Robust and Computationally Efficient Linear Contextual Bandits under Adversarial Corruption and Heavy-Tailed Noise

Naoto Tani[1,2] **Futoshi Futami**[1,2,3]

[1]Graduate School of Engineering Science, The University of Osaka, Osaka, Japan
[2]RIKEN AIP, Tokyo, Japan
[3]Department of Complexity Science and Engineering, The University of Tokyo, Tokyo, Japan

Abstract

We study linear contextual bandits under adversarial corruption and heavy-tailed noise with finite $(1+\epsilon)$-th moments for some $\epsilon \in (0, 1]$. Existing work that addresses both adversarial corruption and heavy-tailed noise relies on a finite variance assumption and suffers from computational inefficiency. We propose the first computationally efficient algorithm based on online mirror descent that is robust to both adversarial corruption and heavy-tailed noise. While the existing algorithm incurs $\mathcal{O}(t \log T)$ computational cost per round, our algorithm reduces this to $\mathcal{O}(1)$ per round. We establish an additive regret bound consisting of a term depending on the $(1+\epsilon)$-moment bound of the noise and a term depending on the total amount of corruption. This bound unifies and extends previous guarantees for (generalized) linear contextual bandits under adversarial corruption and heavy-tailed noise. In particular, when $\epsilon = 1$, it recovers existing guarantees under finite-variance assumptions. When no corruption is present, it achieves the same regret rate as existing results for linear contextual bandits with heavy-tailed noise in the general setting. Moreover, the algorithm requires only upper bounds on the noise moment and the total amount of corruption, while still guaranteeing sublinear regret.

1 INTRODUCTION

Linear contextual bandits provide a general framework for sequential decision-making with side information [Lattimore and Szepesvári, 2020]. Specifically, over a time horizon T, in each round t, the learner selects an action associated with a d-dimensional context vector X_t and observes a stochastic reward $r_t = \langle \theta^*, X_t \rangle + \eta_t$, where θ^* is an d-dimensional unknown parameter and η_t is zero-mean noise. The learner's goal is to select high-reward actions to maximize the cumulative reward over T. Most existing analyses assume that the noise is conditionally sub-Gaussian [Abbasi-Yadkori et al., 2011, Agrawal and Goyal, 2013], which enables the use of self-normalized concentration inequalities.

However, in practical applications, this reward model can be violated due to two major challenges. **First, the stochastic reward may be adversarially corrupted.** Adversarial corruption arises in many real-world systems, including the presence of spam or malicious reviews in recommendation systems [Lykouris et al., 2018]. It can strategically bias parameter estimation, preventing the learner from maximizing cumulative reward. **Second, the noise can be heavy-tailed.** Heavy-tailed noise is commonly observed in domains such as finance [Rachev, 2003]. Unlike sub-Gaussian noise, heavy-tailed noise may yield extreme rewards, preventing standard concentration arguments for least-squares estimators from applying directly.

Since adversarial corruption and heavy-tailed noise can arise simultaneously in practice, algorithms must be robust to both challenges at the same time. In online advertising systems [Li et al., 2010], click fraud introduces adversarial corruption [Lykouris et al., 2018], while value return can also exhibit heavy-tailed behavior [Jebarajakirthy et al., 2021]. In such a situation, existing methods that address only one of these challenges [He et al., 2022, Wang et al., 2025] cannot guarantee robustness to the other challenge.

Although Yu et al. [2025] are the first to study contextual bandits under simultaneous adversarial corruption and heavy-tailed noise, they assume that the noise has a bounded second moment (i.e., finite variance). This requirement is stronger than the bounded $(1+\epsilon)$-moment assumption for some $\epsilon \in (0, 1]$ adopted in linear contextual bandits with general heavy-tailed noise [Huang et al., 2023, Wang et al., 2025]. Therefore, their results do not cover the more general bounded $(1+\epsilon)$-moment setting.

In addition to this statistical limitation, the algorithm of

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:6645–6676.

Yu et al. [2025] is computationally inefficient. At each round t, their estimator is obtained by solving a convex optimization problem over all past observations, resulting in an $\mathcal{O}(t \log T)$ per-round computational cost. Such a per-round computational cost that grows with t is undesirable in large-scale or real-time applications. This contrasts with recent algorithms that achieve $\mathcal{O}(1)$ per-round computational cost when handling either adversarial corruption or heavy-tailed noise [He et al., 2022, Wang et al., 2025].

Taken together, existing works do not provide an algorithm that is simultaneously robust to adversarial corruption and heavy-tailed noise under bounded $(1+\epsilon)$-moment assumptions while also achieving per-round $\mathcal{O}(1)$ computational cost. Robustness, in this context, means achieving sublinear regret despite the presence of adversarial corruption or heavy-tailed noise. In this paper, we study linear contextual bandits under simultaneous adversarial corruption and heavy-tailed noise satisfying a bounded $(1+\epsilon)$-moment assumption. Under this setting, we address both robustness and computational efficiency. We summarize our main contributions below.

- We propose a new algorithm, corruption-robust heavy-tailed UCB (CR-Hvt-UCB, Algorithm 1). It achieves robustness to both adversarial corruption and heavy-tailed noise with bounded $(1+\epsilon)$-th moments for any $\epsilon \in (0, 1]$, while admitting $\mathcal{O}(1)$ per-round updates. This generalizes the finite-variance setting of existing work while improving its computational efficiency. The key idea is to perform an online mirror descent (OMD) update (11) on an adaptively scaled Huber-based loss (5). The adaptive scaling controls the influence of both adversarial corruption and heavy-tailed noise, enabling the algorithm to achieve both robustness and computational efficiency.
- We establish regret guarantees for Algorithm 1 when the noise moment bounds and the total amount of corruption are known (Theorem 4.5). The regret-bound scales with the square root of the cumulative squared $(1+\epsilon)$-moment bounds of the noise, plus an additive term linear in the total amount of corruption. In particular, when $\epsilon = 1$, our bound recovers the regret guarantee of Yu et al. [2025] in the linear case with finite-variance noise. In the absence of adversarial corruption, the bound matches the best-known rate for linear contextual bandits with heavy-tailed noise. Importantly, our algorithm works even when the total amount of adversarial corruption and the $(1+\epsilon)$-moment bounds of the noise are unknown, while retaining theoretical regret guarantees (Corollary 4.7, 4.8, 4.10).

The remainder of the paper is organized as follows. Section 2 discusses related work and Section 3 formalizes the problem setting. Section 4 presents our algorithm together with its regret guarantees. Section 5 offers a proof sketch of the main result. Section 6 reports experimental results, and Section 7 concludes the paper.

2 RELATED WORK

We first position our work relative to prior studies along two key aspects: adversarial corruption and heavy-tailed noise. In our setting, stochastic rewards are corrupted by an additive adversarial corruption with total magnitude constrained by a budget C, and the heavy-tailed noise η_t satisfies a bounded $(1+\epsilon)$-moment condition, i.e., $\mathbb{E}\left[|\eta_t|^{1+\epsilon}\right] \leq \nu_t^{1+\epsilon}$ for some $\epsilon \in (0, 1]$. A detailed and formal problem setting is deferred to Section 3.1. Comparisons with additional existing works are provided in Appendix A.

He et al. [2022] study linear contextual bandits under adversarial corruption and adopt the same additive corruption model as ours. Assuming R-sub-Gaussian noise, they establish a nearly optimal regret bound of $\tilde{\mathcal{O}}(d\sqrt{T} + dC)$. Their analysis relies on self-normalized concentration inequalities for sub-Gaussian noise. Since bounded $(1+\epsilon)$-moment noise need not be sub-Gaussian, their framework does not directly extend to our setting. In contrast, we develop a robust algorithm and analysis that remain valid under bounded $(1+\epsilon)$-moment noise. In the finite-variance case $\epsilon = 1$ with a constant moment bound, our regret bound recovers the same order $\tilde{\mathcal{O}}(d\sqrt{T} + dC)$ (see Table 1).

Wang et al. [2025] study linear contextual bandits with heavy-tailed noise under the same bounded $(1+\epsilon)$-moment assumption as ours. Under this assumption, they establish instance-dependent regret bounds of $\tilde{\mathcal{O}}\big(dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{\sum_{t=1}^{T}\nu_t^2} + dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\big)$. However, their algorithm and analysis cannot be directly applied to our setting. In our setting, the quantities controlled by their concentration inequalities include an additional adversarial corruption, which invalidates the stochastic structure required by their concentration arguments. Therefore, we develop a new algorithm and a corresponding analysis that are robust to both adversarial corruption and heavy-tailed noise. Moreover, when the corruption budget satisfies $C = 0$, our regret bound recovers their rate (see Table 1).

Yu et al. [2025] study generalized linear bandits (GLBs) under simultaneous adversarial corruption and heavy-tailed noise, and adopt the same additive corruption model as ours. Under the finite-variance assumption $\mathbb{E}[\eta_t^2] \leq \nu_t^2$, they establish regret bounds of $\tilde{\mathcal{O}}\big(d\sqrt{\sum_{t=1}^{T}\nu_t^2} + d \cdot 1 \vee C\big)$. Their guarantees are restricted to the finite-variance setting. In contrast, we relax the finite-variance assumption to a bounded $(1+\epsilon)$-moment assumption for some $\epsilon \in (0, 1]$ and establish regret guarantees under this weaker assumption. In the finite-variance case $\epsilon = 1$, our regret bound recovers the same order as theirs (see Table 1).

