M^{-1/2} \widetilde{\theta}. \tag{35b}$$

Note that $\bar{\theta}$ can be computed from the output of Algorithm 2, while $\bar{\theta}_{G_0}$ only serves for the analysis. With these at hand, we have the following decomposition:

$$\left\langle a, \bar{\theta} - \theta^\star \right\rangle = \left\langle a, \bar{\theta} - \bar{\theta}_{G_0} \right\rangle + \left\langle a, \bar{\theta}_{G_0} - \theta_{\mathsf{LS}} \right\rangle + \left\langle a, \theta_{\mathsf{LS}} - \theta^\star \right\rangle \tag{36}$$

In what follows, we bound each term in (36) separately.

A.3.1. Bounding the first term in (36)

The first term in (36) is equal to

$$\begin{aligned}\left\langle M^{-1/2} a, M^{1/2}(\bar{\theta} - \bar{\theta}_{G_0}) \right\rangle &= \left\langle M^{-1/2} a, n\widetilde{\theta} - |G_0| \, \theta_{G_0} \right\rangle = \left\langle M^{-1/2} a, n(\widetilde{\theta} - \theta_{G_0}) \right\rangle + (n - |G_0|) \left\langle M^{-1/2} a, \theta_{G_0} \right\rangle \\ &\le \|a\|_{M^{-1}} \left\| n(\widetilde{\theta} - \theta_{G_0}) \right\| + (n - |G_0|) \left\langle M^{-1/2} a, \theta_{G_0} \right\rangle \end{aligned} \tag{37}$$

In particular, the second term in (37) is given by

$$\begin{aligned}\left\langle M^{-1/2} a, \theta_{G_0} \right\rangle &= \frac{1}{|G_0|} \left\langle M^{-1/2} a, M^{-1/2} \sum_{i \in G_0} y_i a_i \right\rangle \\ &\le \frac{1}{|G_0|} \|a\|_{M^{-1}} \left\| \sum_{i \in G_0} y_i a_i \right\|_{M^{-1}} \\ &\le \frac{1}{|G_0|} \|a\|_{M^{-1}} \left\| \Big(\sum_{i \in G_0} a_i a_i^{\mathsf{T}} \Big)^{-1/2} \sum_{i \in G_0} y_i a_i \right\|, \end{aligned} \tag{38}$$

using the fact that $\sum_{i \in G_0} a_i a_i^{\mathsf{T}} \preceq \sum_{i=1}^n a_i a_i^{\mathsf{T}}$. Let A_{G_0} be a matrix whose rows are the vectors $\{a_i \mid i \in G_0\}$. We have

$$\sum_{i \in G_0} a_i a_i^{\mathsf{T}} = A_{G_0}^{\mathsf{T}} A_{G_0}, \quad \text{and} \quad \sum_{i \in G_0} y_i a_i = A_{G_0}^{\mathsf{T}} \boldsymbol{y}_{G_0}.$$

Letting $A_{G_0} = U \Sigma V^{\mathsf{T}}$ denote the economic SVD of A_{G_0}, we thus have

$$\left\| \Big(\sum_{i \in G_0} a_i a_i^{\mathsf{T}} \Big)^{-1/2} \sum_{i \in G_0} y_i a_i \right\| = \left\| (A_{G_0}^{\mathsf{T}} A_{G_0})^{-1/2} A_{G_0}^{\mathsf{T}} \boldsymbol{y}_{G_0} \right\| = \left\| V \Sigma^{-1} V^{\mathsf{T}} V \Sigma U^{\mathsf{T}} \boldsymbol{y}_{G_0} \right\| \le \|\boldsymbol{y}\|. \tag{39}$$

