Linear Contextual Bandits with Adversarial Corruptions

Anonymous Author(s)
Affiliation
Address
email

Abstract

We study the linear contextual bandits problem in the presence of adversarial corruption, where the interaction between the player and a possibly infinite decision set is contaminated by an adversary that can corrupt the reward up to a corruption level C measured by the sum of the largest alteration on rewards in each round. We present a variance-aware algorithm that is adaptive to the level of adversarial contamination C. The key algorithmic design includes (1) a multi-level partition scheme of the observed data, (2) a cascade of confidence sets that are adaptive to the level of the corruption, and (3) a variance-aware confidence set construction that can take advantage of low-variance reward. We further prove that the regret of the proposed algorithm is $\widetilde{O}(C^2 d\sqrt{\sum_{t=1}^{T}\sigma_t^2} + C^2\sqrt{dT} + CR\sqrt{dT})$, where d is the dimension of context vectors, T is the number of rounds, R is the range of noise and $\sigma_t^2, t = 1 \ldots, T$ are the variances of instantaneous reward. We also prove a gap-dependent regret bound for the proposed algorithm, which is instance-dependent and thus leads to better performance on good practical instances. To the best of our knowledge, this is the first variance-aware corruption robust algorithm for contextual bandits.

1 Introduction

Multi-armed bandits algorithms are widely applied in online advertising (Li et al., 2010), clinical trials (Villar et al., 2015), recommendation system (Deshpande and Montanari, 2012) and many other real-world tasks. In the model of multi-armed bandits, the algorithm needs to decide which action (or arm) to take (or pull) at each round and receive a reward for the chosen action. In the stochastic setting, the reward is subject to a fixed but unknown distribution for each action. In reality, however, these rewards can easily be "corrupted" by some malicious users. A typical example is click fraud (Lykouris et al., 2018), where botnets simulate the legitimate users clicking on an ad to fool the recommendation systems. This motivates the studies of the bandits algorithms that are robust to adversarial corruptions.

For example, Lykouris et al. (2018) introduced a bandit model in which an adversary could corrupt the stochastic reward generated by an arm pull. They proposed an algorithm and show that the regret of this "middle ground" scenario degrades smoothly with the amount of corruption injected by the adversary. Gupta et al. (2019) proposed an alternative algorithm which gives a significant improvement in regret.

While the algorithms that are robust to the corruptions have been studied in the setting of multi-armed bandits in a number of prior works, they are still understudied in the setting of linear contextual bandits. The linear contextual bandits problem can be regarded as an extension of the multi-armed bandit problem to linear optimization, in order to tackle an unfixed and possibly infinite set of feasible

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

actions. There is a large body of literature on efficient algorithms for linear contextual bandits with no corruptions (Abe et al., 2003; Auer, 2002; Chu et al., 2011; Dani et al., 2008; Rusmevichientong and Tsitsiklis, 2010; Abbasi-Yadkori et al., 2011; Li et al., 2019b), to mention a few. The significance of this setting lies in the fact that linear regression approaches are widely used in recommendation systems and advertising (Li et al., 2010; Jhalani et al., 2016; Deshpande and Montanari, 2012). Linear contextual bandits with adversarial corruptions is an arguably more challenging setting since most of the previous corruption-robust algorithms are based on the idea of action elimination (Lykouris et al., 2018; Gupta et al., 2019; Bogunovic et al., 2021), which is not applicable to the contextual bandits settings where the decision set is time varying and possibly infinite at each round. In Garcelon et al. (2020), it is shown that a malicious agent can force a linear contextual bandit algorithm to take any desired action $T - o(T)$ times over T rounds, while applying adversarial corruptions to rewards with a cumulative cost that only grow logarithmically. This poses a big challenge for designing corruption robust algorithms for linear contextual bandits.

In this paper, we make a first attempt to study a linear contextual bandit model where an adversary can corrupt the rewards up to a corruption level C, which is defined as the the sum of biggest alteration the adversary made on rewards in each round. We propose a linear contextual bandits algorithm that is robust to reward corruption, dubbed multi-level optimism-in-the-face-of-uncertainty weighted learning (Multi-level OFUL). More specifically, our algorithm consists of the following novel techniques: (1) We design a multi-level partition scheme and adopt the idea of *sub-sampling* to do the robust estimation of the model parameters; (2) We maintain a cascade of *candidate confidence sets* corresponding to different corruption level (which is unknown) and randomly select a confidence set at each round to take the action; and (3) We design confidence sets that depend on the *variances of rewards*, which lead to a potentially tighter regret bound.

Our contributions are summarized as follows:

- We propose a variance-aware algorithm which is adaptive to the amount of adversarial corruptions C. To the best of our knowledge, it is the first algorithm for the setting of linear contextual bandits with adversarial corruptions which does not rely on the finite number of actions and other additional assumptions.
- We prove that the regret of our algorithm is in $\widetilde{O}\left(C^2 d\sqrt{\sum_{t=1}^T \sigma_t^2} + C^2\sqrt{dT} + CR\sqrt{dT}\right)$, where d is the dimension of context vectors, T is the number of rounds, R is the range of noise and $\sigma_t^2, t = 1\ldots, T$ are the variances of instantaneous reward. Our regret upper bound has a multiplicative dependence on C^2 which indicates that our algorithm achieves a sub-linear regret when the corruption level satisfies $C = o(T^{1/4})$.
- We also derive a gap-dependent regret bound $\widetilde{O}\left(\frac{1}{\Delta}\cdot C^2R^2d + \frac{1}{\Delta}\cdot d^2C^2\max_{t\in[T]}\sigma_t^2\right)$ for our proposed algorithm, which is instance-dependent and thus leads to a better performance on good practical instances.

Notation. We use lower case letters to denote scalars, and use lower and upper case bold face letters to denote vectors and matrices respectively. We denote by $[n]$ the set $\{1,\ldots,n\}$. For a vector $\mathbf{x}\in\mathbb{R}^d$ and matrix $\boldsymbol{\Sigma}\in\mathbb{R}^{d\times d}$, a positive semi-definite matrix, we denote by $\|\mathbf{x}\|_2$ the vector's Euclidean norm and define $\|\mathbf{x}\|_{\boldsymbol{\Sigma}} = \sqrt{\mathbf{x}^\top\boldsymbol{\Sigma}\mathbf{x}}$. For two positive sequences $\{a_n\}$ and $\{b_n\}$ with $n = 1, 2, \ldots,$ we write $a_n = O(b_n)$ if there exists an absolute constant $C > 0$ such that $a_n \le Cb_n$ holds for all $n \ge 1$ and write $a_n = \Omega(b_n)$ if there exists an absolute constant $C > 0$ such that $a_n \ge Cb_n$ holds for all $n \ge 1$. We use $\widetilde{O}(\cdot)$ to further hide the polylogarithmic factors. We use $\mathbb{1}(\cdot)$ to denote the indicator function.

2 Related Work

Bandits with Adversarial Attacks: There is a large body of literature on the problems of multi-armed bandits with adversarial corruptions. Most research in this area aims to design algorithms that achieve desirable regret bound in both stochastic multi-armed bandits and adversarial bandits, known as "the best of both worlds" guarantees (Bubeck and Slivkins, 2012; Seldin and Slivkins, 2014; Auer and Chiang, 2016; Seldin and Lugosi, 2017; Zimmert and Seldin, 2019). These works mainly focus on achieving bounds in the worst case and the case where there is no adversary. As a result, these algorithms are either not robust to instances that moderate amount of corruptions occur, or suffer from restrictive assumptions on adversarial corruptions. Distinctive from the above line of research, Lykouris et al. (2018) focus on a variant of classic multi-armed bandit model in which each pull of an arm generates a stochastic reward that may be contaminated by an adversary before it is revealed to the player. In their work, the corruption level C is defined as $C = \sum_t \max_a |r^t(a) - r^t_{\mathcal{S}}(a)|$ where $r^t_{\mathcal{S}}(a)$ is the stochastic reward of arm a and $r^t(a)$ is the corrupted reward of arm a at round t. They develop algorithms adaptive to the unknown corruption level, which achieves an $O(K^{1.5}C\sqrt{T})$ regret bound. Gupta et al. (2019) proposed an improved algorithm that can achieve a regret bound with only additive dependence on C.

On the other hand, many research efforts have also been devoted into designing adversarial attacks that cause standard algorithms to fail (Jun et al., 2018; Liu and Shroff, 2019; Gupta et al., 2019; Garcelon et al., 2020).

Linear Bandits with Corruptions: Li et al. (2019a) studied stochastic linear bandits with adversarial corruptions and achieved $\widetilde{O}(\frac{1}{\Delta} \cdot d^{5/2}C + \frac{1}{\Delta^2} \cdot d^6)$ regret bound where d is the dimension of the context vectors, Δ is the gap between the rewards of the best and the second best action in the decision set $\mathcal{D}$. The distinction between Li et al. (2019a) and our work is that Li et al. (2019a) considers a fixed decision set $\mathcal{D}$ throughout all T rounds, while we consider contextual bandits with changing decision set observed before each round. Bogunovic et al. (2021) also studied linear bandits with adversarial corruptions and considered the setting under the assumption that context vectors undergo small random perturbations, which is previously introduced by Kannan et al. (2018). Aside from the additional assumption, another major distinction in Bogunovic et al. (2021) is that the number of actions k is finite and the regret bound depends on k in the contextual setting with unknown corruption level C. Recently, Lee et al. (2021) considered corrupted linear bandits with a finite and fixed decision set and achieve an instance-independent regret of $\widetilde{O}(d\sqrt{T} + C)$. Though both their work and the work by Li et al. (2019a) focus on corrupted linear stochastic bandits, Lee et al. (2021) have a slightly different definition of regret and adopt a strong assumption on corruptions that in each round t, the corruptions on rewards are linear in the actions. Neu and Olkhovskaya (2020) studied linear contextual bandits with a finite decision set (i.e., K actions) and an adversary. Unlike our model, they assume that the adversary can add an arbitrary noise to the loss under a limited amount ϵ and prove an $\widetilde{O}((Kd)^{\frac{1}{3}}T^{\frac{2}{3}}) + \epsilon \cdot \sqrt{d}T$ regret bound for their proposed algorithm. Kapoor et al. (2019) considered the corrupted linear contextual bandits setting under a strong assumption on corruptions that for any prefix, at most an η fraction of the rounds are corrupted.

3 Preliminaries

In this paper, we study linear contextual bandits with adversarial corruptions. We will introduce our model and some basic concepts in this section.

Corrupted linear contextual bandits. We consider the the linear contextual bandits model studied in Abbasi-Yadkori et al. (2011) under the same corruption studied by Lykouris et al. (2018). In detail, distinctive from the linear contextual bandits Abbasi-Yadkori et al. (2011), the interaction between the agent and the environment is now contaminated by an adversary. The protocol between the agent and the adversary at each round $t \in [T]$ can be described as follows:

1. At the beginning of round t, the environment generates an arbitrary decision set $\mathcal{D}_t \subseteq \mathbb{R}^d$ where each element represents a feasible action that can be selected by the agent.
2. The environment generates stochastic reward function $r'_t(\mathbf{a}) = \langle \mathbf{a}, \boldsymbol{\mu}^* \rangle + \epsilon_t(\mathbf{a})$ together with an upper bound on the standard variance of $\epsilon_t(\mathbf{a})$, i.e., $\sigma_t(\mathbf{a})$ for all $\mathbf{a} \in \mathcal{D}_t$.
3. The adversary observes $\mathcal{D}_t$, $r'_t(\mathbf{a})$, $\sigma_t(\mathbf{a})$ for all $\mathbf{a} \in \mathcal{D}_t$ and decides a corrupted reward function r_t defined over $\mathcal{D}_t$.
4. The agent observes $\mathcal{D}_t$ and selects $\mathbf{a}_t \in \mathcal{D}_t$.
5. The adversary observes $\mathbf{a}_t$ and then returns $r_t(\mathbf{a}_t)$ and $\sigma_t(\mathbf{a}_t)$.
6. The agent observes $r_t(\mathbf{a}_t), \sigma_t(\mathbf{a}_t)$.