More closely related to our setting, Zhao et al. [2025] pro-

Paper	C-Robust	HT-Robust	Efficiency	Regret
Abbasi-Yadkori et al. [2011]	No	No	$\mathcal{O}(1)$	$\tilde{\mathcal{O}}(d\sqrt{T})$
Zhang et al. [2026]	No	No	$\mathcal{O}(1)$	$\tilde{\mathcal{O}}(d\sqrt{T})$
He et al. [2022]	Yes	No	$\mathcal{O}(1)$	$\tilde{\mathcal{O}}(d\sqrt{T}+dC)$
Wang et al. [2025]	No	Yes	$\mathcal{O}(1)$	$\tilde{\mathcal{O}}\left(dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{\sum_{t=1}^{T}\nu_t^2}+dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\right)$ $\tilde{\mathcal{O}}(dT^{\frac{1}{1+\epsilon}})$
Yu et al. [2025]	Yes	$\epsilon=1$ only	$\mathcal{O}(t\log T)$	$\tilde{\mathcal{O}}\left(d\sqrt{\sum_{t=1}^{T}\nu_t^2}+d\cdot 1\vee C\right)$ $\tilde{\mathcal{O}}(d\sqrt{T}+d\cdot 1\vee C)$
Our work	Yes	Yes	$\mathcal{O}(1)$	$\tilde{\mathcal{O}}\left(dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{\sum_{t=1}^{T}\nu_t^2}+dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\cdot 1\vee C\right)$ $\tilde{\mathcal{O}}(dT^{\frac{1}{1+\epsilon}}+dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\cdot 1\vee C)$

Table 1: We compare our algorithm with existing methods in terms of three key aspects—robustness to adversarial corruption (C-Robust), robustness to heavy-tailed noise (HT-Robust), and per-round computational cost (Efficiency)—and summarize their regret bounds. The second row in the Regret column gives the simplified bound for uniformly bounded moments, $\nu_t \le \nu = \mathcal{O}(1)$ for all t. We define $1 \vee C = \max\{1, C\}$.

vide the first study of heavy-tailed linear bandits with adversarial corruption from a best-of-both-worlds (BOBW) perspective. However, their setting differs from ours in several important respects. First, they consider linear bandits with a fixed finite decision set, where their algorithm employs Follow-the-Regularized-Leader (FTRL) over distributions on the fixed arms. In contrast, we study general contextual decision sets, and it is unclear whether such a finite-action distributional update can be directly applied to our setting. Second, their heavy-tailed assumption is imposed on the observed losses, while we assume bounded $(1+\epsilon)$-moment noise. Moreover, they assume a uniform upper bound on the $(1+\epsilon)$-th moments over the time horizon T, whereas our framework allows time-dependent moment bounds. We do not include a direct comparison with their regret bound, since their guarantee is gap-dependent, whereas our result provides a worst-case regret bound.

We next discuss prior work from the perspective of per-round computational efficiency. Further comparisons are deferred to Appendix A.

In linear bandits, the OFUL algorithm [Abbasi-Yadkori et al., 2011] achieves $\mathcal{O}(1)$ per-round computational cost via closed-form updates based on ridge regression. In linear contextual bandits with adversarial corruption, He et al. [2022] also achieve $\mathcal{O}(1)$ cost through similar closed-form updates. By contrast, Yu et al. [2025] consider GLBs with both adversarial corruption and heavy-tailed noise, where the estimator is defined as the solution to a regularized empirical risk minimization problem over all past observations. Consequently, the estimator is recomputed by solving a convex optimization problem at every round, leading to an $\mathcal{O}(t\log T)$ per-round computational cost.

More broadly, even when closed-form estimators are not available, efficient per-round updates can still be achieved through OMD. Zhang and Sugiyama [2024] first demonstrated $\mathcal{O}(1)$ per-round OMD updates in logistic bandits, and Zhang et al. [2026] extended this framework to GLBs. Building on this development, Wang et al. [2025] applied an OMD-based estimator to linear contextual bandits with heavy-tailed noise, showing that efficient updates remain possible even under heavy-tailed noise. More recently, Kim et al. [2026] improved the computational efficiency of Yu et al. [2025] via OMD updates. Their work considers a more restrictive heteroskedastic exponential-family GLB model than the finite-variance noise model of Yu et al. [2025], which enables sharper instance-wise regret bounds. In contrast, our work develops OMD-based efficient updates under bounded $(1+\epsilon)$-moment noise with adversarial corruption.

3 PROBLEM SETTING AND CHALLENGES

In this section, we introduce the problem setting (Section 3.1) and review the limitations of existing work (Section 3.2).

3.1 PROBLEM SETTING

Notation. Let I_d be the $d \times d$ identity matrix. For $n \in \mathbb{N}$, we define $[n] := \{1, 2, \ldots, n\}$. For $a, b \in \mathbb{R}$, we write $a \vee b := \max\{a, b\}$. For a scalar $x \in \mathbb{R}$ or a vector $\mathbf{x} \in \mathbb{R}^d$ and a positive semidefinite matrix A, we denote the vector's Euclidean norm as $\|\mathbf{x}\|_2$ and the Mahalanobis norm as $\|\mathbf{x}\|_A = \sqrt{\mathbf{x}^\top A\mathbf{x}}$. We denote by $\mathbb{1}\{\cdot\}$ the indicator function, that is, $\mathbb{1}_E = 1$ if E is true and 0 otherwise. For two positive functions f and g defined on the positive integers,

we write $f(T) = \mathcal{O}(g(T))$ if there exist constants $c > 0$ and $T_0 \in \mathbb{N}$ such that $f(T) \le cg(T)$ for all $T \ge T_0$. We also use $\tilde{\mathcal{O}}(\cdot)$ to hide polylogarithmic factors.

We define linear contextual bandits with adversarial corruption and heavy-tailed noise. At each round $t \in [T]$, the learner observes a decision set $\mathcal{X}_t \subset \mathbb{R}^d$ and selects an action $X_t \in \mathcal{X}_t$. After choosing the action X_t, the environment generates the stochastic reward based on the linear model

$$r'_t = \langle X_t, \theta^* \rangle + \eta_t, \tag{1}$$

where $\theta^* \in \mathbb{R}^d$ is an unknown parameter vector and η_t is a heavy-tailed noise term specified later. After generating the stochastic reward, an adversary may corrupt it by adding a corruption term c_t. This corruption may depend on the decision set $\mathcal{X}_t$, the action X_t, and the stochastic reward r'_t. Finally, the learner observes the corrupted reward

$$r_t = r'_t + c_t = \langle X_t, \theta^* \rangle + \eta_t + c_t. \tag{2}$$

To measure the total corruption, we define the corruption level $C = \sum_{t=1}^{T} |c_t|$, which may be known or unknown to the learner.

Following prior work [Wang et al., 2025, Yu et al., 2025], we make the following assumptions on this setting.

Assumption 3.1. *For each round $t \in T$, the heavy-tailed noise η_t satisfies, for some $\epsilon \in (0, 1]$, $\mathbb{E}[\eta_t \mid X_{1:t}, \eta_{1:t-1}, c_{1:t-1}] = 0$ and $\mathbb{E}[|\eta_t|^{1+\epsilon} \mid X_{1:t}, \eta_{1:t-1}, c_{1:t-1}] \le \nu_t^{1+\epsilon}$, where $\nu_t > 0$ is a known or unknown constant to the learner.*

Assumption 3.2. *For all t and all $\mathbf{x} \in \mathcal{X}_t$, we assume $\|\mathbf{x}\|_2 \le L$ for some $L > 0$, and $\theta^* \in \Theta := \{\theta : \|\theta\|_2 \le S\}$ for some $S > 0$.*

Under this model, the goal of the learner is to maximize the cumulative uncorrupted rewards. Accordingly, we define the regret with respect to the stochastic reward model (1) as

$$\mathrm{Reg}(T) = \sum_{t=1}^{T} \max_{\mathbf{x} \in \mathcal{X}_t} \langle \mathbf{x}, \theta^* \rangle - \sum_{t=1}^{T} \langle X_t, \theta^* \rangle.$$

3.2 LIMITATIONS OF EXISTING WORK

To understand the statistical and computational limitations of existing work, we examine the method of Yu et al. [2025]. They employ the Pseudo-Huber loss [Charbonnier et al., 1997, Li and Sun, 2023], which is a smooth approximation to the Huber loss [Huber, 1964].