Plugging Eq. (39) into Eq. (38) and the result into Eq. (37), we obtain

$$\left\langle M^{-1/2}a, M^{1/2}(\bar{\theta} - \bar{\theta}_{G_0})\right\rangle \leq \|a\|_{M^{-1}} \left(\left\| n(\widetilde{\theta} - \theta_{G_0}) \right\| + \frac{n - |G_0|}{|G_0|} \|\boldsymbol{y}\| \right)$$

Using Eq. (34), the bound $\|a\|_{M^{-1}} \leq \sqrt{\mu}$, and Lemma A.3 with $\alpha \gtrsim \frac{\log(1/\delta)}{n}$, the above becomes:

$$\left\langle M^{-1/2}a, M^{1/2}(\bar{\theta} - \bar{\theta}_{G_0})\right\rangle \lesssim \mu \|\boldsymbol{y}\| \sqrt{n} \left(\alpha + \frac{\log(1/\delta)}{n} \right)^{1/2} \tag{40}$$

A.3.2. Bounding the second term in (36)

Recall that $\bar{\theta}_{G_0} = M^{-1/2} |G_0| \theta_{G_0}$. We further decompose the second term in (36) into

$$\begin{aligned}
\left\langle a, \bar{\theta}_{G_0} - \theta_{\mathsf{LS}} \right\rangle &= \left\langle M^{-1/2}a, M^{1/2}(\bar{\theta}_{G_0} - \theta_{\mathsf{LS}})\right\rangle \\
&= \left\langle M^{-1/2}a, M^{-1/2} \left(\sum_{i \in G_0} y_i a_i - \sum_{i=1}^n y_i a_i \right) \right\rangle \\
&= \left\langle M^{-1/2}a, |G_0| \left(\theta_{G_0} - \frac{n}{|G_0|} \theta_n \right) \right\rangle \\
&= \left\langle M^{-1/2}a, |G_0| \left(\theta_{G_0} - \frac{n(1-\alpha)}{|G_0|} \theta_n \right) \right\rangle + \left\langle M^{-1/2}a, n\alpha\theta_n \right\rangle
\end{aligned} \tag{41}$$

The first term in (41) can be upper bounded using Lemma A.4. Indeed,

$$\left\langle M^{-1/2}a, |G_0| \left(\theta_{G_0} - \frac{n(1-\alpha)}{|G_0|} \theta_n \right) \right\rangle \leq \|a\|_{M^{-1}} \|\boldsymbol{y}\| \sqrt{\mu\alpha \log(1/\delta)} \leq \mu \|\boldsymbol{y}\| \sqrt{\alpha \log(1/\delta)}.$$

We now simplify the second term in (41). With $e_i = r_i - \langle a_i, \theta^\star \rangle$, we obtain

$$\begin{aligned}
\left\langle M^{-1/2}a, n\alpha\theta_n \right\rangle &= \alpha \left\langle a, \sum_{i=1}^n M^{-1} y_i a_i \right\rangle \\
&= \alpha \left\langle a, \sum_{i=1}^n M^{-1} a_i (\langle a_i, \theta^\star \rangle + e_i) \right\rangle \\
&= \alpha \left\langle a, M^{-1} \left(\sum_{i=1}^n a_i a_i^\mathsf{T} \right) \theta^\star \right\rangle + \alpha \sum_{i=1}^n \left\langle a, M^{-1} a_i \right\rangle e_i \\
&= \alpha \langle a, \theta^\star \rangle + \alpha \sum_{i=1}^n \left\langle a, M^{-1} a_i \right\rangle e_i.
\end{aligned}$$

Since $\max_{a \in \mathcal{A}} |\langle a, \theta^\star \rangle| \leq 1$, combining the two bounds above yields

$$\left\langle a, \bar{\theta}_{G_0} - \theta_{\mathsf{LS}} \right\rangle \lesssim \mu \|\boldsymbol{y}\| \sqrt{\alpha \log(1/\delta)} + \alpha \left(1 + \sum_{i=1}^n \left\langle a, M^{-1} a_i \right\rangle e_i \right). \tag{42}$$