Let $\mathcal{F}_t$ be the σ-algebra generated by $\mathcal{D}_{1:t}, \mathbf{a}_{1:t-1}, \epsilon_{1:t-1}, r_{1:t-1}$ and $\sigma_{1:t-1}$.

At step 2, $\boldsymbol{\mu}^*$ is a hidden vector unknown to the agent which can be observed by the adversary at the beginning. We assume that for all $t \geq 1$ and all $\mathbf{a} \in \mathcal{D}_t$, $\|\mathbf{a}\|_2 \leq A$, $|\langle \mathbf{a}, \boldsymbol{\mu}^* \rangle| \leq 1$ and $\|\boldsymbol{\mu}^*\|_2 \leq B$ almost surely. $\epsilon_t(\mathbf{a})$ can be any form of random noise as long as it satisfies

$$\forall t \geq 1, \forall \mathbf{a} \in \mathcal{D}_t, |\epsilon_t(\mathbf{a})| \leq R, \quad \mathbb{E}[\epsilon_t(\mathbf{a})|\mathcal{F}_t] = 0, \quad \mathbb{E}[\epsilon_t^2(\mathbf{a})|\mathcal{F}_t] \leq \sigma_t^2(\mathbf{a}). \tag{3.1}$$

This assumption on ϵ_t is a variant of that in Zhou et al. (2020): We now require the noise to be generated for all $\mathbf{a} \in \mathcal{D}_t$ in advance before the adversary decides the corrupted reward function. Our assumption on noises is more general than those in (Li et al., 2019a; Bogunovic et al., 2021; Kapoor et al., 2019) where they are assumed to be 1-sub-Gaussian or Gaussian.

At step 3, we assume that the adversary has observed all the previous information and thus may predict which policy the agent will take at the current round. However, since the agent can take a randomized policy, the adversary may not know exactly which action the agent will take.

Corruption level. We define corruption level

$$C = \frac{1}{R+1} \sum_{t=1}^{T} \sup_{\mathbf{a} \in \mathcal{D}_t} |r_t'(\mathbf{a}) - r_t(\mathbf{a})|. \tag{3.2}$$

to indicate the level of adversarial contamination. We say a model is C-corrupted if the corruption level is no larger than C.

Our definition of corruption level is equivalent to the counterpart in Lykouris et al. (2018) and Gupta et al. (2019) where they define $C = \sum_{t=1}^{T} \max_{\mathbf{a}} |r_t'(\mathbf{a}) - r_t(\mathbf{a})|$ in our notation of rewards. We introduce a factor of $\frac{1}{R+1}$ since the noise is of range R in our model, while they assume all the rewards are in range $[0, 1]$.

Regret. Since the actions selected by the agent may not be deterministic, we define the regret for this model as follows:

$$\mathbf{Regret}(T) = \sum_{t=1}^{T} \langle \mathbf{a}_t^*, \boldsymbol{\mu}^* \rangle - \mathbb{E}\left[\sum_{t=1}^{T} \langle \mathbf{a}_t, \boldsymbol{\mu}^* \rangle\right]. \tag{3.3}$$

Our definition follows from the definition in Gupta et al. (2019) where the standard metric in stochastic multi-armed bandit models of pseudo-regret is adopted. But note that we need to take the expectation on $\sum_{t=1}^{T} r_t'(\mathbf{a}_t)$ (the second term in (3.3)), since a randomized policy is applied in each round.

Gap. Let Δ_t be the gap between the rewards of the best and the second best action in the decision set $\mathcal{D}_t$ as defined in Dani et al. (2008) which can be formally written as

$$\Delta_t = \min_{\mathbf{a} \in \mathcal{D}_t, \mathbf{a} \notin \mathcal{A}_t^*} \left(\langle \mathbf{a}_t^*, \boldsymbol{\mu}^* \rangle - \langle \mathbf{a}, \boldsymbol{\mu}^* \rangle \right). \tag{3.4}$$

where $\mathcal{A}_t^* = \mathrm{argmax}_{\mathbf{a} \in \mathcal{D}_t} \langle \mathbf{a}, \boldsymbol{\mu}^* \rangle$ and $\mathbf{a}_t^*$ is an arbitrary element in $\mathcal{A}_t^*$. Let Δ denotes the smallest gap $\min_{t \in [T]} \Delta_t$.

4 The Proposed Algorithm

In this section, we propose a variance-aware algorithm, Multi-level OFUL, in Algorithm 1, to tackle the corrupted linear contextual bandits problem. At the core of our algorithm is an action partition scheme to group historical selected actions and use them to select the future actions in different groups with different probabilities. Such a scheme is introduced to deal with the unknown corruption level. For simplicity, we denote $r_t(\mathbf{a}_t), \sigma_t(\mathbf{a}_t)$ in Section 3 by r_t, σ_t in our algorithm.

Main difficulty in our setting. We begin with the main difficulty that prevents us from applying existing algorithms to our setting. Consider a simpler setting where the agent knows the corruption level C in prior, and we have $\sigma_t = R$ for all t. Then we can apply OFUL (Abbasi-Yadkori et al., 2011) to solve our problem. In detail, in each round we estimate $\boldsymbol{\mu}^*$ by $\boldsymbol{\mu}_t$, which is the minimizer of the following ridge regression problem:

$$\boldsymbol{\mu}_t = \underset{\boldsymbol{\mu} \in \mathbb{R}^d}{\mathrm{argmin}}\, \lambda \|\boldsymbol{\mu}\|_2^2 + \sum_{i=1}^{t-1} [\langle \boldsymbol{\mu}, \mathbf{a}_i \rangle - r_i]^2. \tag{4.1}$$

Algorithm 1 Multi-level OFUL

1: Set the largest level of confidence sets: $\ell_{\max} \leftarrow \lceil \log_2 2T \rceil$.
2: For $\ell \in [\ell_{\max}]$, set $\mathbf{\Sigma}_{1,\ell} \leftarrow \lambda \mathbf{I}, \boldsymbol{\mu}_{1,\ell} \leftarrow \mathbf{0}, \mathbf{c}_{1,\ell} \leftarrow \mathbf{0}$.
3: Set $\mathbf{\Sigma}_1 \leftarrow \lambda \mathbf{I}, \boldsymbol{\mu}_1 \leftarrow \mathbf{0}, \mathbf{c}_1 \leftarrow \mathbf{0}$.
4: **for** $t = 1, \cdots, T$ **do**
5: Observe $\mathcal{D}_t$.
6: **for** $\ell = 1, \cdots, \ell_{\max}$ **do**
7: Set $\beta_{t,\ell}$ and $\gamma_{t,\ell}$ as defined in (4.5) and (4.6).
8: $\mathcal{C}'_{t,\ell} \leftarrow \{\boldsymbol{\mu} | \|\boldsymbol{\mu} - \boldsymbol{\mu}_t\|_{\mathbf{\Sigma}_t} \le \beta_{t,\ell}\} \cap \left\{\boldsymbol{\mu} | \|\boldsymbol{\mu} - \boldsymbol{\mu}_{t,\ell}\|_{\mathbf{\Sigma}_{t,\ell}} \le \gamma_{t,\ell}\right\}$.
9: $\mathcal{C}_{t,\ell} \leftarrow \begin{cases} \mathcal{C}'_{t,\ell}, & \mathcal{C}'_{t,\ell} \neq \varnothing \\ \mathcal{C}_{t,\ell+1}, & \text{otherwise} \end{cases}$.
10: **end for**
11: Set $f(t) = \begin{cases} \ell & \text{with probability } 2^{-\ell} \quad 1 < \ell \le \ell_{\max} \\ 1 & \text{otherwise} \end{cases}$.
12: Select $\mathbf{a}_t \leftarrow \mathrm{argmax}_{\mathbf{a} \in \mathcal{D}_t} \max_{\boldsymbol{\mu} \in \mathcal{C}_{t,f(t)}} \langle \boldsymbol{\mu}, \mathbf{a} \rangle$ and observe r_t, σ_t.
13: Set $\overline{\sigma}_t = \max\{(R+1)/\sqrt{d}, \sigma_t\}$.
14: $\mathbf{\Sigma}_{t+1} \leftarrow \mathbf{\Sigma}_t + \mathbf{a}_t \mathbf{a}_t^\top / \overline{\sigma}_t^2, \mathbf{c}_{t+1} \leftarrow \mathbf{c}_t + r_t \mathbf{a}_t / \overline{\sigma}_t^2, \boldsymbol{\mu}_{t+1} \leftarrow \mathbf{\Sigma}_{t+1}^{-1} \mathbf{c}_{t+1}$.
15: **for** $\ell \neq f(t)$ **do**
16: $\mathbf{\Sigma}_{t+1,\ell} \leftarrow \mathbf{\Sigma}_{t,\ell}, \mathbf{c}_{t+1,\ell} \leftarrow \mathbf{c}_{t,\ell}, \boldsymbol{\mu}_{t+1,\ell} \leftarrow \boldsymbol{\mu}_{t,\ell}$.
17: **end for**
18: $\mathbf{\Sigma}_{t+1,f(t)} \leftarrow \mathbf{\Sigma}_{t,f(t)} + \mathbf{a}_t \mathbf{a}_t^\top / \overline{\sigma}_t^2, \mathbf{c}_{t+1,f(t)} \leftarrow \mathbf{c}_{t,f(t)} + r_t \mathbf{a}_t / \overline{\sigma}_t^2$.
19: $\boldsymbol{\mu}_{t+1,f(t)} \leftarrow \mathbf{\Sigma}_{t+1,f(t)}^{-1} \mathbf{c}_{t+1,f(t)}$.
20: **end for**

By slightly modifying the self-normalized martingale concentration inequality proposed in Abbasi-Yadkori et al. (2011), we can conclude that $\boldsymbol{\mu}^*$ belongs to the ellipsoid $\|\boldsymbol{\mu} - \boldsymbol{\mu}_t\|_{\mathbf{\Sigma}_t^{-1}} \le \beta_t$ with high probability, where $\beta_t = \widetilde{O}(R\sqrt{d} + C\sqrt{d})$. Such a confidence bound leads to a final regret which has a polynomial dependence on C. However, such a simple approach have two limitations. First, the agent does not know C apriori in our setting, thus it is impossible to set β_t to be dependent on C. Second, vanilla ridge regression estimator does not consider different variances σ_t in each round, thus it only gives a very conservative estimation.