Definition 3.3 (Pseudo-Huber loss)**.** *Let $\tau > 0$ be a robustification parameter. For $x \in \mathbb{R}$, the Pseudo-Huber loss $\phi_\tau(x)$ is defined as*

$$\phi_\tau(x) = \tau \left(\sqrt{\tau^2 + x^2} - \tau \right).$$

Based on this loss, their estimator is constructed at each round t by performing adaptive Huber regression [Sun et al., 2020] using all past observations.

$$\hat{\theta}_t = \arg\min_{\theta \in \Theta} \sum_{s=1}^{t} \phi_{\tau_s} \left(\frac{r_s - \langle X_s, \theta \rangle}{\sigma_s} \right) + \frac{\lambda}{2} \|\theta\|_2^2, \tag{3}$$

where $\lambda > 0$ is a regularization parameter and $\sigma_s > 0$ denotes a scale parameter. They propose an OFUL-type algorithm built upon this estimator.

Their robustness to heavy-tailed noise guarantees relies on a finite-variance assumption on the noise, namely $\mathbb{E}[\eta_t^2 \mid X_{1:t}, \eta_{1:t-1}, c_{1:t-1}] \le \nu_t^2$. This is because their analysis builds upon that of Li and Sun [2023], which fundamentally assumes bounded second moments. In particular, the scale parameter σ_t in their estimator (3) explicitly requires an upper bound on the variance. As a result, it does not naturally extend to our setting where only a bounded $(1+\epsilon)$-moment assumption holds.

Their algorithm is computationally expensive. This computational burden stems from its reliance on the adaptive pseudo-Huber regression of Li and Sun [2023], in which the estimator is defined as the solution to a regularized empirical risk minimization problem over all previously collected observations. Consequently, the estimator is recomputed by solving a convex optimization problem at every round. In particular, Wang et al. [2025] show that, for this class of optimization-based estimators, achieving the best-known regret bound incurs an $\mathcal{O}(t \log T)$ per-round computational cost.

4 METHOD

In this section, we present our algorithm (CR-Hvt-UCB, Algorithm 1). Section 4.1 describes its design, and Section 4.2 establishes the corresponding regret guarantee.

4.1 OUR ALGORITHM

Our algorithm achieves robustness to adversarial corruption and heavy-tailed noise with bounded $(1+\epsilon)$ moments while achieving $\mathcal{O}(1)$ per-round updates. It is built upon a Huber-based robust estimator combined with an update rule based on OMD.

Following prior Huber-based approaches for linear contextual bandits under bounded $(1+\epsilon)$-moment assumptions [Huang et al., 2023, Wang et al., 2025], we adopt the Huber loss [Huber, 1964].

Definition 4.1 (Huber loss)**.** *Let $\tau > 0$ be a threshold parameter. For $x \in \mathbb{R}$, the Huber loss is defined as*

$$f_\tau(x) := \begin{cases} \frac{x^2}{2}, & |x| \le \tau, \\ \tau|x| - \frac{\tau^2}{2}, & |x| > \tau. \end{cases} \tag{4}$$

The Huber loss coincides with the squared loss when $|x| \le \tau$ and transitions to the absolute loss when $|x| > \tau$. Consequently, its gradient is bounded by τ, which prevents extreme rewards from dominating the parameter estimation.

We employ the following Huber loss applied to the normalized prediction error $z_t(\theta) := (r_t - \theta^\top X_t)/\sigma_t$, where the scale parameter $\sigma_t > 0$ will be specified later in (6).

$$\ell_t(\theta) := \begin{cases} \frac{1}{2} z_t(\theta)^2, & |z_t(\theta)| \le \tau_t, \\ \tau_t |z_t(\theta)| - \frac{\tau_t^2}{2}, & |z_t(\theta)| > \tau_t. \end{cases} \tag{5}$$

The threshold parameter τ_t will be specified later in (7).

Existing Huber-based approaches [Huang et al., 2023, Wang et al., 2025] design the parameters σ_t and τ_t to ensure robustness to heavy-tailed noise. To achieve robustness to both adversarial corruption and heavy-tailed noise, we redesign σ_t and τ_t as follows:

$$\sigma_t := \max\Bigg\{ \nu_t, \sigma_{\min}, \sqrt{\frac{2\beta_{t-1}}{\tau_0 \sqrt{\alpha} t^{\frac{1-\epsilon}{2(1+\epsilon)}}}} \|X_t\|_{V_{t-1}^{-1}}, \sqrt{C} \kappa^{-1/4} \|X_t\|_{V_{t-1}^{-1}}^{1/2} \Bigg\}, \tag{6}$$

$$\tau_t := \tau_0 \frac{\sqrt{1 + w_t^2}}{w_t} t^{\frac{1-\epsilon}{2(1+\epsilon)}}. \tag{7}$$

The matrix V_t denotes the data-dependent positive definite matrix specified in (10). Here, $\sigma_{\min} > 0$ is a small constant specified in Lemma 4.4, and β_t denotes the confidence radius defined in the same lemma. The parameter $\alpha > 0$ will be specified following (10). We further define the problem-dependent constant $\kappa = d \log\Big(1 + \frac{L^2 T}{\sigma_{\min}^2 \lambda \alpha d}\Big)$. The auxiliary quantities τ_0 and w_t are given by

$$\tau_0 = \frac{\sqrt{2\kappa} (\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}}{\big(\log \frac{2T^2}{\delta}\big)^{\frac{1}{1+\epsilon}}}, w_t = \frac{1}{\sqrt{\alpha}} \left\| \frac{X_t}{\sigma_t} \right\|_{V_{t-1}^{-1}}. \tag{8}$$

Intuitively, the robustness to corruption stems from the adaptive normalization scale σ_t. When large corruption is present, the corruption-dependent term in σ_t inflates the normalization scale. Consequently, the normalized prediction error $z_t(\theta) = (r_t - \theta^\top X_t)/\sigma_t$ shrinks in magnitude, leading to a smaller gradient of the loss (5). This smaller gradient suppresses the influence of corruption on the parameter estimation. With this design, the loss (5) is robust to both adversarial corruption and heavy-tailed noise.

Remark 4.2. *Our definition of σ_t can be viewed as a direct generalization of the scheme of Wang et al. [2025]. When $C = 0$, the corruption-dependent component in* (6) *vanishes, and σ_t reduces exactly to the form used by Wang et al. [2025].*

Remark 4.3. *Since $\kappa = \tilde{\mathcal{O}}(d)$, the corruption-dependent component of σ_t in* (6) *scales as $\tilde{\mathcal{O}}(\sqrt{C} d^{-1/4} \|X_t\|_{V_{t-1}^{-1}}^{1/2})$. This matches the order of the corruption-dependent component in the scale parameter σ_t of Yu et al. [2025] defined in* (3)*. Moreover, the corruption-dependent weight in the weighted squared loss of He et al. [2022] can be expressed in terms of $z_t(\theta)$, under which it matches the order of the corruption-dependent component of σ_t in* (6)*.*

While the loss (5) ensures robustness to both adversarial corruption and heavy-tailed noise, empirical-loss minimization at each round over all past observations would incur the same computational cost as in Yu et al. [2025]. To overcome this issue, we build on the estimator proposed by Wang et al. [2025], which is based on the online mirror descent (OMD) update [Orabona, 2019].

Given the current estimate $\hat{\theta}_t \in \Theta$, the OMD update takes the form

$$\hat{\theta}_{t+1} = \arg\min_{\theta \in \Theta} \Big\langle \theta, \nabla \ell_t(\hat{\theta}_t) \Big\rangle + \mathcal{D}_{\psi_t}(\theta, \hat{\theta}_t), \tag{9}$$

where $\mathcal{D}_{\psi_t}(\theta, \theta') := \psi_t(\theta) - \psi_t(\theta') - \langle \nabla \psi_t(\theta'), \theta - \theta' \rangle$ denotes the Bregman divergence induced by the regularizer ψ_t. We choose a quadratic regularizer

$$\psi_t(\theta) = \frac{1}{2} \|\theta\|_{V_t}^2, \qquad V_t := V_{t-1} + \frac{1}{\alpha} \frac{X_t X_t^\top}{\sigma_t^2}, \tag{10}$$

with $V_0 = \lambda I_d$, where $\lambda > 0$ is a regularization parameter and $\alpha > 0$ is the step-size parameter of the OMD update. The positive definite matrix V_t modulates the contribution of each round via the weight $1/\sigma_t^2$, ensuring that observations more strongly influenced by adversarial corruption or heavy-tailed noise exert a smaller impact on the update. This choice of regularizer aligns the update geometry with the elliptical norm used for constructing confidence sets in Lemma 4.4. The OMD update (9) admits an equivalent two-step representation:

$$\begin{aligned} \tilde{\theta}_{t+1} &= \hat{\theta}_t - V_t^{-1} \nabla \ell_t(\hat{\theta}_t), \\ \hat{\theta}_{t+1} &= \arg\min_{\theta \in \Theta} \|\theta - \tilde{\theta}_{t+1}\|_{V_t}. \end{aligned} \tag{11}$$

Focusing on the dependence on t and T, our update (11) admits a per-round computational cost of $\mathcal{O}(1)$, whereas the update of Yu et al. [2025] in (3) requires $\mathcal{O}(t \log T)$ per round. Our update only requires maintaining the inverse matrix via the Sherman–Morrison–Woodbury formula [Horn and Johnson, 2012] and performing a projection, which can be computed in $\mathcal{O}(d^2)$ and $\mathcal{O}(d^3)$ time, respectively.