A.3.3. Bounding the third term in (36)

The last term is straightforward to bound. Let $e_i = y_i - \langle a_i, \theta^\star \rangle$ and note that

$$\theta_{\mathsf{LS}} - \theta^\star = M^{-1} \sum_{i=1}^n y_i a_i - \theta^\star = M^{-1} \sum_{i=1}^n a_i (\langle a_i, \theta^\star \rangle + e_i) - \theta^\star = M^{-1} \sum_{i=1}^n a_i e_i. \tag{43}$$

A.3.4. PUTTING EVERYTHING TOGETHER

Combining Eqs. (40), (42) and (43) yields the following robust confidence intervals:

$$\left|\left\langle a, \bar{\theta} - \theta^\star \right\rangle\right| \lesssim \mu \left\|\boldsymbol{y}\right\| \left[\sqrt{n}\left(\alpha + \frac{\log(1/\delta)}{n}\right)^{1/2} + \sqrt{\alpha \log(1/\delta)}\right] + \sum_{i=1}^{n} e_i \left\langle a, M^{-1} a_i \right\rangle + \alpha. \tag{44}$$

B. Missing proofs from Section 3

B.1. Missing proofs from Section 3.1

B.1.1. PROOF OF LEMMA 3.1

Proof. We write $e_i = \mathcal{M}(r_i) - \langle a_i, \theta^\star \rangle = \eta_i + \xi_i,\ \ \eta_i \sim \mathrm{SubG}(1),\ \xi_i \sim \mathsf{Lap}\left(\frac{2}{\varepsilon_{\mathsf{priv}}}\right)$. Now, define the random variables

$$X_i := \eta_i \left\langle a, M^{-1} a_i \right\rangle; \qquad Y_i := \xi_i \left\langle a, M^{-1} a_i \right\rangle.$$

The family $\{X_i\}$ is subgaussian with $\left\|X_i\right\|_{\psi_2} \le \left|\left\langle a, M^{-1} a_i \right\rangle\right|$. Consequently,

$$\begin{aligned}
\sum_{i=1}^{n} \left\|X_i\right\|_{\psi_2}^2 &\le \sum_{i=1}^{n} \left\langle a, M^{-1} a_i \right\rangle^2 \\
&= \sum_{i=1}^{n} \mathsf{Tr}(a^\mathsf{T} M^{-1} a_i a_i^\mathsf{T} M^{-1} a) \\
&= \left\langle M^{-1} a, \sum_{i=1}^{n} a_i a_i^\mathsf{T} M^{-1} a \right\rangle \\
&= \left\langle a, M^{-1} a \right\rangle \\
&= \left\|a\right\|_{M^{-1}}^2.
\end{aligned}$$

Therefore, applying the Hoeffding inequality (Vershynin, 2018, Theorem 2.6.2) yields:

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} \eta_i \left\langle a, M^{-1} a_i \right\rangle\right| \ge c_1 \left\|a\right\|_{M^{-1}} \sqrt{\log(1/\delta)}\right) \le \delta \tag{45}$$

On the other hand, when $\xi_i \sim \mathsf{Lap}(2/\varepsilon_{\mathsf{priv}})$, we have the Bernstein-style bound

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda \sum_{i=1}^{n} \xi_i \left\langle a, M^{-1} a_i \right\rangle}\right] &= \prod_{i=1}^{n} \mathbb{E}\left[\exp\left(\lambda \xi_i \left\langle a, M^{-1} a_i \right\rangle\right)\right] \\
&\le \prod_{i=1}^{n} \exp\left(\frac{\lambda^2 \left\langle a, M^{-1} a_i \right\rangle^2}{2\varepsilon_{\mathsf{priv}}^2}\right), \quad \forall \lambda \in \left(0, \frac{b}{\left\|\alpha\right\|_\infty}\right],
\end{aligned}$$