Action partition scheme. To address the unknown C issue, besides the original estimator $\boldsymbol{\mu}_t$ which uses all previous data, Algorithm 1 maintains several additional learners to learn $\boldsymbol{\mu}^*$ at different accuracy level simultaneously, and it *randomly* selects one of the learners with different probabilities at each round. Such a "parallel learning" idea is inspired by Lykouris et al. (2018). In detail, we partition the observed data into $\ell_{\max}$ levels indexed by $[\ell_{\max}]$ and maintain $\ell_{\max}$ sub-sampled estimators $\boldsymbol{\mu}_{t,1}, \cdots, \boldsymbol{\mu}_{t,\ell_{\max}}$. According to line 11, the observed data in round t goes into level ℓ with probability $2^{-\ell}$ if $1 < \ell \le \ell_{\max}$ and it goes to level 1 with probability $1 - \sum_{\ell=2}^{\ell_{\max}} 2^{-\ell} = 1/2 + 2^{-\ell_{\max}}$. The intuition is that if $2^\ell \ge C$, then the corruption level experienced by level ℓ

$$\text{Corruption}_{t,\ell} = \sum_{i=1}^{t} \frac{\mathbb{1}(f(i) = \ell)}{R+1} \cdot \sup_{\mathbf{a} \in \mathcal{D}_i} |r_i(\mathbf{a}) - r'_i(\mathbf{a})| \tag{4.2}$$

can be bounded by some quantity that is *independent of* C. That says, the individual learners whose level is greater than $\log C$ can learn $\boldsymbol{\mu}^*$ successfully, even with the corruption. For the learners whose level is less than $\log C$, we can also control the error by controlling the probability for the agent to select them.

Weighted regression estimator. After introducing the partition scheme, we still need to deal with the varying variance (heteroscedastic) case. Similar to (Kirschner and Krause, 2018; Zhou et al., 2020), we proposed the following *weighted ridge regression estimator*, which incorporates the variance information of the rewards into estimation:

$$\boldsymbol{\mu}_t = \underset{\boldsymbol{\mu} \in \mathbb{R}^d}{\mathrm{argmin}}\, \lambda \|\boldsymbol{\mu}\|_2^2 + \sum_{i=1}^{t-1} [\langle \boldsymbol{\mu}, \mathbf{a}_i \rangle - r_i]^2 / \overline{\sigma}_i^2. \tag{4.3}$$

Here $\overline{\sigma}_t$ is defined as the upper bound of the true variance σ_t in line 13. The closed-form solution to (4.3) is calculated at each round in line 14. The use of $\overline{\sigma}_t$, as we will show later, makes our estimator more efficient in the heteroscedastic case. Meanwhile, we also apply our weighted regression estimator to each individual learner, and their estimator $\boldsymbol{\mu}_{t,\ell}$ can be written as follows:

$$\boldsymbol{\mu}_{t,\ell} = \operatorname*{argmin}_{\boldsymbol{\mu}\in\mathbb{R}^d} \lambda\|\boldsymbol{\mu}\|_2^2 + \sum_{i=1}^{t-1} \mathbb{1}(f(i)=\ell)\cdot[\langle\boldsymbol{\mu},\mathbf{a}_i\rangle - r_i]^2/\overline{\sigma}_i^2. \tag{4.4}$$

The closed-form solution to (4.4) is calculated at each round in lines 15–20.

Final Multi-Level confidence sets. With the estimators $\boldsymbol{\mu}_t, \boldsymbol{\mu}_{t,1}, \cdots, \boldsymbol{\mu}_{t,\ell_{\max}}$ at the beginning of round t, we define a cascade of candidate confidence sets as in lines 6–10, where

$$\beta_{t,\ell} = 8\sqrt{d\log\frac{(R+1)^2\lambda + tA^2}{(R+1)^2\lambda}\log(4t^2/\delta)} + 4\sqrt{d}\log(4t^2/\delta) + 2^\ell\sqrt{d} + \sqrt{\lambda}B, \tag{4.5}$$

$$\gamma_{t,\ell} = 8\sqrt{d\log\frac{(R+1)^2\lambda + tA^2}{(R+1)^2\lambda}\log(8t^2T/\delta)} + 4\sqrt{d}\log(8t^2T/\delta) + \overline{C}_\ell\sqrt{d} + \sqrt{\lambda}B, \tag{4.6}$$

with $\overline{C}_\ell = \log(2\ell^2/\delta) + 3$. For simplicity, we define

$$\ell^* = \max\{2, \lceil\log_2 C\rceil\} \tag{4.7}$$

as an important threshold in our later proof for regret bound analysis. Later we will prove that $\mathcal{C}_{t,\ell}$ contains $\boldsymbol{\mu}^*$ for all $\ell \geq \ell^*$, $t \geq 1$ with high probability.

Note that each candidate confidence set can be written as the intersection of two ellipsoids. The intuition behind our construction of candidate confidence sets is that we hope that $\mathcal{C}_{t,\ell}$ is robust enough to handle the 2^ℓ-corrupted case, i.e., $\boldsymbol{\mu}^* \in \mathcal{C}_{t,\ell}$ with high probability. To achieve this, the first ellipsoid makes use of the global information and the "radius" $\beta_{t,\ell}$ need to contain a factor of 2^ℓ to tolerate a corruption level of 2^ℓ, and the second ellipsoid makes use of the observed data in level ℓ since this level only contain a few times of corruptions in 2^ℓ-corrupted case.

Action selection. With the candidate confidence sets, we use line 11 to randomly decide one confidence set and select an action based on the optimism-in-the-face-of-uncertainty (OFU) principle in line 12. Then we update the estimators for the next round $t+1$.

Remark 4.1. Our algorithm shares a similar strategy for partitioning the observed data with the algorithm in Lykouris et al. (2018) but note that there is a major difference in that: Lykouris et al. (2018) regard the partition scheme as a "layer structure", i.e., their algorithm further uses different estimators in layers of parallel learners and do action elimination layer by layer in each round. In contrast, the sub-sampled estimators in our algorithm are used independently, i.e., the selected action only relies on one of the partitions. As a result, Algorithm 1 does not need to do action elimination, thus is capable of handling the cases where the number of actions is huge or even infinite.

5 Main Results

In this section we present our main theorem, which establishes the regret bound for Multi-level OFUL.

Theorem 5.1. Set $\lambda = 1/B^2$. Suppose that $C = \Omega(1)$, $R = \Omega(1)$, for all $t \geq 1$ and all $\mathbf{a} \in \mathcal{D}_t$, $\langle\mathbf{a}, \boldsymbol{\mu}^*\rangle \in [-1, 1]$. Then with probability at least $1 - 3\delta$, the regret of Algorithm 1 is bounded as follows:

$$\mathbf{Regret}(T) = \widetilde{O}\left(C^2 d\sqrt{\sum_{t=1}^T \sigma_t^2} + C^2\sqrt{dT} + CR\sqrt{dT}\right).$$

Remark 5.2. When $\sigma_t, R = \Omega(1)$, the regret bound in Theorem 5.1 matches the regret bound of OFUL proposed in Zhou et al. (2020) when the corruption level C is a constant.

Remark 5.3. Compared with the $\widetilde{O}(d\sqrt{T} + C)$ result in Lee et al. (2021), our result has a multiplicative quadratic dependence on C, which seems to be worse. However, we want to emphasize that we focus on a more challenging contextual bandits setting where the decision sets $\mathcal{D}_t$ at each round are not identical, which is different from that in Lee et al. (2021). Therefore, our result and that in Lee et al. (2021) are not directly comparable.

Remark 5.4. Note that this instance-independent regret upper bound also holds in a stronger model than the one described in Section 3, where the adversary can even decide the decision set $\mathcal{D}_t$ at each round t since our regret bound can hold without any assumption on the decision sets.

Corollary 5.5. Under the same conditions as in Theorem 5.1, if σ_t given by the environment are all R, the regret of Algorithm 1 is bounded by:

$$\mathbf{Regret}(T) = \widetilde{O}\left(C^2 dR\sqrt{T}\right).$$

We also provide a gap-dependent regret bound.

Theorem 5.6. Suppose that $C = \Omega(1)$, $R = \Omega(1)$, for all $t \geq 1$ and all $\mathbf{a} \in \mathcal{D}_t$, $\langle \mathbf{a}, \boldsymbol{\mu}^* \rangle \in [-1, 1]$. Then with probability at least $1 - 3\delta$, the regret of Algorithm 1 is bounded as follows:

$$\mathbf{Regret}(T) = \widetilde{O}\left(\frac{1}{\Delta} \cdot C^2 R^2 d + \frac{1}{\Delta} \cdot d^2 C^2 \max_{t \in [T]} \sigma_t^2\right).$$

Remark 5.7. Theorem 5.6 automatically suggests an $\widetilde{O}(R^2 d^2 C^2/\Delta)$ regret bound, by the fact $\sigma_t = O(R)$. Compared with previous result $\widetilde{O}(d^{5/2}C/\Delta + d^6/\Delta^2)$ (Lee et al., 2021), our result has a better dependence on the dimension d but a worse dependence on the corruption level C. As Remark 5.3 suggests, we focus on a more challenging contextual bandits setting, and the worse dependence on C might be due to this.

6 Proof Outline

First we have the following lemma which is a corruption-tolerant variant of Bernstein inequality for self-normalized vector-valued martingales introduced in Zhou et al. (2020).

Lemma 6.1 (Bernstein inequality for vector-valued martingales with corruptions)**.** Let $\{\mathcal{G}_t\}_{t=1}^\infty$ be a filtration, $\{\mathbf{x}_t, \eta_t\}_{t \geq 1}$ a stochastic process so that $\mathbf{x}_t \in \mathbb{R}^d$ is $\mathcal{G}_t$-measurable and $\eta_t \in \mathbb{R}$ is $\mathcal{G}_{t+1}$-measurable. Fix $R, L, \sigma, \lambda > 0$, $\boldsymbol{\mu}^* \in \mathbb{R}^d$. For $t \geq 1$ let $y_t^{\text{stoch}} = \langle \boldsymbol{\mu}^*, \mathbf{x}_t \rangle + \eta_t$ and suppose that $\eta_t, \mathbf{x}_t$ also satisfy

$$|\eta_t| \leq R, \mathbb{E}[\eta_t|\mathcal{G}_t] = 0, \mathbb{E}[\eta_t^2|\mathcal{G}_t] \leq \sigma^2, \|\mathbf{x}_t\|_2 \leq L.$$

Suppose $\{y_t\}$ is a sequence such that $\sum_{i=1}^t |y_i - y_i^{\text{stoch}}| = C(t)$ for all $t \geq 1$. Then, for any $0 < \delta < 1$, with probability at least $1 - \delta$ we have $\forall t > 0$,

$$\|\boldsymbol{\mu}_t - \boldsymbol{\mu}^*\|_{\mathbf{Z}_t} \leq \beta_t + C(t) + \sqrt{\lambda}\|\boldsymbol{\mu}^*\|_2,$$

where for $t \geq 1$, $\boldsymbol{\mu}_t = \mathbf{Z}_t^{-1}\mathbf{b}_t$, $\mathbf{Z}_t = \lambda \mathbf{I} + \sum_{i=1}^t \mathbf{x}_i \mathbf{x}_i^\top$, $\mathbf{b}_t = \sum_{i=1}^t y_i \mathbf{x}_i$, and

$$\beta_t = 8\sigma\sqrt{d \log \frac{d\lambda + tL^2}{d\lambda} \log(4t^2/\delta)} + 4R\log(4t^2/\delta).$$

Next, we have that with high probability, all the levels satisfying $\ell \geq \ell^*$ are only influenced by a limited amount of corruptions as mentioned in Section 4.

Lemma 6.2. Let $\text{Corruption}_{t,\ell}$ be defined in (4.2). Then we have with probability at least $1 - \delta$, for all $\ell \geq \ell^*, t \geq 1$:

$$\text{Corruption}_{t,\ell} \leq \overline{C}_\ell = \log(2\ell^2/\delta) + 3.$$

We denote by $\mathcal{E}_{\text{sub}}$ the event that the above inequality holds.