By combining the loss (5) using two parameters (6) and (7) with the OMD-based update (11), we obtain the estimator that is robust to both adversarial corruption and heavy-tailed noise, while maintaining $\mathcal{O}(1)$ per-round computational cost. With this estimator (11) in place, we now specify the arm-selection rule.

Algorithm 1 CR-Hvt-UCB

Input: time horizon T, corruption level C, moment parameter ϵ, norm bounds S, L, confidence δ, regularizer λ, algorithm parameters $\sigma_{\min}, \alpha$.
Set $\kappa = d\log\left(1+\frac{L^2T}{\sigma_{\min}^2\lambda\alpha d}\right)$, τ_0 as (8), $V_0 = \lambda I_d$, $\hat{\theta}_1 = \mathbf{0}$, and compute β_0 by (13).
for $t = 1, 2, \ldots, T$ **do**
  $X_t = \arg\max_{\mathbf{x}\in\mathcal{X}_t} \langle \mathbf{x}, \hat{\theta}_t\rangle + \beta_{t-1}\|\mathbf{x}\|_{V_{t-1}^{-1}}$. (ties broken arbitrarily.)
  Observe r_t and ν_t.
  Set σ_t as (6).
  $w_t = \frac{1}{\sqrt{\alpha}}\left\|\frac{X_t}{\sigma_t}\right\|_{V_{t-1}^{-1}}$.
  $\tau_t = \tau_0 \frac{\sqrt{1+w_t^2}}{w_t} t^{\frac{1-\epsilon}{2(1+\epsilon)}}$.
  $V_t = V_{t-1} + \alpha^{-1}\sigma_t^{-2} X_t X_t^\top$.
  Update $\hat{\theta}_{t+1}$ via (11), and compute β_t by (13).
end for

We adopt a standard UCB-based arm-selection rule [Auer et al., 2002, Abbasi-Yadkori et al., 2011].

$$X_t = \arg\max_{\mathbf{x}\in\mathcal{X}_t} \left\{ \langle \mathbf{x}, \hat{\theta}_t\rangle + \beta_{t-1}\|\mathbf{x}\|_{V_{t-1}^{-1}} \right\}. \quad (12)$$

Here, when the corruption level C and the $(1+\epsilon)$-moment bound ν_t are known, the confidence radius β_t is specified in the following lemma.

Lemma 4.4. *Let σ_t and τ_t be defined as in* (6) *and* (7)*, where $\sigma_{\min} = 1/\sqrt{T}$ and $\kappa = d\log\big(1+\frac{L^2T}{4\sigma_{\min}^2\lambda d}\big)$. Set the auxiliary parameters τ_0 and w_t as in* (8)*. Let the step size $\alpha = 8$. For any $\delta \in (0, \frac{1}{4})$, with probability at least $1-4\delta$, for any $t \in [T]$, $\|\hat{\theta}_{t+1} - \theta^*\|_{V_t} \le \beta_t$ holds, where*

$$\beta_t = 409\log\left(\frac{2T^2}{\delta}\right)\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} + \sqrt{\lambda(2+4S^2)}. \quad (13)$$

The overall procedure is summarized in Algorithm 1.

4.2 REGRET GUARANTEE

In this section, we now present the regret guarantee for Algorithm 1. We first establish the result under the assumption that the corruption level C and $(1+\epsilon)$-moment bounds ν_t are known to the learner. We then extend the analysis to the case where one or both of these quantities are unknown.

When both C and ν_t are known to the learner, our algorithm achieves the following instance-dependent regret bound.

Theorem 4.5. *Let σ_t, τ_t, τ_0, w_t, and α be set according to Lemma 4.4. Choose $\lambda = d$, $\sigma_{\min} = \frac{1}{\sqrt{T}}$, and $\delta = \frac{1}{8T}$. Then, with probability at least $1-1/T$, the regret of Algorithm 1 satisfies*

$$\mathrm{Reg}(T) = \tilde{\mathcal{O}}\left(dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{\sum_{t=1}^{T}\nu_t^2} + dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\cdot 1 \vee C \right).$$

A complete proof is provided in Appendix C.1.

Remark 4.6. *Our result can be viewed as a generalization of existing bounds. When $\epsilon = 1$, our regret bound coincides with the bound established by Yu et al. [2025] for finite-variance noise. Moreover, in the absence of adversarial corruption ($C = 0$), our bound recovers the regret order for linear contextual bandits with heavy-tailed noise [Huang et al., 2023, Wang et al., 2025].*

We now consider the setting where C and/or ν_t are unknown to the learner. This includes the cases where only C is unknown, only ν_t is unknown, or both are unknown. In each case, the unknown quantities are replaced by available upper bounds $(\bar{C}, \nu)$ in the definition of σ_t, where $\bar{C}$ and ν are a positive constant such that $\bar{C} \ge C$ and $\mathbb{E}[|\eta_t|^{1+\epsilon} \mid X_{1:t}, \eta_{1:t-1}, c_{1:t-1}] \le \nu^{1+\epsilon}$. Under these substitutions, Lemma 4.4 continues to hold.

When C is unknown, the regret is bounded as follows.

Corollary 4.7. *Under the parameter choices of Theorem 4.5, replace C by $\bar{C}$ in the definition of σ_t. Then, with probability at least $1-1/T$,*

$$\mathrm{Reg}(T) = \tilde{\mathcal{O}}\left(dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{\sum_{t=1}^{T}\nu_t^2} + dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\cdot 1 \vee \bar{C} \right).$$

A complete proof is provided in Appendix C.2.

When ν_t is unknown, the regret satisfies the following upper bound.

Corollary 4.8. *Under the parameter choices of Theorem 4.5, replace ν_t with ν in the definition of σ_t. Then, with probability at least $1-1/T$,*

$$\mathrm{Reg}(T) = \tilde{\mathcal{O}}\left(dT^{\frac{1}{1+\epsilon}} + dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\cdot 1 \vee C \right).$$

A complete proof is provided in Appendix C.3.

Remark 4.9. *If $C = \mathcal{O}(\sqrt{T})$, the regret reduces to $\tilde{\mathcal{O}}(dT^{\frac{1}{1+\epsilon}})$, which coincides with the uncorrupted rate ($C = 0$). This bound does not match the lower bound of order $\tilde{\Omega}(d^{\frac{2\epsilon}{1+\epsilon}}T^{\frac{1}{1+\epsilon}})$ under bounded $(1+\epsilon)$-moment assumptions [Tajdini et al., 2025]. Nevertheless, it matches the best-known regret order achieved by algorithms for linear contextual bandits with heavy-tailed noise.*

When both C and ν_t are unknown, we obtain the following regret guarantee.

Corollary 4.10. *Under the parameter choices of Theorem 4.5, replace (C, ν_t) by $(\bar{C}, \nu)$ in the definition of σ_t. Then, with probability at least $1 - 1/T$,*

$$\mathrm{Reg}(T) = \tilde{\mathcal{O}}\left(dT^{\frac{1}{1+\epsilon}} + dT^{\frac{1-\epsilon}{2(1+\epsilon)}} \cdot 1 \vee \bar{C}\right).$$

Proof. The claim follows by combining the proofs of Corollary 4.7 and Corollary 4.8. □

5 PROOF SKETCH

In this section, we focus on the technical ideas underlying the proof of Lemma 4.4. Once this lemma is obtained, the regret guarantee follows from standard UCB arguments.

Lemma 4.4 is established by adapting the OMD-based error decomposition introduced in Lemma 3 of Wang et al. [2025]. The main additional challenge is that the gradients of the Huber loss (5) depend not only on the heavy-tailed stochastic noise but also on the adversarial corruption through $z_t(\theta)$. Therefore, unlike in Wang et al. [2025], we cannot directly apply concentration arguments based only on the stochastic-noise assumptions. To overcome this difficulty, we separate the corruption-independent and corruption-dependent contributions throughout the OMD analysis. This separation allows us to handle the corruption-independent terms using arguments similar to Wang et al. [2025], while controlling the corruption-dependent terms via the scale parameters $\{\sigma_t\}_{t=1}^T$ and the corruption level C.

Adapting Lemma 3 of Wang et al. [2025], we first implement this separation at the level of the estimation error. The estimation error can be decomposed as follows (formal statement in Lemma B.1):

$$\begin{aligned}\|\hat{\theta}_{t+1} - \theta^*\|_{V_t}^2 \le\ & \underbrace{4\lambda S^2}_{\text{regularization term}} + \underbrace{\sum_{s=1}^{t} \left\|\nabla \ell_s(\hat{\theta}_s)\right\|_{V_s^{-1}}^2}_{\text{stability term}} \\ & + \underbrace{2\sum_{s=1}^{t} \left\langle \nabla \tilde{\ell}_s(\hat{\theta}_s) - \nabla \ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^* \right\rangle}_{\text{generalization-gap term}} \\ & + \underbrace{\left(\frac{1}{\alpha} - 1\right) \sum_{s=1}^{t} \left\|\hat{\theta}_s - \theta^*\right\|_{X_s X_s^\top / \sigma_s^2}}_{\text{negative term}}. \end{aligned} \tag{14}$$

Here, $\ell_s(\theta)$ is the loss in (5), while $\tilde{\ell}_s(\theta)$ denotes the noise- and corruption-free Huber loss (4), evaluated at $\tilde{z}_s(\theta) = \langle X_s, \theta^* - \theta \rangle / \sigma_s$.