using (Vershynin, 2018, Proposition 2.7.1(e)) in the last step. Collecting terms we obtain

$$\prod_{i=1}^{n} \exp\left(\frac{\lambda^2 \left\langle a, M^{-1} a_i \right\rangle^2}{2\varepsilon_{\mathsf{priv}}^2}\right) = \exp\left(\lambda^2 \frac{\sum_{i=1}^{n} \left\langle a, M^{-1} a_i \right\rangle^2}{\varepsilon_{\mathsf{priv}}^2}\right) \le \exp\left(\lambda^2 c_1 \left(\frac{\left\|a\right\|_{M^{-1}}}{\varepsilon_{\mathsf{priv}}}\right)^2\right).$$

Now, appealing to (Vershynin, 2018, Proposition 2.7.1(a)), we obtain the concentration bound

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} \xi_i \left\langle a, M^{-1} a_i \right\rangle\right| \ge c_2 \frac{\left\|a\right\|_{M^{-1}} \log(1/\delta)}{\varepsilon_{\mathsf{priv}}}\right) \le \delta. \tag{46}$$

Combining the two bounds yields the result. □

B.1.2. PROOF OF LEMMA 3.3

Proof. Let π below denote an approximate G-optimal design in the sense of Definition 1.6. We have

$$\begin{aligned} M &= \sum_{i=1}^{n} a_i a_i^\mathsf{T} \\ &= \sum_{a \in \mathrm{supp}(\pi)} n_a a a^\mathsf{T} \\ &= n \cdot \sum_{a \in \mathrm{supp}(\pi)} \pi(a) a a^\mathsf{T} \\ &= nM(\pi). \end{aligned}$$

Consequently, we have the inequality

$$\begin{aligned} \|a\|_{M^{-1}}^2 &= \left\langle a, M^{-1} a \right\rangle \\ &= \left\langle a, (nM(\pi))^{-1} a \right\rangle \\ &= \frac{1}{n} \left\langle a, M^{-1}(\pi) a \right\rangle \\ &= \frac{\|a\|_{M^{-1}(\pi)}^2}{n} \\ &\leq \frac{2d}{n}, \end{aligned}$$

using the fact that π is an approximate G-optimal design in the last inequality. □

B.1.3. PROOF OF LEMMA 3.3

Proof. With $\boldsymbol{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^\mathsf{T}$, we have $\|\boldsymbol{y}\| \leq \sqrt{n} \|\boldsymbol{y}\|_\infty$. To control the latter, we note

$$\max_i |\langle a_i, \theta^\star \rangle + \eta_i + \xi_i| \leq \max_i \{|\langle a_i, \theta^\star \rangle| + |\eta_i| + |\xi_i|\} \leq 1 + \max_i |\eta_i| + \max_i |\xi_i| .$$

Since $\eta_i \sim \mathrm{SubG}(1)$, standard concentration inequalities for subgaussian maxima yield

$$\mathbb{P}\left(\max_i |\eta_i| \geq C \sqrt{\log(n/\delta)} \right) \leq \delta. \tag{47}$$

Similarly, ξ_i are subexponential with parameter $2/\varepsilon_{\mathsf{priv}}$. By a union bound and (Vershynin, 2018, Proposition 2.7.1),

$$\mathbb{P}\left(\max_i |\xi_i| \geq t \right) \leq \sum_{i=1}^{n} \mathbb{P}\left(|\xi_i| \geq t \right) \leq n \exp\left(- \min\left\{ \frac{\varepsilon_{\mathsf{priv}}^2 t^2}{8}, \frac{\varepsilon_{\mathsf{priv}} t}{4} \right\} \right)$$

Setting $t := \frac{4 \log(n/\delta)}{\varepsilon_{\mathsf{priv}}}$ above yields $\max_i |\xi_i| \leq \frac{4 \log(n/\delta)}{\varepsilon_{\mathsf{priv}}}$ with probability at least $1 - \delta$.