We define the following event to further show that our candidate confidence sets with $\ell \geq \ell^*$ are "robust" enough, i.e., $\mathcal{C}_{t,\ell}$ contains $\boldsymbol{\mu}^*$ with high probability.

Definition 6.3. Let ℓ^* be defined in (4.7). We introduce the event $\mathcal{E}_1$ as follows.

$$\mathcal{E}_1 := \left\{\forall \ell \geq \ell^* \text{ and } t \geq 1, \|\boldsymbol{\mu}^* - \boldsymbol{\mu}_t\|_{\boldsymbol{\Sigma}_t} \leq \beta_{t,\ell} \text{ and } \|\boldsymbol{\mu}^* - \boldsymbol{\mu}_{t,\ell}\|_{\boldsymbol{\Sigma}_{t,\ell}} \leq \gamma_{t,\ell}\right\}. \tag{6.1}$$

where $\beta_{t,\ell}$, $\gamma_{t,\ell}$ are defined in (4.5) and (4.6).

Next lemma suggests that the event $\mathcal{E}_1$ happens with high probability.

Lemma 6.4. Let $\mathcal{E}_1$ be defined in (6.1). For any $0 < \delta < 1/3$, we have $\mathbb{P}(\mathcal{E}_1) \geq 1 - 3\delta$.

For simplicity, we define $\mathbf{a}_{t,\ell} = \mathrm{argmax}_{\mathbf{a}\in\mathcal{D}_t} \max_{\boldsymbol{\mu}\in\mathcal{C}_{t,\ell}} \langle \boldsymbol{\mu}, \mathbf{a}\rangle$ for each level ℓ. $\mathbf{a}_t$ can be seen as an action vector randomly chosen from $\mathbf{a}_{t,\ell}$, $\ell \in [\ell_{\max}]$. Next two lemmas suggest that under event $\mathcal{E}_1$, at each round, the gap between the optimal reward and the selected reward can be upper bounded by some bonus terms related to $\mathbf{a}_{t,\ell}$.

Lemma 6.5. On event $\mathcal{E}_1$, if $f(t) \leq \ell^*$, we have $\langle \mathbf{a}_t^* - \mathbf{a}_t, \boldsymbol{\mu}^*\rangle \leq 2\beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}} + 2\beta_{t,\ell^*}\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}}$.

Lemma 6.6. On event $\mathcal{E}_1$, if $f(t) = \ell > \ell^*$, we have $\langle \mathbf{a}_t^* - \mathbf{a}_t, \boldsymbol{\mu}^*\rangle \leq 2\gamma_{t,\ell}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_{t,\ell}^{-1}}$.

Now we provide the proof sketch of Theorem 5.1.

Proof sketch of Theorem 5.1 . Suppose $\mathcal{E}_1$ occurs. The main idea to bound the regret is to decompose the total rounds $[T]$ into two non-overlapping parts, based on which individual learner is selected at that round. In detail, we have

$$\begin{aligned}
\mathbf{Regret}(T) &= \mathbb{E}\left[\sum_{t=1}^{T}\left(\langle \mathbf{a}_t^*, \boldsymbol{\mu}^*\rangle - \langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle\right)\right] \\
&= \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(f(t)\leq \ell^*)\left(\langle \mathbf{a}_t^*, \boldsymbol{\mu}^*\rangle - \langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle\right)\right]}_{I_1} \\
&\quad + \sum_{\ell=\ell^*+1}^{\ell_{\max}} \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(f(t)=\ell)\left(\langle \mathbf{a}_t^*, \boldsymbol{\mu}^*\rangle - \langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle\right)\right]}_{I_2(\ell)}. \qquad (6.2)
\end{aligned}$$

Here I_1 represents the regret where the the "low-level" learner is selected, and the corruption level is beyond the learner level. In this case, by Lemma 6.5, we can directly show that

$$I_1 \leq \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(f(t)\leq \ell^*)\min\left\{2, 2\beta_{t,\ell^*}\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}} + 2\beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}\right\}\right]. \qquad (6.3)$$

We further bound (6.3). Let $\mathcal{F}_t$ be the σ-algebra generated by $\mathbf{a}_s, r_s, \sigma_s, f(s)$ for $s \leq t-1$. Then by the property of our partition scheme (note that $\mathbb{P}(f(t) = \ell^*) = 2^{-\ell^*}$), we can show that $\mathbb{E}\left[\mathbb{1}(f(t)\leq \ell^*)\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}}|\mathcal{F}_t\right] \leq 2^{\ell^*}\mathbb{E}\left[\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}|\mathcal{F}_t\right]$. Therefore, we can further bound I_1 by

$$I_1 \leq 4\cdot 2^{\ell^*}\underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\min\left\{2, \beta_{T,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}\right\}\right]}_{I_3}. \qquad (6.4)$$

To further bound I_3, we split $[T]$ into 2 parts, $\mathcal{I}_1 = \{t \in [T] | \|\mathbf{a}_t/\overline{\sigma}_t\|_{\boldsymbol{\Sigma}_t^{-1}} > 1\}, \mathcal{I}_2 = \{t \in [T] | \|\mathbf{a}_t/\overline{\sigma}_t\|_{\boldsymbol{\Sigma}_t^{-1}} \leq 1\}$. To bound $\mathcal{I}_1$ part, the intuition is that the cardinality of $\mathcal{I}_1$ is bounded, and the sum of terms with $t \in \mathcal{I}_2$ can be bounded using Cauchy-Schwarz inequality.

$$\sum_{t\in\mathcal{I}_1}\min\left\{2, \beta_{T,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}\right\} \leq 2|\mathcal{I}_1| \leq 2\sum_{t=1}^{T}\min\left\{1, \|\mathbf{a}_t/\overline{\sigma}_t\|^2_{\boldsymbol{\Sigma}_t^{-1}}\right\} \leq 4d\log\frac{(R+1)^2\lambda + TA^2}{(R+1)^2\lambda}, \qquad (6.5)$$

where the first inequality holds since $\min\left\{2, \beta_{T,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}\right\} \leq 2$, the second inequality follows from the definition of $\mathcal{I}_1$, and the third inequality holds by Lemma C.2. To bound $\mathcal{I}_2$ part, we have

$$\sum_{t\in\mathcal{I}_2}\min\left\{2, \beta_{T,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}\right\} \leq \beta_{T,\ell^*}\sqrt{\sum_{t\in\mathcal{I}_2}\overline{\sigma}_t^2}\cdot\sqrt{\sum_{t\in\mathcal{I}_2}\min\left\{1, \|\mathbf{a}_t/\overline{\sigma}_t\|^2_{\boldsymbol{\Sigma}_t^{-1}}\right\}}$$

$$\leq \beta_{T,\ell^*}\sqrt{(R+1)^2T/d+\sum_{t=1}^{T}\sigma_t^2}\cdot\sqrt{2d\log\frac{(R+1)^2\lambda+TA^2}{(R+1)^2\lambda}}, \tag{6.6}$$

where the first inequality follows from Cauchy-Schwarz inequality, the second inequality follows from the definition of $\overline{\sigma}_t$ and Lemma C.2.

Substituting (6.5) and (6.6) into (6.3), we have

$$I_1 = \widetilde{O}\left(C^2 d\sqrt{\sum_{t=1}^{T}\sigma_t^2} + C^2\sqrt{dT} + CR\sqrt{dT}\right). \tag{6.7}$$

Now it remains to bound $I_2(\ell)$. By Lemma 6.6, we have

$$I_2(\ell) \leq 2\mathbb{E}\underbrace{\left[\sum_{t=1}^{T}\mathbb{1}(f(t)=\ell)\min\left\{1, \gamma_{t,\ell}\|\mathbf{a}_{t,\ell}\|_{\mathbf{\Sigma}_{t,\ell}^{-1}}\right\}\right]}_{I_4} = \widetilde{O}\left(R\sqrt{Td} + d\sqrt{\sum_{t=1}^{T}\sigma_t^2}\right), \tag{6.8}$$

where the second equality can be proved by an analysis similar to that of (6.5) and (6.6). Finally, substituting (6.7) and (6.8) into (6.2) completes our proof.

□

7 Conclusion and Future Work

In this paper, we have considered the linear contextual bandits problem in the presence of adversarial corruptions. We propose a Multi-level OFUL algorithm, which is provably robust to the adversarial attacks. We prove a gap-independent regret bound of $\widetilde{O}\left(C^2 d\sqrt{\sum_{t=1}^{T}\sigma_t^2} + C^2\sqrt{dT} + CR\sqrt{dT}\right)$ together with a gap-dependent bound of $\widetilde{O}\left(\frac{1}{\Delta}\cdot C^2R^2d + \frac{1}{\Delta}\cdot d^2C^2\max_{t\in[T]}\sigma_t^2\right)$.

We leave it as an open question that whether the multiplicative dependence on C^2 in the regret upper bounds can be removed without making additional assumptions in our setting.

References

ABBASI-YADKORI, Y., PÁL, D. and SZEPESVÁRI, C. (2011). Improved algorithms for linear stochastic bandits. In *NIPS*, vol. 11.

ABE, N., BIERMANN, A. W. and LONG, P. M. (2003). Reinforcement learning with immediate rewards and linear hypotheses. *Algorithmica* **37** 263–293.

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

AUER, P. and CHIANG, C.-K. (2016). An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits. In *Conference on Learning Theory*. PMLR.

BOGUNOVIC, I., LOSALKA, A., KRAUSE, A. and SCARLETT, J. (2021). Stochastic linear bandits robust to adversarial attacks. In *International Conference on Artificial Intelligence and Statistics*. PMLR.

BUBECK, S. and SLIVKINS, A. (2012). The best of both worlds: Stochastic and adversarial bandits. In *Conference on Learning Theory*. JMLR Workshop and Conference Proceedings.

CHU, W., LI, L., REYZIN, L. and SCHAPIRE, R. (2011). Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings.

Dani, V., Hayes, T. P. and Kakade, S. (2008). Stochastic linear optimization under bandit feedback. In *COLT*.

Deshpande, Y. and Montanari, A. (2012). Linear bandits in high dimension and recommendation systems. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE.

Garcelon, E., Roziere, B., Meunier, L., Tarbouriech, J., Teytaud, O., Lazaric, A. and Pirotta, M. (2020). Adversarial attacks on linear contextual bandits. *arXiv preprint arXiv:2002.03839* .

Gupta, A., Koren, T. and Talwar, K. (2019). Better algorithms for stochastic bandits with adversarial corruptions. In *Conference on Learning Theory*. PMLR.

Jhalani, T., Kant, V. and Dwivedi, P. (2016). A linear regression approach to multi-criteria recommender system. In *International Conference on Data Mining and Big Data*. Springer.

Jun, K.-S., Li, L., Ma, Y. and Zhu, X. J. (2018). Adversarial attacks on stochastic bandits. In *NeurIPS*.

Kannan, S., Morgenstern, J. H., Roth, A., Waggoner, B. and Wu, Z. (2018). A smoothed analysis of the greedy algorithm for the linear contextual bandit problem. In *NeurIPS*.

Kapoor, S., Patel, K. K. and Kar, P. (2019). Corruption-tolerant bandit learning. *Machine Learning* **108** 687–715.

Kirschner, J. and Krause, A. (2018). Information directed sampling and bandits with heteroscedastic noise. In *Conference On Learning Theory*. PMLR.