In this decomposition, the effect of adversarial corruption appears in both the stability term and the generalization-gap term through the gradient $\nabla \ell_s(\hat{\theta}_s)$. Thus, the key step is to separate the corruption-independent and corruption-dependent contributions within these two terms. We explain this separation and the resulting bounds for each term below.

Stability term analysis. To isolate the contribution of adversarial corruption, we first decompose the squared gradient norm at each round s as follows:

$$\begin{aligned}\left\|\nabla \ell_s(\hat{\theta}_s)\right\|_{V_s^{-1}}^2 \le\ & 3 \underbrace{\left(\min\left\{\left|\frac{\eta_s}{\sigma_s}\right|, \tau_s\right\}\right)^2 \left\|\frac{X_s}{\sigma_s}\right\|_{V_s^{-1}}^2}_{\text{stochastic term}} \\ & + 3 \underbrace{\left(\frac{\langle X_s, \theta^* - \hat{\theta}_s \rangle}{\sigma_s}\right)^2 \left\|\frac{X_s}{\sigma_s}\right\|_{V_s^{-1}}^2}_{\text{deterministic term}} \\ & + 3 \underbrace{\left(\frac{c_s}{\sigma_s}\right)^2 \left\|\frac{X_s}{\sigma_s}\right\|_{V_s^{-1}}^2}_{\text{corruption term}}. \end{aligned}$$

Summing over $s = 1, \ldots, t$, the first two terms coincide with the corresponding quantities in Wang et al. [2025] and therefore admit the same bound $\tilde{\mathcal{O}}(dt^{\frac{1-\epsilon}{1+\epsilon}})$. The main new difficulty is the third term, which captures the cumulative effect of adversarial corruption and has no counterpart in Wang et al. [2025]. Although bounding each c_s by the corruption level C introduces a C^2 factor, this factor is offset by the corruption-dependent component in the scale parameter σ_s. As a result, the cumulative contribution of the corruption term is only $\tilde{\mathcal{O}}(d)$ and does not affect the overall order of the stability bound. The complete proof of this analysis is given in Appendix B.2.

Generalization-gap term analysis. For the generalization-gap term, we first follow Wang et al. [2025] and further decompose it into the Huber loss term and the self-regularization term at each round s:

$$\begin{aligned}& \left\langle \nabla \tilde{\ell}_s(\hat{\theta}_s) - \nabla \ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^* \right\rangle \\ &= \underbrace{\left\langle \nabla \tilde{\ell}_s(\hat{\theta}_s) + \nabla \ell_s(\theta^*) - \nabla \ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^* \right\rangle}_{\text{Huber loss term}} \\ &\quad + \underbrace{\left\langle -\nabla \ell_s(\theta^*), \hat{\theta}_s - \theta^* \right\rangle}_{\text{Self-regularization term}}. \end{aligned}$$

As in Wang et al. [2025], the Huber loss term is supported only on the event $\{|z_s(\theta^*)| > \tau_s/2\}$. The main difficulty is that $z_s(\theta^*) = (\eta_s + c_s)/\sigma_s$ depends on both the stochastic noise and the adversarial corruption. Consequently, the tail-event argument of Wang et al. [2025] cannot be applied directly. To separate these two effects, we apply a union bound:

$$\mathbb{1}\left\{|z_s(\theta^*)| > \tfrac{\tau_s}{2}\right\} \le \mathbb{1}\left\{\left|\tfrac{\eta_s}{\sigma_s}\right| > \tfrac{\tau_s}{4}\right\} + \mathbb{1}\left\{\left|\tfrac{c_s}{\sigma_s}\right| > \tfrac{\tau_s}{4}\right\}.$$

The corruption-independent part can be bounded exactly as in Wang et al. [2025]. Meanwhile, our corruption-aware choices of τ_s and σ_s ensure that the corruption-dependent part is no larger in order. Consequently, the Huber loss term admits the same bound $\tilde{\mathcal{O}}(\sqrt{d}t^{\frac{1-\varepsilon}{2(1+\varepsilon)}}\beta_t)$ as in Wang et al. [2025]. For the self-regularization term, by the chain rule, we have $\nabla \ell_s(\theta) = -h_s\left(z_s(\theta)\right)\frac{X_s}{\sigma_s}$, where $h_s : \mathbb{R} \to \mathbb{R}$ denotes the derivative function of the Huber loss (4). Although h_s satisfies $|h_s(\cdot)| \le \tau_s$, directly using this boundedness to control the corrupted gradient would introduce an additional dependence on τ_s, leading to a looser dependence on t. To address this issue, we instead exploit the 1-Lipschitz property of h_s: $\left|h_s((\eta_s + c_s)/\sigma_s) - h_s(\eta_s/\sigma_s)\right| \le |c_s|/\sigma_s$. This inequality explicitly isolates the corruption contribution:

$$\left\langle -\nabla \ell_s(\theta^*), \hat{\theta}_s - \theta^* \right\rangle = \left| \left\langle h_s\left(\frac{\eta_s + c_s}{\sigma_s}\right)\frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^* \right\rangle \right|$$
$$\le \left| h_s\left(\frac{\eta_s}{\sigma_s}\right)\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^* \right\rangle \right| + \frac{|c_s|}{\sigma_s}\left| \left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^* \right\rangle \right|.$$

Summing over $s = 1, \ldots, t$, the corruption-independent part is bounded exactly as in Wang et al. [2025]. For the corruption-dependent part, we apply Cauchy–Schwarz to separate the feature-dependent factor from the parameter-error factor. The former is controlled together with $|c_s|/\sigma_s$ using the corruption level C and the corruption-aware scale parameter σ_s, while the latter is bounded by the confidence radius β_s. Combining these bounds, the resulting contribution is of the same order as the Huber loss term. Consequently, the overall bound for the self-regularization term coincides with that of Wang et al. [2025]. The complete proof of this analysis is given in Appendix B.3.

Therefore, despite the presence of adversarial corruption, both the stability and generalization-gap terms are controlled at the same order as in Wang et al. [2025]. Combining these bounds in the OMD-based error decomposition (14) yields the same overall rate, so a confidence radius of $\tilde{\mathcal{O}}(\sqrt{d}t^{\frac{1-\epsilon}{2(1+\epsilon)}})$ suffices to establish Lemma 4.4. The complete proof of Lemma 4.4 is given in Appendix B.4.

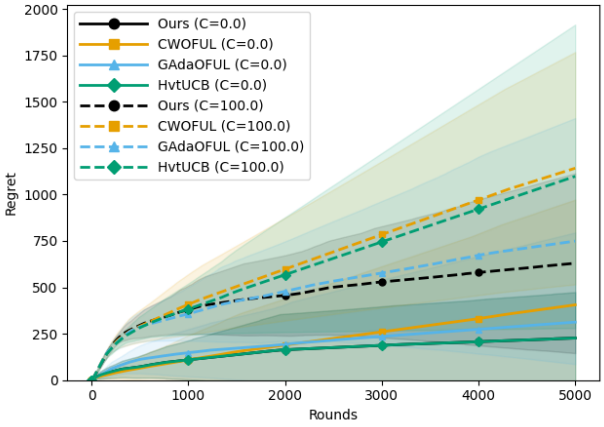
6 EXPERIMENTS

In this section, we numerically evaluate the effectiveness of our algorithm (Algorithm 1) by comparing it with existing algorithms. Additional experiments comparing our method with a rarely-switching variant of GAdaOFUL [Yu et al., 2025] based on Abbasi-Yadkori et al. [2011] are presented in Appendix E. We further evaluate the robustness of our algorithm under different heavy-tailed noise distributions, including Fisk and Lomax distributions, in Appendix E. We consider a linear contextual bandit with time horizon $T = 5000$ and feature dimension $d = 10$. We sample θ^* uniformly from the unit sphere in $\mathbb{R}^d$. At each round t, the decision set $\mathcal{X}_t$ consists of 20 independent random unit vectors in $\mathbb{R}^d$ generated in the same manner. Thus, in accordance with Assumption 3.2, we set $L = 1$ and $S = 1$. The stochastic reward r'_t is generated as in (2). Following Bogunovic et al. [2021], we adopt the θ-flipping corruption mechanism. The adversary flips the sign of $\langle X_t, \theta^* \rangle$ until the cumulative amount of corruption reaches C. Specifically, for corrupted rounds we observe $r_t = -\langle X_t, \theta^* \rangle + \eta_t$, while the remaining rounds are left uncorrupted, i.e., $r_t = r'_t$. In our experiments, we consider the cases $C = 0$ and $C = 100$. We investigate two noise models:

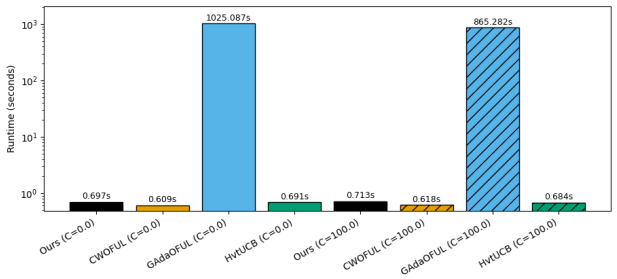
- **In the case of $\epsilon = 1$:** Following Yu et al. [2025], we generate the noise from a Student's t-distribution with 3 degrees of freedom. This distribution has finite variance ($\mathbb{E}[\eta_t^2] = 3$) and satisfies Assumption 3.1 with $\epsilon = 1$ and $\nu_t = \nu = \sqrt{3}$.
- **In the case of $\epsilon < 1$:** We generate the noise from a Pareto distribution with shape parameter $\alpha = 1.5$ and minimum value $x_{\min} = 1$. This distribution has a finite $(1+\epsilon)$-th moment for $\epsilon = 0.4$, and satisfies Assumption 3.1 with $\nu_t = \nu = 15^{\frac{1}{1+\epsilon}}$.