Finally, taking a union bound and relabelling yields the result. □

B.1.4. PROOF OF THEOREM 3.4

Proof. We perform a regret analysis under the LDP model (M1). First, we simplify Proposition 2.1 using Lemmas 3.1 and 3.3 and the assumption $\alpha \gtrsim \log(1/\delta)/n$. Letting $\mu := \max_{a \in \mathcal{A}} \|a\|_{M^{-1}}^2$, we have

$$\left| \langle a, \widetilde{\theta} - \theta^\star \rangle \right| \lesssim \mu \left(n\sqrt{\alpha} + \sqrt{\alpha n \log(1/\delta)} \right) \left(1 + \sqrt{\log(n/\delta)} + \frac{\log(n/\delta)}{\varepsilon_{\mathsf{priv}}} \right) + \sqrt{\mu \log(1/\delta)} \left(1 + \frac{\sqrt{\log(1/\delta)}}{\varepsilon_{\mathsf{priv}}} \right) + \alpha, \tag{48}$$

for any fixed a with probability at least $1 - \delta$ by suitably adjusting constants.

In particular, when $a_1, \dots, a_n$ are drawn from an approximate G-optimal design, Lemma 3.3 implies that

$$\mu \equiv \max_{a \in \mathcal{A}} \|a\|^2_{M_n^{-1}} \le \frac{2d}{n}, \tag{49}$$

so the bound in (48) can be written as

$$\left|\langle a, \widetilde{\theta} - \theta^\star \rangle\right| \lesssim d \left(\sqrt{\alpha} + \sqrt{\frac{\alpha \log(1/\delta)}{n}}\right) \left(1 + \sqrt{\log(n/\delta)} + \frac{\log(n/\delta)}{\varepsilon_{\mathsf{priv}}}\right) + \sqrt{\frac{d \log(1/\delta)}{n}} \left(1 + \frac{\sqrt{\log(1/\delta)}}{\varepsilon_{\mathsf{priv}}}\right), \tag{50}$$

By standard arguments (see, e.g., the proof of (Esfandiari et al., 2021, Theorem 5.1)), we may focus on bounding the regret conditioned on the "good" event where all the invocations to the coreset construction and robust filtering algorithms succeed. This requires us to choose failure probability δ proportional to $\delta'/(KT^2)$, where T is the number of rounds, K is the size of the action space, and δ' is an overall desired failure probability. To ease notation, we relabel δ in this manner below.

Now, recalling the width of the confidence interval

$$\gamma_i := \sqrt{d} \left(\sqrt{\log(q^i/\delta)} + \frac{\log(q^i/\delta)}{\varepsilon_{\mathsf{priv}}}\right) (\sqrt{\alpha} + \alpha\sqrt{d}) + \alpha + \sqrt{\frac{d \log(1/\delta)}{q^i}} \left(1 + \frac{\sqrt{\log(1/\delta)}}{\varepsilon_{\mathsf{priv}}}\right),$$

we have the following expression for the regret:

$$\begin{aligned}
\mathsf{Regret} &= \sum_{i=1}^{B} (\text{arms pulled}) \times (\text{instantaneous regret}) \\
&\le \sum_{i=1}^{B} q^i 4\gamma_{i-1} \\
&\lesssim \sum_{i=1}^{B} q^i \sqrt{\frac{d \log(1/\delta)}{q^{i-1}}} \left(1 + \frac{\sqrt{\log(1/\delta)}}{\varepsilon_{\mathsf{priv}}}\right) + \sqrt{d}(\sqrt{\alpha} + \alpha\sqrt{d}) \sum_{i=1}^{B} q^i \left(\sqrt{\log(q^{i-1}/\delta)} + \frac{\log(q^{i-1}/\delta)}{\varepsilon_{\mathsf{priv}}}\right)
\end{aligned} \tag{51}$$