Lee, C.-W., Luo, H., Wei, C.-Y., Zhang, M. and Zhang, X. (2021). Achieving near instance-optimality and minimax-optimality in stochastic and adversarial linear bandits simultaneously. *arXiv preprint arXiv:2102.05858* .

Li, L., Chu, W., Langford, J. and Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*.

Li, Y., Lou, E. Y. and Shan, L. (2019a). Stochastic linear optimization with adversarial corruption. *arXiv preprint arXiv:1909.02109* .

Li, Y., Wang, Y. and Zhou, Y. (2019b). Nearly minimax-optimal regret for linearly parameterized bandits. In *Conference on Learning Theory*. PMLR.

Liu, F. and Shroff, N. (2019). Data poisoning attacks on stochastic bandits. In *International Conference on Machine Learning*. PMLR.

Lykouris, T., Mirrokni, V. and Paes Leme, R. (2018). Stochastic bandits robust to adversarial corruptions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*.

Neu, G. and Olkhovskaya, J. (2020). Efficient and robust algorithms for adversarial linear contextual bandits. In *Conference on Learning Theory*. PMLR.

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

Seldin, Y. and Lugosi, G. (2017). An improved parametrization and analysis of the exp3++ algorithm for stochastic and adversarial bandits. In *Conference on Learning Theory*. PMLR.

Seldin, Y. and Slivkins, A. (2014). One practical algorithm for both stochastic and adversarial bandits. In *International Conference on Machine Learning*. PMLR.

Villar, S. S., Bowden, J. and Wason, J. (2015). Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. *Statistical science: a review journal of the Institute of Mathematical Statistics* **30** 199.

Zhou, D., Gu, Q. and Szepesvari, C. (2020). Nearly minimax optimal reinforcement learning for linear mixture markov decision processes. *arXiv preprint arXiv:2012.08507* .

Zimmert, J. and Seldin, Y. (2019). An optimal algorithm for stochastic and adversarial bandits. In *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A] Our work studies the regret bounds for contextual linear bandits with corruption. That is a pure theoretical problem, thus it does not have any negative social impact.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Proof of the Main Results

A.1 Proof of Theorem 5.1

We first prove the following lemma which is a corruption-tolerant variant of Bernstein inequality for self-normalized vector-valued martingales introduced in Zhou et al. (2020).

Lemma A.1 (Restatement of Lemma 6.1)**.** Let $\{\mathcal{G}_t\}_{t=1}^{\infty}$ be a filtration, $\{\mathbf{x}_t, \eta_t\}_{t\geq 1}$ a stochastic process so that $\mathbf{x}_t \in \mathbb{R}^d$ is $\mathcal{G}_t$-measurable and $\eta_t \in \mathbb{R}$ is $\mathcal{G}_{t+1}$-measurable. Fix $R, L, \sigma, \lambda > 0$, $\boldsymbol{\mu}^* \in \mathbb{R}^d$. For $t \geq 1$ let $y_t^{\text{stoch}} = \langle \boldsymbol{\mu}^*, \mathbf{x}_t \rangle + \eta_t$ and suppose that $\eta_t, \mathbf{x}_t$ also satisfy

$$|\eta_t| \leq R, \mathbb{E}[\eta_t|\mathcal{G}_t] = 0, \mathbb{E}[\eta_t^2|\mathcal{G}_t] \leq \sigma^2, \|\mathbf{x}_t\|_2 \leq L.$$

Suppose $\{y_t\}$ is a sequence such that $\sum_{i=1}^t |y_i - y_i^{\text{stoch}}| = C(t)$ for all $t \geq 1$. Then, for any $0 < \delta < 1$, with probability at least $1 - \delta$ we have $\forall t > 0$,

$$\|\boldsymbol{\mu}_t - \boldsymbol{\mu}^*\|_{\mathbf{Z}_t} \leq \beta_t + C(t) + \sqrt{\lambda}\|\boldsymbol{\mu}^*\|_2,$$

where for $t \geq 1$, $\boldsymbol{\mu}_t = \mathbf{Z}_t^{-1}\mathbf{b}_t$, $\mathbf{Z}_t = \lambda \mathbf{I} + \sum_{i=1}^t \mathbf{x}_i \mathbf{x}_i^\top$, $\mathbf{b}_t = \sum_{i=1}^t y_i \mathbf{x}_i$, and

$$\beta_t = 8\sigma\sqrt{d \log \frac{d\lambda + tL^2}{d\lambda} \log(4t^2/\delta)} + 4R\log(4t^2/\delta).$$

Proof. See Appendix B.1. □

Then we prove that with high probability, all the level $\ell \geq \ell^*$ only influenced by limited amount of corruptions as mentioned in Section 4.

Lemma A.2 (Restatement of Lemma 6.2)**.** Let $\text{Corruption}_{t,\ell}$ be defined in (4.2). Then we have with probability at least $1 - \delta$, for all $\ell \geq \ell^*$, $t \geq 1$:

$$\text{Corruption}_{t,\ell} \leq \overline{C}_\ell = \log(2\ell^2/\delta) + 3.$$

We denote by $\mathcal{E}_{\text{sub}}$ the event that the above inequality holds.

Proof. The proof of this lemma is based on Lemma B.1 introduced by Lykouris et al. (2018); for details see Appendix B.2. □

We define the following event to further show that our candidate confidence sets with $\ell \geq \ell^*$ are "robust" enough, i.e. $\mathcal{C}_{t,\ell}$ contains $\boldsymbol{\mu}^*$ with high probability.

Definition A.3. Let ℓ^* be defined in (4.7). We introduce the event $\mathcal{E}_1$ as follows.

$$\mathcal{E}_1 := \left\{\forall \ell \geq \ell^* \text{ and } t \geq 1, \|\boldsymbol{\mu}^* - \boldsymbol{\mu}_t\|_{\boldsymbol{\Sigma}_t} \leq \beta_{t,\ell} \text{ and } \|\boldsymbol{\mu}^* - \boldsymbol{\mu}_{t,\ell}\|_{\boldsymbol{\Sigma}_{t,\ell}} \leq \gamma_{t,\ell}\right\}. \tag{A.1}$$

Recall that

$$\beta_{t,\ell} = 8\sqrt{d \log \frac{(R+1)^2\lambda + tA^2}{(R+1)^2\lambda} \log(4t^2/\delta)} + 4\sqrt{d}\log(4t^2/\delta) + 2^\ell\sqrt{d} + \sqrt{\lambda}B, \tag{A.2}$$

$$\gamma_{t,\ell} = 8\sqrt{d \log \frac{(R+1)^2\lambda + tA^2}{(R+1)^2\lambda} \log(8t^2T/\delta)} + 4\sqrt{d}\log(8t^2T/\delta) + \overline{C}_\ell\sqrt{d} + \sqrt{\lambda}B. \tag{A.3}$$

Lemma A.4 (Restatement of Lemma 6.4)**.** Let $\mathcal{E}_1$ be defined in (A.1). For any $0 < \delta < 1$, we have $\mathbb{P}(\mathcal{E}_1) \geq 1 - 3\delta$.

Proof. See Appendix B.3. □

Definition A.5. For simplicity, we define $\mathbf{a}_{t,\ell} = \text{argmax}_{\mathbf{a}\in\mathcal{D}_t} \max_{\boldsymbol{\mu}\in\mathcal{C}_{t,\ell}} \langle \boldsymbol{\mu}, \mathbf{a} \rangle$ for each level ℓ.

With this definition, $\mathbf{a}_t$ can be seen as an action vector randomly chosen from $\mathbf{a}_{t,\ell}$, $\ell \in [\ell_{\max}]$. In the following part of this section, we show how to derive the instance-independent regret upper bound using this notation.

Lemma A.6 (Restatement of Lemms 6.5)**.** Suppose $\mathcal{E}_1$ occurs. If $f(t) \le \ell^*$, we have $\langle \mathbf{a}_t^* - \mathbf{a}_t, \boldsymbol{\mu}^* \rangle \le 2\beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}} + 2\beta_{t,\ell^*}\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}}$.

Lemma A.7 (Restatement of Lemms 6.6)**.** On event $\mathcal{E}_1$, if $f(t) = \ell > \ell^*$, we have $\langle \mathbf{a}_t^* - \mathbf{a}_t, \boldsymbol{\mu}^* \rangle \le 2\gamma_{t,\ell}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_{t,\ell}^{-1}}$.

Proof of Theorem 5.1. Suppose $\mathcal{E}_1$ occurs. We divide regret into two parts,

$$\begin{aligned}
\mathbf{Regret}(T) &= \mathbb{E}\left[\sum_{t=1}^{T}\left(\langle \mathbf{a}_t^*, \boldsymbol{\mu}^*\rangle - \langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle\right)\right] \\
&= \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(f(t) \le \ell^*)\left(\langle \mathbf{a}_t^*, \boldsymbol{\mu}^*\rangle - \langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle\right)\right]}_{I_1} \\
&\quad + \sum_{\ell=\ell^*+1}^{\ell_{\max}} \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(f(t) = \ell)\left(\langle \mathbf{a}_t^*, \boldsymbol{\mu}^*\rangle - \langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle\right)\right]}_{I_2(\ell)},
\end{aligned} \tag{A.4}$$

where the first equality holds by definition in (3.3).

By Lemma B.3, we have

$$I_1 \le \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(f(t) \le \ell^*)\min\left\{2, 2\beta_{t,\ell^*}\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}} + 2\beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}\right\}\right]. \tag{A.5}$$

Let $\mathcal{F}_t$ be the σ-algebra generated by $\mathbf{a}_s, r_s, \sigma_s, f(s)$ for $s \le t-1$. Note that

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}(f(t) \le \ell^*)\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}}|\mathcal{F}_t\right] &= \mathbb{P}(f(t) \le \ell^*)\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}} \\
&\le 2^{\ell^*}\mathbb{P}(f(t) = \ell^*)\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}} \\
&\le 2^{\ell^*}\mathbb{E}\left[\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}|\mathcal{F}_t\right],
\end{aligned} \tag{A.6}$$

where the first equality holds since $\mathbf{a}_{t,\ell^*}$ and $\boldsymbol{\Sigma}_t$ is deterministic given $\mathcal{F}_t$, the first inequality holds since $\mathbb{P}(f(t) = \ell^*) = 2^{-\ell^*}$, the last inequality holds due to the fact that $\mathbb{P}(f(t) = \ell^*)\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}} = \mathbb{E}\left[\mathbb{1}(f(t) = \ell^*)\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}|\mathcal{F}_t\right]$.

Substituting (A.6) into (A.5), we have

$$I_1 \le 2^{\ell^*}\mathbb{E}\left[\sum_{t=1}^{T}4\min\{\beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}, 2\}\right] \le 4\cdot 2^{\ell^*}\,\mathbb{E}\underbrace{\left[\sum_{t=1}^{T}\min\left\{2, \beta_{T,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}\right\}\right]}_{I_3} \tag{A.7}$$

We split $[T]$ into 2 parts to bound I_3.