We first compare CR-Hvt-UCB with representative existing methods under finite-variance noise, corresponding to $\epsilon = 1$, namely CWOFUL [He et al., 2022], HvtUCB [Wang et al., 2025], and GAdaOFUL [Yu et al., 2025] (Figures 1a and 1b). We aim to evaluate whether CR-Hvt-UCB can achieve regret performance comparable to existing corruption-robust methods, while retaining the computational efficiency of $\mathcal{O}(1)$-update methods. We then compare CR-Hvt-UCB with GAdaOFUL under heavier-tailed noise with $\epsilon < 1$ (Figure 1c and 1d). This experiment highlights the effect of noise robustness on empirical performance, as GAdaOFUL relies on a finite-variance assumption while CR-Hvt-UCB only requires bounded $(1+\epsilon)$ moments. We also compare their running times to assess the computational efficiency of the two approaches. Each experiment is averaged over 10 independent runs, and the shaded region indicates one standard deviation from the mean.

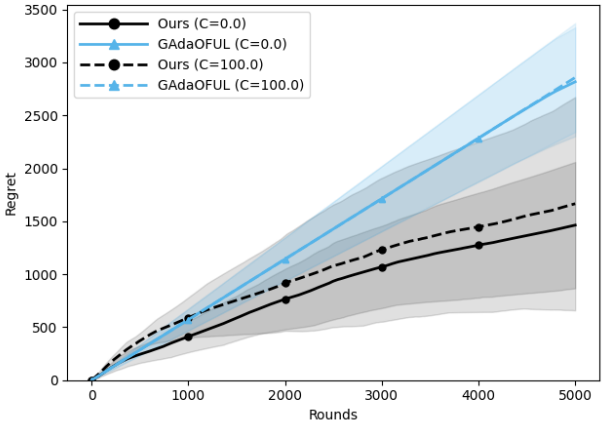
Figure 1a shows the regret comparison. When $C = 0$, Ours and HvtUCB exhibit almost identical regret, as expected, since the two algorithms coincide in the absence of corruption. GAdaOFUL also achieves a comparable regret in this setting, whereas CWOFUL suffers nearly linear regret. When $C = 100$, Ours and GAdaOFUL maintain similar regret performance, while HvtUCB deteriorates significantly and performs comparably to CWOFUL. This indicates that HvtUCB is not robust to adversarial corruption, whereas Ours preserves robustness to corruption comparable to GAdaOFUL. Figure 1b shows the computational cost. In both the $C = 0$ and $C = 100$ settings, the algorithms with $\mathcal{O}(1)$ per-round updates have comparable running times. In contrast, GAdaOFUL requires substantially larger running time, despite achieving similar regret to Ours. Thus, Ours achieves regret performance comparable to GAdaOFUL



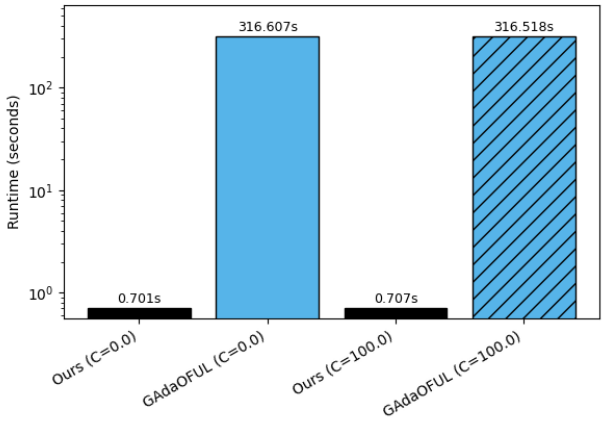
(a) Regret ($\epsilon = 1$)



(b) Runtime ($\epsilon = 1$)



(c) Regret ($\epsilon < 1$)



(d) Runtime ($\epsilon < 1$)

Figure 1: Comparison of Ours and existing works

while being much more computationally efficient.

Figure 1c reveals a clear contrast between the two methods. For GAdaOFUL, the regret exhibits nearly linear growth both in the absence of corruption ($C = 0$) and under corruption ($C = 100$). This indicates that GAdaOFUL fails to achieve sublinear regret even without corruption in the case of $\epsilon < 1$. In contrast, although corruption worsens the regret, our algorithm achieves sublinear regret for both settings. Moreover, Figure 1d further confirms the computational advantage of Ours in the heavy-tailed setting. Even under heavy-tailed noise, Ours runs several hundred times faster than GAdaOFUL.

7 CONCLUSION

In this work, we study linear contextual bandits under adversarial corruption and heavy-tailed noise. We highlight the limitations of existing approaches in terms of robustness to heavy-tailed noise and computational efficiency. To address these limitations, we propose CR-Hvt-UCB (Algorithm 1). The algorithm is designed to be robust to both adversarial corruption and heavy-tailed noise and is computationally efficient via OMD update. The central mechanism underlying our algorithm is the adaptive scaling σ_t, which mitigates the effect of adversarial corruption on parameter estimation. When both the corruption level C and the $(1+\epsilon)$-moment upper bounds ν_t are known, our algorithm achieves $\tilde{\mathcal{O}}(dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{\sum_{t=1}^{T}\nu_t^2} + dT^{\frac{1-\epsilon}{2(1+\epsilon)}}(1 \vee C))$ regret. We further provide regret guarantees for the settings where C or ν_t, or both, are unknown.

In the corruption-free case ($C = 0$), our regret rate is optimal among algorithms designed for linear contextual bandits with general decision sets under bounded $(1+\epsilon)$-moment assumptions. However, it does not attain the lower bound for heavy-tailed linear bandits [Tajdini et al., 2025]. Achieving this lower bound while allowing general decision sets remains an important direction for future work. Another direction is to extend our framework beyond linear models to GLBs [Yu et al., 2025].

Acknowledgements

This work was supported by Japan Science and Technology Agency (JST) as part of Adopting Sustainable Partnerships for Innovative Research Ecosystem (ASPIRE), Grant Number JPMJAP25B1. TN was also supported by JST SPRING, Grant Number JPMJSP2138.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 127–135, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2):235–256, 2002.

Ilija Bogunovic, Arpan Losalka, Andreas Krause, and Jonathan Scarlett. Stochastic linear bandits robust to adversarial attacks. In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 991–999. PMLR, 13–15 Apr 2021.

Sébastien Bubeck, Nicolò Cesa-Bianchi, and Gábor Lugosi. Bandits with heavy tail. *IEEE Transactions on Information Theory*, 59(11):7711–7717, 2013. doi: 10.1109/TIT.2013.2277869.

P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2): 298–311, 1997. doi: 10.1109/83.551699.

Yu Chen, Jiatai Huang, Yan Dai, and Longbo Huang. uniinf: Best-of-both-worlds algorithm for parameter-free heavy-tailed mabs. In *International Conference on Learning Representations*, volume 2025, pages 71762–71798, 2025.

Duo Cheng, Xingyu Zhou, and Bo Ji. Taming heavy-tailed losses in adversarial bandits and the best-of-both-worlds setting. *Advances in Neural Information Processing Systems*, 37:48083–48129, 2024.

Chris Dann, Chen-Yu Wei, and Julian Zimmert. A blackbox approach to best of both worlds in bandits and beyond. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 5503–5570. PMLR, 2023.

Qin Ding, Cho-Jui Hsieh, and James Sharpnack. Robust stochastic linear contextual bandits under adversarial attacks. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 7111–7123. PMLR, 28–30 Mar 2022.

Anupam Gupta, Tomer Koren, and Kunal Talwar. Better algorithms for stochastic bandits with adversarial corruptions. In *Conference on Learning Theory*, pages 1562–1578. PMLR, 2019.

Jiafan He, Dongruo Zhou, Tong Zhang, and Quanquan Gu. Nearly optimal algorithms for linear contextual bandits with adversarial corruptions. *Advances in neural information processing systems*, 35:34614–34625, 2022.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

Jiatai Huang, Yan Dai, and Longbo Huang. Adaptive best-of-both-worlds algorithm for heavy-tailed multi-armed bandits. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 9173–9200. PMLR, 17–23 Jul 2022.