To bound the first sum above, we notice that

$$\sum_{i=1}^{B} q^i \sqrt{\frac{1}{q^{i-1}}} = q \sum_{i=0}^{B-1} \sqrt{q^i} = q \cdot \frac{q^{B/2} - 1}{q^{1/2} - 1}.$$

For the second sum, we first bound $\log(q^{i-1}/\delta) \le \log(T^{\frac{B-1}{B}}/\delta) \le \log(T/\delta)$, followed by

$$\sum_{i=1}^{B} q^i \left(\sqrt{\log(q^{i-1}/\delta)} + \frac{\log(q^{i-1}/\delta)}{\varepsilon_{\mathsf{priv}}}\right) \le \left(\sqrt{\log(T/\delta)} + \frac{\log(T/\delta)}{\varepsilon_{\mathsf{priv}}}\right) T.$$

Finally, we note that when $B \ge \log(T)$ we have $\frac{q^{B/2}-1}{q^{1/2}-1} \lesssim \sqrt{T}$ and $q = T^{1/B} \le e$. Therefore,

$$\mathsf{Regret} \lesssim \sqrt{Td \log(1/\delta)} \left(1 + \frac{\sqrt{\log(1/\delta)}}{\varepsilon_{\mathsf{priv}}}\right) + T \max\left\{\sqrt{\alpha d}, \alpha d\right\} \left(\sqrt{\log(T/\delta)} + \frac{\log(T/\delta)}{\varepsilon_{\mathsf{priv}}}\right).$$

□

B.2. Missing proofs from Section 3.2

B.2.1. Proof of Lemma 3.6

Proof. We have $e_v = \mathcal{M}(r_v) - \langle v, \theta^\star \rangle = \eta_v + \xi_v$, where $\eta_v \sim \mathrm{SubG}(1/n_a)$ and $\xi_v \sim \mathsf{Lap}(2/n_a\varepsilon_{\mathsf{priv}})$. Therefore,

$$\eta_v + \xi_v \overset{(d)}{=} \frac{1}{\sqrt{n_a}} \tilde{\eta}_v + \frac{1}{n_a} \tilde{\xi}_v, \quad \tilde{\eta}_v \sim \mathrm{SubG}(1),\ \tilde{\xi}_v \sim \mathsf{Lap}(2/\varepsilon_{\mathsf{priv}}).$$

Consequently, we may trace the proof of Lemma 3.1 to arrive at

$$\sum_{v\in\mathcal{H}} \eta_v \left\langle a, M^{-1}v\right\rangle \lesssim \|a\|_{M^{-1}} \sqrt{\log(1/\delta)} \left(\frac{c_1}{\sqrt{n_a}} + \frac{c_2\sqrt{\log(1/\delta)}}{n_a\varepsilon_{\mathsf{priv}}}\right). \tag{52}$$

This completes the proof after noticing that $n_a \geq \nu m$. □

B.2.2. Proof of Lemma 3.7

Proof. Recall that $M = \sum_{v\in\mathcal{H}} vv^\mathsf{T}$. In particular, we have

$$\sum_{v\in\mathcal{H}} vv^\mathsf{T} = \sum_{v\in\mathrm{supp}(\hat{\pi})} \succeq \sum_{v\in\mathcal{H}} \hat{\pi}(v) vv^\mathsf{T} \implies \|a\|^2_{M^{-1}} \leq \|a\|^2_{M^{-1}(\hat{\pi})} \leq 2d, \tag{53}$$

where the last inequality follows since π is an approximate G-optimal design. □

B.2.3. Proof of Lemma 3.8

Proof. Let $k = |\mathrm{supp}(\hat{\pi})|$ and let $y_1, \ldots, y_k$ be an enumeration of the elements $y_v,\ v \in \mathrm{supp}(\hat{\pi})$. With $\boldsymbol{y} = \begin{bmatrix} y_1 & \ldots & y_k \end{bmatrix}^\mathsf{T}$, we have $\|\boldsymbol{y}\| \leq \sqrt{k}\,\|\boldsymbol{y}\|_\infty$. To control the latter, we note

$$\max_v |\langle v, \theta^\star\rangle + \eta_v + \xi_v| \leq \max_v \{|\langle v, \theta^\star\rangle| + |\eta_v| + |\xi_v|\} \leq 1 + \max_v |\eta_v| + \max_i |\xi_v|.$$