Let $\mathcal{I}_1 = \{t \in [T] | \|\mathbf{a}_t/\overline{\sigma}_t\|_{\boldsymbol{\Sigma}_t^{-1}} > 1\}, \mathcal{I}_2 = \{t \in [T] | \|\mathbf{a}_t/\overline{\sigma}_t\|_{\boldsymbol{\Sigma}_t^{-1}} \le 1\}$.

$$\sum_{t\in\mathcal{I}_1}\min\left\{2, \beta_{T,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}\right\} \le 2|\mathcal{I}_1| \le 2\sum_{t=1}^{T}\min\left\{1, \|\mathbf{a}_t/\overline{\sigma}_t\|_{\boldsymbol{\Sigma}_t^{-1}}^2\right\} \le 4d\log\frac{(R+1)^2\lambda + TA^2}{(R+1)^2\lambda}, \tag{A.8}$$

where the first inequality holds since $\min\left\{2, \beta_{T,\ell^*}\|\mathbf{a}_t\|_{\mathbf{\Sigma}_t^{-1}}\right\} \le 2$, the second inequality follows from the definition of $\mathcal{I}_1$, the third inequality holds by Lemma C.2.

$$\begin{aligned}\sum_{t\in\mathcal{I}_2}\min\left\{2, \beta_{T,\ell^*}\|\mathbf{a}_t\|_{\mathbf{\Sigma}_t^{-1}}\right\} &\le \beta_{T,\ell^*}\sqrt{\sum_{t\in\mathcal{I}_2}\overline{\sigma}_t^2}\cdot\sqrt{\sum_{t\in\mathcal{I}_2}\min\left\{1, \|\mathbf{a}_t/\overline{\sigma}_t\|^2_{\mathbf{\Sigma}_t^{-1}}\right\}}\\ &\le \beta_{T,\ell^*}\sqrt{(R+1)^2T/d+\sum_{t=1}^T\sigma_t^2}\cdot\sqrt{2d\log\frac{(R+1)^2\lambda+TA^2}{(R+1)^2\lambda}},\end{aligned} \tag{A.9}$$

where the first inequality follows from Cauchy-Schwarz inequality, the second inequality follows from the definition of $\overline{\sigma}_t$ and Lemma C.2.

Substituting (A.8) and (A.9) into (A.7), we have

$$I_1 = \widetilde{O}\left(C^2d\sqrt{\sum_{t=1}^T\sigma_t^2}+C^2\sqrt{dT}+CR\sqrt{dT}\right). \tag{A.10}$$

By Lemma B.4,

$$\begin{aligned}I_2(\ell) &\le \mathbb{E}\left[\sum_{t=1}^T\mathbb{1}(f(t)=\ell)\min\left\{2, 2\gamma_{t,\ell}\|\mathbf{a}_{t,\ell}\|_{\mathbf{\Sigma}_{t,\ell}^{-1}}\right\}\right]\\ &\le 2\underbrace{\mathbb{E}\left[\sum_{t=1}^T\mathbb{1}(f(t)=\ell)\min\left\{1, \gamma_{t,\ell}\|\mathbf{a}_{t,\ell}\|_{\mathbf{\Sigma}_{t,\ell}^{-1}}\right\}\right]}_{I_4}\end{aligned} \tag{A.11}$$

Again, we divide $[T]$ into two parts. Let $\mathcal{J}_1 = \{t \in [T] | \|\mathbf{a}_{t,\ell}/\overline{\sigma}_t\|_{\mathbf{\Sigma}_{t,\ell}^{-1}} > 1\}, \mathcal{J}_2 = \{t \in [T] | \|\mathbf{a}_{t,\ell}/\overline{\sigma}_t\|_{\mathbf{\Sigma}_{t,\ell}^{-1}} \le 1\}$.

$$\begin{aligned}\sum_{t\in\mathcal{J}_1}\mathbb{1}(f(t)=\ell)\min\left\{1, \gamma_{t,\ell}\|\mathbf{a}_{t,\ell}\|_{\mathbf{\Sigma}_{t,\ell}^{-1}}\right\} &\le \sum_{t\in\mathcal{J}_1}\mathbb{1}(f(t)=\ell)\cdot 1\\ &\le \sum_{t=1}^T\mathbb{1}(f(t)=\ell)\min\left\{1, \|\mathbf{a}_{t,\ell}\|^2_{\mathbf{\Sigma}_{t,\ell}^{-1}}\right\}\\ &\le 2d\log\frac{(R+1)^2\lambda+TA^2}{(R+1)^2\lambda},\end{aligned} \tag{A.12}$$

where the second inequality follows from the definition of $\mathcal{J}_1$, the second inequality holds due to Lemma C.2.

$$\begin{aligned}&\sum_{t\in\mathcal{J}_2}\mathbb{1}(f(t)=\ell)\min\left\{1, \gamma_{t,\ell}\|\mathbf{a}_{t,\ell}\|_{\mathbf{\Sigma}_{t,\ell}^{-1}}\right\}\\ &\le \gamma_{T,\ell}\sqrt{\sum_{t=1}^T\overline{\sigma}_t^2}\sqrt{\sum_{t\in\mathcal{J}_2}\min\left\{1, \|\mathbf{a}_t/\overline{\sigma}_t\|^2_{\mathbf{\Sigma}_{t,\ell}^{-1}}\right\}}\\ &\le \gamma_{T,\ell}\sqrt{(R+1)^2T/d+\sum_{t=1}^T\sigma_t^2}\sqrt{2d\log\frac{(R+1)^2\lambda+TA^2}{(R+1)^2\lambda}},\end{aligned} \tag{A.13}$$

where the first inequality follows from Cauchy-Schwarz inequality, the second inequality follows from the definition of $\overline{\sigma}_t$ and Lemma C.2.

Substituting (A.12) and (A.13) into (A.11), we have

$$I_2(\ell) \leq 4d \log \frac{(R+1)^2\lambda + TA^2}{(R+1)^2\lambda} + 2\gamma_{T,\ell}\sqrt{(R+1)^2T/d + \sum_{t=1}^{T}\sigma_t^2}\sqrt{2d \log \frac{(R+1)^2\lambda + TA^2}{(R+1)^2\lambda}} \tag{A.14}$$

$$I_2(\ell) = \widetilde{O}\left(R\sqrt{Td} + d\sqrt{\sum_{t=1}^{T}\sigma_t^2}\right). \tag{A.15}$$

Substituting (A.10) and (A.15) into (A.4), we have

$$\mathbf{Regret}(T) = \widetilde{O}\left(C^2 d\sqrt{\sum_{t=1}^{T}\sigma_t^2} + C^2\sqrt{dT} + CR\sqrt{dT}\right).$$

$\square$

A.2 Proof of Theorem 5.6

Proof of Theorem 5.6. First we decompose the regret as follows.

$$\begin{aligned}
\mathbf{Regret}(T) &= \mathbb{E}\left[\sum_{t=1}^{T}\left(\langle \mathbf{a}_t^*, \boldsymbol{\mu}^*\rangle - \langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle\right)\right] \\
&\leq \underbrace{\frac{1}{\Delta}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(f(t) \leq \ell^*)\left(\langle \mathbf{a}_t^*, \boldsymbol{\mu}^*\rangle - \langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle\right)^2\right]}_{I_1} \\
&\quad + \sum_{\ell=\ell^*+1}^{\ell_{\max}}\frac{1}{\Delta}\underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(f(t) = \ell)\left(\langle \mathbf{a}_t^*, \boldsymbol{\mu}^*\rangle - \langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle\right)^2\right]}_{I_2(\ell)},
\end{aligned} \tag{A.16}$$

where the first equality holds due to the definition in (3.3), the last inequality follows from the fact that either $\langle \mathbf{a}_t^*, \boldsymbol{\mu}^*\rangle - \langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle = 0$ or $\overline{\Delta}_T \leq \langle \mathbf{a}_t^*, \boldsymbol{\mu}^*\rangle - \langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle$. To bound I_1, we have

$$\begin{aligned}
I_1 &\leq \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}(f(t) \leq \ell^*)\min\left\{4, \left(2\beta_{t,\ell^*}\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}} + 2\beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}\right)^2\right\}\right] \\
&\leq 2^{\ell^*}\underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\min\left\{4, 16\beta_{t,\ell^*}^2\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}^2\right\}\right]}_{I_3},
\end{aligned} \tag{A.17}$$

where the first inequality holds due to Lemma A.6 and the second inequality follows from a similar argument as (A.6). To further bound I_3, we decompose $[T]$ into two non-overlapping sets: $\mathcal{I}_1 = \{t \in [T] | \|\mathbf{a}_t/\overline{\sigma}_t\|_{\boldsymbol{\Sigma}_t^{-1}} > 1\}, \mathcal{I}_2 = \{t \in [T] | \|\mathbf{a}_t/\overline{\sigma}_t\|_{\boldsymbol{\Sigma}_t^{-1}} \leq 1\}$. For $\mathcal{I}_1$, we have

$$\begin{aligned}
\sum_{t\in\mathcal{I}_1}\min\left\{4, 16\beta_{t,\ell^*}^2\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}^2\right\} &\leq 4|\mathcal{I}_1| \\
&\leq 4\sum_{t=1}^{T}\min\left\{1, \|\mathbf{a}_t/\overline{\sigma}_t\|_{\boldsymbol{\Sigma}_t^{-1}}^2\right\}
\end{aligned}$$

$$\leq 8d \log \frac{(R+1)^2\lambda + TA^2}{(R+1)^2\lambda}, \tag{A.18}$$

where the third inequality holds due to Lemma C.2. For $\mathcal{I}_2$, we have

$$\begin{aligned}\sum_{t\in\mathcal{I}_2} \min\left\{4, 9\beta_{t,\ell^*}^2 \|\mathbf{a}_t\|^2_{\boldsymbol{\Sigma}_t^{-1}}\right\} &\leq 16\beta_{T,\ell^*}^2 \max_{t\in[T]} \overline{\sigma}_t^2 \sum_{t\in\mathcal{I}_2} \min\left\{1, \|\mathbf{a}_t/\overline{\sigma}_t\|^2_{\boldsymbol{\Sigma}_t^{-1}}\right\} \\ &\leq 32\beta_{T,\ell^*}^2 (\max_{t\in[T]} \sigma_t^2 + (R+1)^2/d) d \log \frac{(R+1)^2\lambda + TA^2}{(R+1)^2\lambda},\end{aligned} \tag{A.19}$$

where the first inequality follows from the definition of $\mathcal{I}_2$, the second inequality follows from Lemma C.2.