Jiayi Huang, Han Zhong, Liwei Wang, and Lin Yang. Tackling heavy-tailed rewards in reinforcement learning with function approximation: Minimax optimal and instance-dependent regret bounds. *Advances in Neural Information Processing Systems*, 36:56576–56588, 2023.

Peter J Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

Shinji Ito and Kei Takemura. An exploration-by-optimization approach to best of both worlds in linear bandits. *Advances in Neural Information Processing Systems*, 36:71582–71602, 2023a.

Shinji Ito and Kei Takemura. Best-of-three-worlds linear bandit algorithm with variance-adaptive regret bounds. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 2653–2677. PMLR, 12–15 Jul 2023b.

Shinji Ito, Taira Tsuchiya, and Junya Honda. Adaptive learning rate for follow-the-regularized-leader: Competitive analysis and best-of-both-worlds. In Shipra Agrawal and Aaron Roth, editors, *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 2522–2563. PMLR, 30 Jun–03 Jul 2024.

Charles Jebarajakirthy, Haroon Iqbal Maseeh, Zakir Morshed, Amit Shankar, Denni Arli, and Robin Pentecost. Mobile advertising: A systematic literature review and future research agenda. *International Journal of Consumer Studies*, 45(6):1258–1291, 2021.

Minhyun Kang and Gi-Soo Kim. Heavy-tailed linear bandit with Huber regression. In Robin J. Evans and Ilya Shpitser, editors, *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pages 1027–1036. PMLR, 31 Jul–04 Aug 2023.

Masahiro Kato and Shinji Ito. Best-of-both-worlds linear contextual bandits. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Masahiro Kato and Shinji Ito. Lc-tsallis-inf: Generalized best-of-both-worlds linear contextual bandits. In Yingzhen Li, Stephan Mandt, Shipra Agrawal, and Emtiyaz Khan, editors, *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 3655–3663. PMLR, 03–05 May 2025.

Sanghwa Kim, Junghyun Lee, and Se-Young Yun. A jointly efficient and optimal algorithm for heteroskedastic generalized linear bandits with adversarial corruptions. *arXiv preprint arXiv:2602.10971*, 2026.

Fang Kong, Canzhe Zhao, and Shuai Li. Best-of-three-worlds analysis for linear bandits with follow-the-regularized-leader algorithm. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 657–673. PMLR, 12–15 Jul 2023.

Yuko Kuroki, Alberto Rumi, Taira Tsuchiya, Fabio Vitale, and Nicolò Cesa-Bianchi. Best-of-both-worlds algorithms for linear contextual bandits. In Sanjoy Dasgupta, Stephan Mandt, and Yingzhen Li, editors, *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 1216–1224. PMLR, 02–04 May 2024.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Chung-Wei Lee, Haipeng Luo, Chen-Yu Wei, Mengxiao Zhang, and Xiaojin Zhang. Achieving near instance-optimality and minimax-optimality in stochastic and adversarial linear bandits simultaneously. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 6142–6151. PMLR, 18–24 Jul 2021.

Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, page 661–670, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605587998. doi: 10.1145/1772690.1772758.

Xiang Li and Qiang Sun. Variance-aware robust reinforcement learning with linear function approximation under heavy-tailed rewards. *arXiv preprint arXiv:2303.05606*, 2023.

Yingkai Li, Edmund Y Lou, and Liren Shan. Stochastic linear optimization with adversarial corruption. *arXiv preprint arXiv:1909.02109*, 2019.

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

Andres Munoz Medina and Scott Yang. No-regret algorithms for heavy-tailed linear bandits. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1642–1650, New York, New York, USA, 20–22 Jun 2016. PMLR.

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

Svetlozar Todorov Rachev. *Handbook of heavy tailed distributions in finance: Handbooks in finance, Book 1*, volume 1. Elsevier, 2003.

Han Shao, Xiaotian Yu, Irwin King, and Michael R Lyu. Almost optimal algorithms for linear stochastic bandits with heavy-tailed payoffs. *Advances in Neural Information Processing Systems*, 31, 2018.

Qiang Sun, Wen-Xin Zhou, and Jianqing Fan. Adaptive huber regression. *Journal of the American Statistical Association*, 115(529):254–265, 2020.

Artin Tajdini, Jonathan Scarlett, and Kevin Jamieson. Improved regret bounds for linear bandits with heavy-tailed rewards. In *NeurIPS 2025 Workshop: Second Workshop on Aligning Reinforcement Learning Experimentalists and Theorists*, 2025.

Jing Wang, Yu-Jie Zhang, Peng Zhao, and Zhi-Hua Zhou. Heavy-tailed linear bandits: Huber regression with one-pass update. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj,

Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 65595–65617. PMLR, 13–19 Jul 2025.

Chen-Yu Wei, Christoph Dann, and Julian Zimmert. A model selection approach for corruption robust reinforcement learning. In *International Conference on Algorithmic Learning Theory*, pages 1043–1096. PMLR, 2022.

Bo Xue, Guanghui Wang, Yimu Wang, and Lijun Zhang. Nearly optimal regret for stochastic linear bandits with heavy-tailed payoffs. In *In Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, IJCAI'20, 2021. ISBN 9780999241165.

Bo Xue, Yimu Wang, Yuanyu Wan, Jinfeng Yi, and Lijun Zhang. Efficient algorithms for generalized linear bandits with heavy-tailed rewards. *Advances in Neural Information Processing Systems*, 36:70880–70891, 2023.

Qingyuan Yu, Euijin Baek, Xiang Li, and Qiang Sun. Corruption-robust variance-aware algorithms for generalized linear bandits under heavy-tailed rewards. In Silvia Chiappa and Sara Magliacane, editors, *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pages 4826–4843. PMLR, 21–25 Jul 2025.

Yu-Jie Zhang and Masashi Sugiyama. Online (multinomial) logistic bandit: Improved regret and constant computation cost. *Advances in Neural Information Processing Systems*, 36, 2024.

Yu-Jie Zhang, Sheng-An Xu, Peng Zhao, and Masashi Sugiyama. Generalized linear bandits: Almost optimal regret with one-pass update. *Advances in Neural Information Processing Systems*, 38:69244–69277, 2026.

Canzhe Zhao, Shinji Ito, and Shuai Li. Heavy-tailed linear bandits: Adversarial robustness, best-of-both-worlds, and beyond. *arXiv preprint arXiv:2508.13679*, 2025.

Heyang Zhao, Dongruo Zhou, and Quanquan Gu. Linear contextual bandits with adversarial corruptions. *arXiv preprint arXiv:2110.12615*, 2021.

Julian Zimmert and Yevgeny Seldin. Tsallis-inf: An optimal algorithm for stochastic and adversarial bandits. *Journal of Machine Learning Research*, 22(28):1–49, 2021.

Robust and Computationally Efficient Linear Contextual Bandits under Adversarial Corruption and Heavy-Tailed Noise (Supplementary Material)

Naoto Tani[1,2] **Futoshi Futami**[1,2,3]

[1]Graduate School of Engineering Science, The University of Osaka, Osaka, Japan
[2]RIKEN AIP, Tokyo, Japan
[3]Department of Complexity Science and Engineering, The University of Tokyo, Tokyo, Japan

A ADDITIONAL RELATED WORK

This appendix provides a broader overview of related work beyond the main comparisons presented in Section 2. Rather than focusing only on the closest comparisons, we summarize existing studies according to three themes: (1) bandits with adversarial corruption, (2) heavy-tailed bandits, (3) computational efficiency.

A.1 BANDIT WITH ADVERSARIAL CORRUPTION

Research on adversarial corruption in bandits has mainly followed two modeling paradigms. One models an underlying stochastic bandit corrupted by an adversary with a bounded corruption budget. The other approaches the problem from the best-of-both-worlds (BOBW) perspective, aiming to perform well in both stochastic and adversarial environments. We briefly review both directions below.

Adversarial corruption under a bounded corruption budget was first studied in the multi-armed bandit setting by [Lykouris et al., 2018, Gupta et al., 2019]. Lykouris et al. [2018] derived regret bounds in which the total corruption level C appears multiplicatively with the time horizon T. In contrast, Gupta et al. [2019] proposed an algorithm achieving an additive regret bound of the form $\mathrm{Reg}(T) = o(T) + \mathcal{O}(C)$. This additive regret bound has since become a standard benchmark for corruption-robust bandit algorithms. This line of work was later extended to linear bandits [Li et al., 2019, Bogunovic et al., 2021]. Li et al. [2019] proposed the Support Basis Exploration algorithm for stochastic linear bandits under adversarial corruption. Bogunovic et al. [2021] later developed an arm-elimination algorithm. Both algorithms assume a fixed decision set and therefore do not directly extend to the general linear contextual bandit setting considered in this work. In particular, the arm-elimination approach of Bogunovic et al. [2021] additionally relies on a finite decision set. Subsequent works [Zhao et al., 2021, Ding et al., 2022] proposed OFUL-type algorithms that are robust to adversarial corruption, and Wei et al. [2022] also studied corruption robustness under a different definition of the corruption level C. However, their regret guarantees are not optimal. These limitations were overcome by He et al. [2022], who proposed an algorithm achieving the optimal regret bound $\tilde{\mathcal{O}}(d\sqrt{T} + dC)$. More recently, Yu et al. [2025] studied generalized linear bandits (GLBs) under the same additive corruption model. Our work follows the same corruption model, while differing in the noise assumptions and computational efficiency, which are discussed in the following section.