Since $\eta_v \sim \mathrm{SubG}(1/n_v)$, standard concentration inequalities for subgaussian maxima yield

$$\mathbb{P}\left(\max_v |\eta_v| \geq C\sqrt{\frac{\log(k/\delta)}{n_v}}\right) \leq \delta. \tag{54}$$

Similarly, ξ_v are subexponential with parameter $\frac{2}{n_v\varepsilon_{\mathsf{priv}}}$. By a union bound and (Vershynin, 2018, Proposition 2.7.1),

$$\begin{aligned}
\mathbb{P}\left(\max_v |\xi_v| \geq t\right) &\leq \sum_{v\in\mathrm{supp}(\hat{\pi})} \mathbb{P}\left(|\xi_v| \geq t\right) \\
&\leq \sum_{v\in\mathrm{supp}(\hat{\pi})} \mathbb{P}\left(|\xi_v| \geq t\right) \\
&\leq k\exp\left(-\min\left\{\frac{\min_v n_v^2\varepsilon_{\mathsf{priv}}^2 t^2}{8}, \frac{\min_v n_v\varepsilon_{\mathsf{priv}} t}{4}\right\}\right) \\
&\leq k\exp\left(-\min\left\{\frac{\nu^2 m^2\varepsilon_{\mathsf{priv}}^2 t^2}{8}, \frac{\nu m\varepsilon_{\mathsf{priv}} t}{4}\right\}\right).
\end{aligned}$$

Setting $t := \frac{4\log(k/\delta)}{\nu m\varepsilon_{\mathsf{priv}}}$ above yields $\max_v |\xi_v| \leq \frac{4\log(k/\delta)}{\nu m\varepsilon_{\mathsf{priv}}}$ with probability at least $1-\delta$.

Finally, taking another union bound and relabelling yields the result. □

B.2.4. Proof of Theorem 3.9

Proof. We proceed with deriving an expression for the robust confidence interval from Proposition 2.1 under (M2). Indeed, with probability at least $1-\delta$, for any fixed $a \in \mathcal{A}$ we have:

$$\begin{aligned}
\left|\langle a, \widetilde{\theta} - \theta^\star\rangle\right| \lesssim{}& \sqrt{\frac{d\log(1/\delta)}{\nu m}}\left(1 + \frac{1}{\varepsilon_{\mathsf{priv}}}\sqrt{\frac{\log(1/\delta)}{\nu m}}\right) \\
&+ 2d\left(1 + \sqrt{\frac{\log(k/\delta)}{\nu m}} + \frac{\log(k/\delta)}{\nu m\varepsilon_{\mathsf{priv}}}\right)\left(\sqrt{k\alpha} + \sqrt{\alpha\log(1/\delta)}\right) \\
&+ \alpha,
\end{aligned} \tag{55}$$

where $k := |\mathrm{supp}(\pi)|$. Recall we can find (in poly-time) an approximate G-optimal design π satisfying

$$k := |\mathrm{supp}(\pi)| \lesssim d \log\log d. \tag{56}$$

Therefore, we may proceed with the regret analysis. Similarly to the proof of Theorem 3.4, we condition on the case where all randomized algorithms and invocations to random events succeed with high probability.