Substituting (A.18) and (A.19) into (A.17), we have

$$I_1 = \widetilde{O}\left(C^2R^2d + d^2C^2 \max_{t\in[T]} \sigma_t^2\right). \tag{A.20}$$

To bound $I_2(\ell)$, by Lemma A.7, we have

$$\begin{aligned} I_2(\ell) &\leq \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}(f(t)=\ell) \min\left\{4, 4\gamma_{t,\ell}^2 \|\mathbf{a}_t\|^2_{\boldsymbol{\Sigma}_{t,\ell}^{-1}}\right\}\right] \\ &\leq 4\underbrace{\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}(f(t)=\ell) \min\left\{1, \gamma_{t,\ell}^2 \|\mathbf{a}_t\|^2_{\boldsymbol{\Sigma}_{t,\ell}^{-1}}\right\}\right]}_{I_4}. \end{aligned} \tag{A.21}$$

We divide $[T]$ into two parts to calculate I_4. Let $\mathcal{J}_1 = \{t \in [T] | \|\mathbf{a}_t/\overline{\sigma}_t\|_{\boldsymbol{\Sigma}_{t,\ell}^{-1}} > 1\}, \mathcal{J}_2 = \{t \in [T] | \|\mathbf{a}_t/\overline{\sigma}_t\|_{\boldsymbol{\Sigma}_{t,\ell}^{-1}} \leq 1\}$. For $\mathcal{J}_1$, we have

$$\begin{aligned}\sum_{t\in\mathcal{J}_1} \mathbb{1}(f(t)=\ell) \min\left\{1, \gamma_{t,\ell}^2 \|\mathbf{a}_t\|^2_{\boldsymbol{\Sigma}_{t,\ell}^{-1}}\right\} &\leq |\mathcal{J}_1| \\ &\leq \sum_{t\in[T]} \min\left\{1, \|\mathbf{a}_t\|^2_{\boldsymbol{\Sigma}_{t,\ell}^{-1}}\right\} \\ &\leq 2d \log \frac{(R+1)^2\lambda + TA^2}{(R+1)^2\lambda},\end{aligned} \tag{A.22}$$

where the second inequality follows from the fact that $\mathcal{J}_1 \subseteq [T]$, the third inequality holds due to Lemma C.2. For $\mathcal{J}_2$, we have

$$\begin{aligned}\sum_{t\in\mathcal{J}_2} \mathbb{1}(f(t)=\ell) \min\left\{1, \gamma_{t,\ell}^2 \|\mathbf{a}_t\|^2_{\boldsymbol{\Sigma}_{t,\ell}^{-1}}\right\} &\leq \gamma_{T,\ell}^2 \max_{t\in[T]} \overline{\sigma}_t^2 \sum_{t\in\mathcal{J}_2} \|\mathbf{a}_t\|^2_{\boldsymbol{\Sigma}_{t,\ell}^{-1}} \\ &\leq \gamma_{T,\ell}^2 (\max_{t\in[T]} \sigma_t^2 + (R+1)^2/d) \sum_{t\in[T]} \min\left\{1, \|\mathbf{a}\|^2_{\boldsymbol{\Sigma}_{t,\ell}^{-1}}\right\} \\ &\leq \gamma_{T,\ell}^2 (\max_{t\in[T]} \sigma_t^2 + (R+1)^2/d) 2d \log \frac{(R+1)^2\lambda + TA^2}{(R+1)^2\lambda},\end{aligned} \tag{A.23}$$

where the second inequality follows from the definition of $\overline{\sigma}_t$ and $\mathcal{J}_2$ and the third inequality holds due to Lemma C.2.

Substituting (A.22) and (A.23) into (A.21), we have

$$I_2(\ell) = \widetilde{O}(dR^2 + d^2 \max_{t\in[T]} \sigma_t^2). \tag{A.24}$$

Finally, substituting (A.24) and (A.20) into (A.16), we have

$$\mathbf{Regret}(T) = \frac{1}{\Delta}\widetilde{O}\left(C^2R^2d + d^2C^2\max_{t\in[T]}\sigma_t^2\right).$$

□

B Proof of Technical Lemmas in Section A

B.1 Proof of Lemma A.1

Proof. Let $\mathcal{S}(t) = \{1 \le i \le t | y_i \neq y_i^{\text{stoch}}\}$, $\mathbf{b}_t^{\text{stoch}} = \sum_{i=1}^t y_i^{\text{stoch}}\mathbf{x}_i$ and $\boldsymbol{\mu}_t^{\text{stoch}} = \mathbf{Z}_t^{-1}\mathbf{b}_t^{\text{stoch}}$. By Lemma C.1, we have that with probability at least $1-\delta$, $\|\boldsymbol{\mu}_t^{\text{stoch}} - \boldsymbol{\mu}^*\|_{\mathbf{Z}_t} \le \beta_t + \sqrt{\lambda}\|\boldsymbol{\mu}^*\|_2$ holds for all $t \ge 1$.

Also, we have

$$\begin{aligned}
\|\boldsymbol{\mu}_t - \boldsymbol{\mu}_t^{\text{stoch}}\|_{\mathbf{Z}_t} &= \|\mathbf{Z}_t^{-1}(\mathbf{b}_t - \mathbf{b}_t^{\text{stoch}})\|_{\mathbf{Z}_t} \\
&\le \sum_{i=1}^t \|\mathbf{Z}_t^{-1}(y_i^{\text{stoch}} - y_i)\mathbf{x}_i\|_{\mathbf{Z}_t} \\
&\le \sum_{i=1}^t |y_i^{\text{stoch}} - y_i| \cdot \|\mathbf{x}_i\|_{\mathbf{Z}_t^{-1}} \\
&\le C(t).
\end{aligned}$$

where the first inequality holds due to the triangle inequality and the last inequality holds due to $\|\mathbf{x}_i\|_{\mathbf{Z}_t^{-1}} \le 1$.

Hence, we can obtain

$$\|\boldsymbol{\mu}_t - \boldsymbol{\mu}^*\|_{\mathbf{Z}_t} \le \|\boldsymbol{\mu}_t^{\text{stoch}} - \boldsymbol{\mu}^*\|_{\mathbf{Z}_t} + \|\boldsymbol{\mu}_t - \boldsymbol{\mu}_t^{\text{stoch}}\|_{\mathbf{Z}_t} \le \beta_t + C(t) + \sqrt{\lambda}\|\boldsymbol{\mu}^*\|_2.$$

□

B.2 Proof of Lemma A.2

Lemma B.1 (Lemma 3.3, Lykouris et al. 2018)**.** Define the corruption level for a level ℓ:

$$\text{Corruption}_{t,\ell} = \sum_{i=1}^t \frac{\mathbb{1}(f(i)=\ell)}{R+1} \cdot \sup_{\mathbf{a}\in\mathcal{D}_i} |r_i(\mathbf{a}) - r_i'(\mathbf{a})|.$$

Then we have for all $\ell \ge \ell^*$, with probability at least $1-\delta$:

$$\text{Corruption}_{t,\ell} \le \log(1/\delta) + 3, \qquad \forall t \ge 1.$$

Proof of Lemma A.2. Applying Lemma B.1, we have for all $\ell \ge \ell^*$, with probability at least $1-\delta/(2\ell^2)$: $\text{Corruption}_{t,\ell} \le \log(2\ell^2/\delta) + 3, \forall t \ge 1$.

Using a union bound over all $\ell \ge \ell^*$, we can prove the lemma. □

B.3 Proof of Lemma A.4

To prove the lemma, we first define the following two events:

$$\mathcal{E}_2 := \{\forall \ell \ge \ell^* \text{ and } t \ge 1, \|\boldsymbol{\mu}^* - \boldsymbol{\mu}_t\|_{\boldsymbol{\Sigma}_t} \le \beta_{t,\ell}\} \tag{B.1}$$

$$\mathcal{E}_3 := \left\{\forall \ell \ge \ell^* \text{ and } t \ge 1, \|\boldsymbol{\mu}^* - \boldsymbol{\mu}_{t,\ell}\|_{\boldsymbol{\Sigma}_{t,\ell}} \le \gamma_{t,\ell}\right\} \tag{B.2}$$

Lemma B.2. Let $\mathcal{E}_2$ be defined in (B.1). For any $0 < \delta < 1$, we have $\mathbb{P}(\mathcal{E}_2) \ge 1-\delta$.

Proof. Applying Lemma A.1, we have that $\|\boldsymbol{\mu}_t - \boldsymbol{\mu}^*\|_{\boldsymbol{\Sigma}_t} \le 8\sqrt{d\log\frac{(R+1)^2\lambda+tA^2}{(R+1)^2\lambda}\log(4t^2/\delta)} + 4\sqrt{d}\log(4t^2/\delta) + C\sqrt{d} + \sqrt{\lambda}\|\boldsymbol{\mu}^*\|_2$ for all $t \ge 1$ with probability at least $1-\delta$. Note that $2^\ell \ge C$ for all $\ell \ge \ell^*$, which indicates that $\mathcal{E}_2$ occurs with probability at least $1-\delta$. □

Lemma B.3. Let $\mathcal{E}_3$ be defined in (B.2). For any $0 < \delta < 1$, we have $\mathbb{P}(\mathcal{E}_3) \ge 1 - 2\delta$.

Proof. Applying Lemma A.1, we have that $\|\boldsymbol{\mu}_{t,\ell} - \boldsymbol{\mu}^*\|_{\boldsymbol{\Sigma}_t} \le 8\sqrt{d\log\frac{(R+1)^2\lambda+tA^2}{(R+1)^2\lambda}\log(4t^2T/\delta)} + 4\sqrt{d}\log(4t^2T/\delta) + \text{Corruption}_{t,\ell}\sqrt{d} + \sqrt{\lambda}\|\boldsymbol{\mu}^*\|_2$ for all $t \ge 1$ with probability at least $1-\delta/\ell$. Here we use the fact that $\ell \le T$. Applying Lemma A.2 and a union bound, we have $\mathcal{E}_3$ occurs with probability at least $1-2\delta$. □

Proof of Lemma A.4. This lemma can be proved by a union bound on $\mathcal{E}_2$ and $\mathcal{E}_3$ with Lemmas B.2 and B.3. □

B.4 Proof of Lemma A.6

Proof. For simplicity, let $\mathcal{A}_{t,\ell} = \{\boldsymbol{\mu} | \|\boldsymbol{\mu} - \boldsymbol{\mu}_t\|_{\boldsymbol{\Sigma}_t} \le \beta_{t,\ell}\}, \mathcal{B}_{t,\ell} = \left\{\boldsymbol{\mu} | \|\boldsymbol{\mu} - \boldsymbol{\mu}_{t,\ell}\|_{\boldsymbol{\Sigma}_{t,\ell}} \le \gamma_{t,\ell}\right\}$. Let $\boldsymbol{\mu}_t^m = \operatorname{argmax}_{\boldsymbol{\mu}\in\mathcal{C}_{t,f(t)}}\langle \mathbf{a}_t, \boldsymbol{\mu}\rangle$. Then we have

$$
\begin{aligned}
\langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle &\ge \langle \mathbf{a}_t, \boldsymbol{\mu}_t\rangle - \beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}} \\
&\ge \langle \mathbf{a}_t, \boldsymbol{\mu}_t^m\rangle - 2\beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}} \\
&\ge \langle \mathbf{a}_{t,\ell^*}, \boldsymbol{\mu}_t^m\rangle - 2\beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}} \\
&\ge \langle \mathbf{a}_{t,\ell^*}, \boldsymbol{\mu}_t\rangle - \beta_{t,\ell^*}\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}} - 2\beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}} \\
&\ge \max_{\boldsymbol{\mu}\in\mathcal{A}_{t,\ell^*}}\langle \mathbf{a}_{t,\ell^*}, \boldsymbol{\mu}\rangle - 2\beta_{t,\ell^*}\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}} - 2\beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}}, \\
&\ge \max_{\boldsymbol{\mu}\in\mathcal{C}_{t,\ell^*}}\langle \mathbf{a}_{t,\ell^*}, \boldsymbol{\mu}\rangle - 2\beta_{t,\ell^*}\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}} - 2\beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}},
\end{aligned} \tag{B.3}
$$

where the first inequality holds since $\boldsymbol{\mu}^* \in \mathcal{C}_{t,\ell^*} \subseteq \mathcal{A}_{t,l^*}$, the second inequality holds since $\boldsymbol{\mu}_t^m \in \mathcal{C}_{t,f(t)} \subseteq \mathcal{A}_{t,l^*}$, the third inequality holds by the definition of $\mathbf{a}_t$ and $\boldsymbol{\mu}_t^m$, the fourth inequality holds since $\boldsymbol{\mu}_t^m \in \mathcal{A}_{t,\ell^*}$, the fifth inequality holds since $\boldsymbol{\mu}_t \in \mathcal{A}_{t,\ell^*}$, the last one holds since $\mathcal{C}_{t,\ell^*} \subseteq \mathcal{A}_{t,\ell^*}$. By the definition of $\mathcal{E}_1$ and $\mathbf{a}_{t,\ell^*}$, we have

$$
\max_{\boldsymbol{\mu}\in\mathcal{C}_{t,\ell^*}}\langle \mathbf{a}_{t,\ell^*}, \boldsymbol{\mu}\rangle = \max_{\mathbf{a}\in\mathcal{D}_t}\max_{\boldsymbol{\mu}\in\mathcal{C}_{t,\ell^*}}\langle \mathbf{a}, \boldsymbol{\mu}\rangle \ge \max_{\mathbf{a}\in\mathcal{D}_t}\langle \mathbf{a}, \boldsymbol{\mu}^*\rangle = \langle \mathbf{a}_t^*, \boldsymbol{\mu}^*\rangle. \tag{B.4}
$$