Then, with $m = q^i$ at round i, we have the following bound:

$$\begin{aligned} n_i &= \sum_{v \in \mathrm{supp}(\pi)} n_v \\ &= \sum_{v \in \mathrm{supp}(\pi)} \left\lceil q^i \max\{\pi(v), \nu\} \right\rceil \\ &\leq \sum_{v \in \mathrm{supp}(\pi)} q^i \max\{\pi(v), \nu\} + 1 \\ &= \mathrm{supp}(\pi) + q^i \sum_v \max\{\pi(v), \nu\} \\ &\lesssim q^i (1 + \nu d \log\log d). \end{aligned}$$

In particular, we have the following property for the sum $\sum_i q^i$:

$$\sum_{i=1}^{B} q^i = \frac{1}{1 + \nu d \log\log d} \sum_{i=1}^{B} n_i = \frac{T}{1 + \nu d \log\log d}. \tag{57}$$

Consequently, the regret of the algorithm conditioned on the good event is given by

$$\mathsf{Regret} \leq 4 \sum_{i=1}^{B} n_i \gamma_{i-1} \lesssim q(1 + \nu d \log\log d) \sum_{i=0}^{B-1} q^i \gamma_i, \tag{58}$$

where γ_i is the width of the confidence interval at round i. The first term in the sum $\sum_i q^i \gamma_i$ is

$$\begin{aligned} \sum_{i=0}^{B-1} q^i \sqrt{\frac{d \log(1/\delta)}{\nu q^i}} + q^i \frac{\sqrt{d} \log(1/\delta)}{\nu q^i \varepsilon_{\mathsf{priv}}} &\leq \sqrt{\frac{d \log(1/\delta)}{\nu}} \left(\sum_{i=0}^{B-1} q^{i/2} + \frac{\sqrt{\log(1/\delta)}}{\varepsilon_{\mathsf{priv}} \sqrt{\nu}} \sum_{i=0}^{B-1} \right) \\ &\leq \sqrt{\frac{d \log(1/\delta)}{\nu}} \left(\frac{q^{B/2} - 1}{q^{1/2} - 1} + \sqrt{\frac{\log(1/\delta)}{\nu}} \frac{B-1}{\varepsilon_{\mathsf{priv}}} \right), \end{aligned} \tag{59}$$

which is a term independent of the corruption fraction. The second group of summands in $\sum_i q^i \gamma_i$ is

$$\begin{aligned} &\sum_{i=0}^{B-1} q^i \left(1 + \sqrt{\frac{\log(d \log\log d/\delta)}{\nu q^i}} + \frac{\log(d \log\log d/\delta)}{\nu q^i \varepsilon_{\mathsf{priv}}} \right) \\ &= \frac{T}{1 + \nu d \log\log d} + \sqrt{\frac{\log(d \log\log d/\delta)}{\nu}} \cdot \frac{q^{B/2} - 1}{q^{1/2} - 1} + \frac{\log(d \log\log d/\delta)}{\nu} \cdot \frac{B-1}{\varepsilon_{\mathsf{priv}}}. \end{aligned} \tag{60}$$

Finally, summing over i using the last term of the confidence interval yields

$$\sum_{i=0}^{B-1} q^i \alpha = \frac{\alpha T}{1 + \nu d \log\log d}. \tag{61}$$

Putting everything together, we arrive at the claimed regret bound:

$$\begin{aligned} \mathsf{Regret} &\lesssim (1 + \nu d \log\log d) \left(\sqrt{\frac{dT \log(1/\delta)}{\nu}} + \frac{\log(1/\delta) \log(T) \sqrt{d}}{\varepsilon_{\mathsf{priv}} \sqrt{\nu}} \right) \\ &\quad + 2d \left(\sqrt{\alpha d \log\log d} + \sqrt{\alpha \log(1/\delta)} \right) \left(T + \sqrt{\frac{Td \log\log d/\delta}{\nu}} + \frac{\log(d \log\log d/\delta)}{\nu} \frac{\log T}{\varepsilon_{\mathsf{priv}}} \right) \\ &\quad + \alpha T. \end{aligned} \tag{62}$$

□