Combining (B.3) with (B.4), we have $\langle \mathbf{a}_t^* - \mathbf{a}_t, \boldsymbol{\mu}^*\rangle \le 2\beta_{t,\ell^*}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_t^{-1}} + 2\beta_{t,\ell^*}\|\mathbf{a}_{t,\ell^*}\|_{\boldsymbol{\Sigma}_t^{-1}}$. □

B.5 Proof of Lemma A.7

Proof. We have

$$
\langle \mathbf{a}_t^* - \mathbf{a}_t, \boldsymbol{\mu}^*\rangle \le \max_{\boldsymbol{\mu}\in\mathcal{C}_{t,\ell}}\langle \mathbf{a}_t, \boldsymbol{\mu}\rangle - \langle \mathbf{a}_t, \boldsymbol{\mu}^*\rangle \le 2\gamma_{t,\ell}\|\mathbf{a}_t\|_{\boldsymbol{\Sigma}_{t,\ell}^{-1}},
$$

where the first inequality follows from the fact that $\boldsymbol{\mu}^* \in \mathcal{C}_{t,\ell}$ and the definition of $\mathbf{a}_t$, the second inequality holds since $\boldsymbol{\mu}^* \in \mathcal{C}_{t,\ell}$ on the event $\mathcal{E}_1$. □

C Auxiliary Lemmas

Lemma C.1 (Theorem 4.1, Zhou et al. 2020)**.** Let $\{\mathcal{G}_t\}_{t=1}^\infty$ be a filtration, $\{\mathbf{x}_t, \eta_t\}_{t\ge1}$ a stochastic process so that $\mathbf{x}_t \in \mathbb{R}^d$ is $\mathcal{G}_t$-measurable and $\eta_t \in \mathbb{R}$ is $\mathcal{G}_{t+1}$-measurable. Fix $R, L, \sigma, \lambda > 0$, $\boldsymbol{\mu}^* \in \mathbb{R}^d$. For $t \ge 1$ let $y_t = \langle \boldsymbol{\mu}^*, \mathbf{x}_t\rangle + \eta_t$ and suppose that $\eta_t, \mathbf{x}_t$ also satisfy

$$
|\eta_t| \le R, \mathbb{E}[\eta_t|\mathcal{G}_t] = 0, \mathbb{E}[\eta_t^2|\mathcal{G}_t] \le \sigma^2, \|\mathbf{x}_t\|_2 \le L.
$$

Then, for any $0 < \delta < 1$, with probability at least $1-\delta$ we have $\forall t > 0$,

$$
\|\boldsymbol{\mu}_t - \boldsymbol{\mu}^*\|_{\mathbf{Z}_t} \le \beta_t + \sqrt{\lambda}\|\boldsymbol{\mu}^*\|_2,
$$

where for $t \geq 1$, $\boldsymbol{\mu}_t = \mathbf{Z}_t^{-1}\mathbf{b}_t$, $\mathbf{Z}_t = \lambda\mathbf{I} + \sum_{i=1}^t \mathbf{x}_i\mathbf{x}_i^\top$, $\mathbf{b}_t = \sum_{i=1}^t y_i\mathbf{x}_i$, and

$$\beta_t = 8\sigma\sqrt{d\log\frac{d\lambda + tL^2}{d\lambda}\log(4t^2/\delta)} + 4R\log(4t^2/\delta).$$

Lemma C.2 (Lemma 11, Abbasi-Yadkori et al. 2011)**.** For any $\lambda > 0$ and sequence $\{\mathbf{x}_t\}_{t=1}^T \subset \mathbb{R}^d$ for $t \in 0 \cup [T]$, define $\mathbf{Z}_t = \lambda\mathbf{I} + \sum_{i=1}^t \mathbf{x}_i\mathbf{x}_i^\top$. Then, provided that $\|\mathbf{x}_t\|_2 \leq L$ holds for all $t \in [T]$, we have

$$\sum_{t=1}^T \min\{1, \|\mathbf{x}_t\|^2_{\mathbf{Z}_{t-1}^{-1}}\} \leq 2d\log\frac{d\lambda + TL^2}{d\lambda}.$$

D Experiments

In this section, we conduct experiments and evaluate the performance our algorithm Multi-level OFUL, along with the baselines, OFUL (Abbasi-Yadkori et al., 2011), weighted OFUL (Zhou et al., 2020) and the greedy algorithm proposed by Bogunovic et al. (2021) under different corruption levels. We repeat each baseline algorithm for 10 times and plot their regrets w.r.t. number of rounds in Figure 1.

D.1 Experimental Setup

Following Bogunovic et al. (2021), we let the adversary always corrupt the first k rounds, and leave the rest $T - k$ rounds intact. According to our definition in (3.2), our design can simulate the cases where corruption level is $2k$.

Model parameters. Recall that corrupted linear contextual bandits defined in Section 3, we consider $B = 1$, $A = 1$ $d = 20$ and $R = 0.5$ and fix $\boldsymbol{\mu}^*$ as $\left(\frac{1}{\sqrt{d}}, \cdots, \frac{1}{\sqrt{d}}\right)^\top$. We set σ_t as a random variable which is independently and uniformly chosen from $[0, 0.05]$ in each round t. Note that $\langle \mathbf{a}, \boldsymbol{\mu}^* \rangle \in [-1, 1]$ always hold for any eligible $\mathbf{a}$ under our setting of parameters.

Attack method. In the first k rounds, the adversary always trick the learner by flipping the value of $\boldsymbol{\mu}^*$, i.e., $r_t(\mathbf{a}) = -\langle \mathbf{a}, \boldsymbol{\mu}^* \rangle + \epsilon_t(\mathbf{a})$ for all $t \in [k]$ and $\mathbf{a} \in \mathcal{D}_t$.

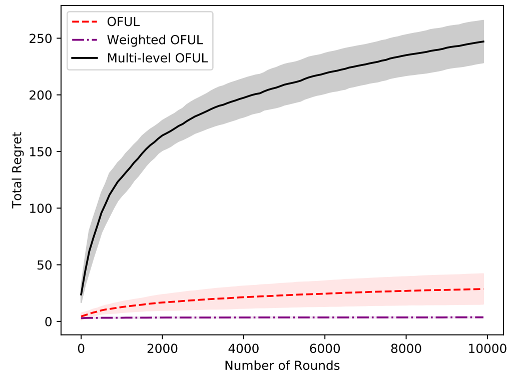
Decision set. We consider $|\mathcal{D}_t| = 20$ for all $t \geq 1$. In each of the first k rounds, we generate the 20 actions in $\mathcal{D}_t$ independently, each having entries drawn i.i.d. from the uniform distribution on $\left[-\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}\right]$. For the following uncorrupted rounds, however, we use a fixed $\mathcal{D}$ generated in the same way.

Intuitively, non-robust algorithm will "learn" the flipped $\boldsymbol{\mu}^*$ faster with diversified action vectors. As a result, the learner is likely to select the same nonoptimal action for a huge number of rounds afterwards, making it even more difficult to learn the true $\boldsymbol{\mu}^*$.

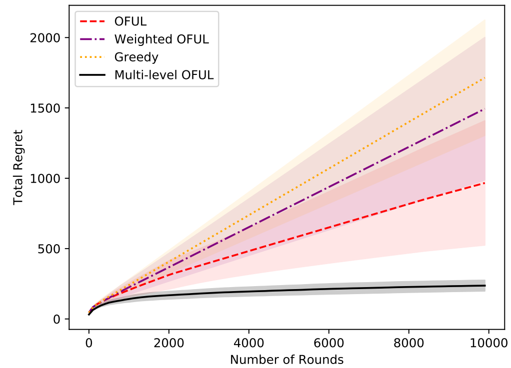
Noise synthesis. We generate identical noises ϵ_t for all $\mathbf{a} \in \mathcal{D}_t$ at each round t, i.e., $\epsilon_t(\mathbf{a}) = \epsilon_t$. To generate ϵ_t, we first generate ϵ_t' subject to $\mathcal{N}(0, \sigma_t^2)$ and let

$$\epsilon_t = \begin{cases} -R, & \epsilon_t' < -R \\ R, & \epsilon_t' > R \\ \epsilon_t', & \text{otherwise} \end{cases}.$$

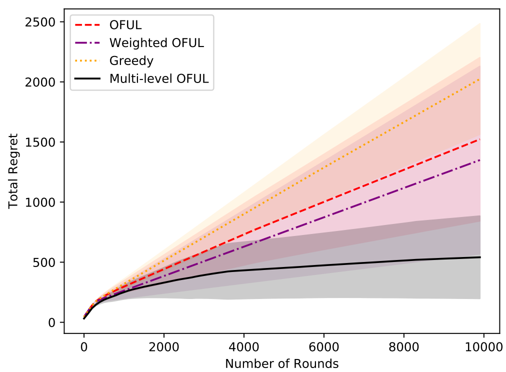
D.2 Results and Discussion



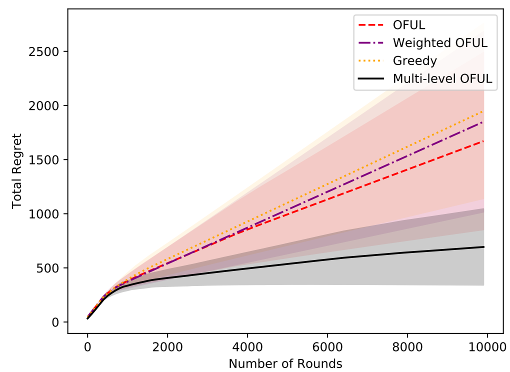
(a) $C = 0$



(b) $C = 300$



(c) $C = 600$



(d) $C = 900$

Figure 1: Regret plot against number of rounds under corruption level from 0 to 900 averaged in 10 trials.

We plot the regret with respect to the number of rounds in Figure 1. The results are averaged over 10 trials. In the setting where $C = 0$ (Figure 1(a)), we only plot the regret of OFUL, weighted OFUL and Multi-level OFUL, and do not plot the regret of the greedy algorithm since its regret is much worse than the other three algorithms.

We have the following observations from Figure 1. For the corruption-free case $C = 0$ (Figure 1(a)), our proposed Multi-level OFUL behaves worse than weighted OFUL and OFUL, which is not surprising since Multi-level OFUL has additional algorithm design to deal with the corruption and it may pay additional price in regret in the absence of corruption. Weighted OFUL outperforms OFUL remarkably since it takes advantage of the information concerning the variance of noise. For the corruption case (Figure 1(b) to 1(d)), our Multi-level OFUL outperforms other baseline algorithms by a large margin, which suggests that it can deal with the corruption successfully.