A complementary line of research studies adversarial corruption from the BOBW perspective. Zimmert and Seldin [2021] showed that adversarial corruption can be expressed through the self-constraint framework used in BOBW analyses. This perspective has also been studied in several structured bandit settings, including linear and linear contextual bandits [Lee et al., 2021, Dann et al., 2023, Kong et al., 2023, Ito and Takemura, 2023a,b, Ito et al., 2024, Kato and Ito, 2024, Kuroki et al., 2024, Kato and Ito, 2025]. More recently, Zhao et al. [2025] extended this line of research to heavy-tailed linear bandits. Unlike our work, this work considers linear bandits with a fixed finite decision set under the BOBW framework. Their algorithms are based on FTRL over distributions on a fixed finite decision set, so the policy update reduces to a finite-dimensional optimization over the simplex. In contrast, our setting considers general contextual decision sets, which may be infinite and vary with the context, rather than a fixed finite decision set. As a result, it is unclear whether such finite-arm distributional updates can be applied to our setting. We discuss the differences in the noise assumptions in the

next section.

A.2 HEAVY-TAILED BANDITS

Research on heavy-tailed bandits began in the multi-armed bandit setting. Bubeck et al. [2013] studied multi-armed bandits under the assumption that the rewards admit bounded $(1+\epsilon)$-th moments for some $\epsilon \in (0, 1]$. They proposed robust algorithms based on truncation and median-of-means (MoM) estimators. Building on the truncation- and MoM-based approaches of Bubeck et al. [2013], Medina and Yang [2016] studied linear bandits under the assumption that the noise admits bounded $(1+\epsilon)$-th moments. Under the same bounded $(1+\epsilon)$-th moment assumption on the noise, Shao et al. [2018] established a lower bound of order $\Omega(dT^{\frac{1}{1+\epsilon}})$ for heavy-tailed linear contextual bandits. This was the best-known lower bound at the time. They also proposed truncation- and MoM-based algorithms achieving regret of the same order as this lower bound. For linear bandits with a finite decision set, Xue et al. [2021] established a lower bound of order $\Omega(d^{\frac{\epsilon}{1+\epsilon}}T^{\frac{1}{1+\epsilon}})$ and proposed truncation- and MoM-based algorithms attaining the regret rate of $\tilde{\mathcal{O}}(d^{\frac{1}{2}}T^{\frac{1}{1+\epsilon}})$. Subsequently, Xue et al. [2023] extended this line of work to GLBs with general (possibly infinite) decision sets and proposed truncation- and MoM-based algorithms achieving the best-known regret rate, matching the lower bound of Shao et al. [2018]. An alternative line of work is based on Huber-type estimators. Under a fixed finite arm set with i.i.d. contexts, Kang and Kim [2023] proposed a Huber-type algorithm and established regret guarantees under bounded $(1+\epsilon)$-moment noise. In the general linear contextual bandit setting, Li and Sun [2023] proposed an algorithm based on adaptive Huber regression [Sun et al., 2020], assuming finite-variance noise and deriving variance-aware regret guarantees. This direction was further developed by Huang et al. [2023] and Wang et al. [2025], who relaxed the finite-variance assumption to bounded $(1+\epsilon)$-th moments. Under this weaker assumption, they obtained regret bounds of order $\tilde{\mathcal{O}}(dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{\sum_{t=1}^{T}\nu_t^2} + dT^{\frac{1-\epsilon}{2(1+\epsilon)}})$, where ν_t denotes an upper bound on the $(1+\epsilon)$-th moment of the noise at round t. In particular, when the moment bounds are uniformly bounded as $\nu_t \le \nu = \mathcal{O}(1)$ for all t, this regret bound reduces to $\tilde{\mathcal{O}}(dT^{\frac{1}{1+\epsilon}})$. More recently, Tajdini et al. [2025] derived an improved lower bound of order $\Omega(d^{\frac{2\epsilon}{1+\epsilon}}T^{\frac{1}{1+\epsilon}})$ for linear contextual bandits with heavy-tailed noise, and proposed an algorithm achieving $\tilde{\mathcal{O}}(d^{\frac{1+3\epsilon}{2(1+\epsilon)}}T^{\frac{1}{1+\epsilon}})$. However, their algorithm assumes a fixed decision set and therefore cannot be directly applied to general contextual settings. Consequently, among algorithms applicable to general linear contextual bandits, the order $\tilde{\mathcal{O}}(dT^{\frac{1}{1+\epsilon}})$ remains the best-known regret rate. On the other hand, Yu et al. [2025] studied linear contextual bandits under both heavy-tailed noise and adversarial corruption, but their analysis assumes finite-variance noise. We extend this line of work to the more general bounded $(1+\epsilon)$-moment noise setting.

Another line of work studies BOBW algorithms under heavy-tailed noise. Huang et al. [2022] first studied multi-armed bandits with heavy-tailed losses (rewards) under bounded $(1+\epsilon)$-moment assumptions, and proposed the first best-of-both-worlds algorithm in this setting. Subsequent works further developed this direction [Cheng et al., 2024, Chen et al., 2025]. More recently, Zhao et al. [2025] extended this line of work to heavy-tailed linear bandits and established the first BOBW guarantees in this setting. However, unlike our setting, their heavy-tailed assumption is imposed on the losses, whereas we assume bounded $(1+\epsilon)$-moment noise. Moreover, they assume a uniform upper bound on the $(1+\epsilon)$-th moments throughout the entire horizon, whereas our framework allows time-dependent moment bounds.

A.3 COMPUTATIONAL EFFICIENCY

In linear bandits, the OFUL algorithm of Abbasi-Yadkori et al. [2011] maintains a regularized least-squares estimator with a closed-form update, which admits $\mathcal{O}(1)$ computational cost per round. Similar OFUL-based methods for contextual bandits under adversarial corruption also rely on least-squares-based updates, including weighted least-squares estimators in Zhao et al. [2021] and He et al. [2022], and a bandit-over-bandit scheme combining least-squares-based UCB with EXP3 in Ding et al. [2022]. Bogunovic et al. [2021] proposed an arm-elimination-based algorithm, whose action selection can be implemented by a polynomial-time Frank-Wolfe procedure, while its parameter estimation is still based on least squares. In GLBs with heavy-tailed noise, Xue et al. [2023] proposed algorithms based on truncation and MoM estimators that admit $\mathcal{O}(1)$ per-round computational cost. In contrast, algorithms that compute a Huber-type estimator by solving a regularized empirical risk minimization problem at each round generally require iterative optimization [Huang et al., 2023, Li and Sun, 2023, Yu et al., 2025]. These methods incur a per-round cost of $\mathcal{O}(t \log T)$. To avoid such repeated optimization, recent work has used online updates for parameter estimation. Zhang and Sugiyama [2024] first demonstrated that OMD-based estimators can achieve $\mathcal{O}(1)$ per-round updates in logistic bandits, and Zhang et al. [2026] extended this framework to GLBs. Building on this idea, Wang et al. [2025] developed an OMD-based estimator for heavy-tailed linear contextual bandits under

bounded $(1+\epsilon)$-moment noise. More recently, Kim et al. [2026] improved the computational efficiency of Yu et al. [2025] by introducing OMD-based updates for GLBs with adversarial corruption. As discussed in Remark 3 of Kim et al. [2026], however, their model is more structured than that of Yu et al. [2025]: they assume a heteroskedastic exponential-family distribution, whereas Yu et al. [2025] only require an upper bound on the conditional variance of the noise. Under these stronger structural assumptions, including self-concordance, they obtain sharper instance-wise regret bounds. Consequently, their proof technique does not directly carry over to our moment-based heavy-tailed setting. In contrast, our work develops OMD-based efficient updates under bounded $(1+\epsilon)$-moment noise with adversarial corruption.

B ANALYSIS FOR THE CONFIDENCE INTERVAL

We first introduce the notation used in the proof of Lemma 4.4. In particular, we formalize the properties of the Huber loss, define the clean (corruption- and noise-free) counterpart of the loss, and specify the scale and threshold parameters that control the effect of heavy-tailed noise and adversarial corruption.

Property 1 (Property 1 of Huang et al. [2023] and Wang et al. [2025])**.** *Let $f_\tau(x)$ be the Huber loss defined in* (4)*. The following properties hold:*

1. $|f'_\tau(x)| = \min\{|x|, \tau\}$.
2. $f'_\tau(x) = \tau f'_1\left(\frac{x}{\tau}\right)$.
3. *For any* $\epsilon \in (0,1]$, $f'_1(x)$ *satisfies* $-\log\left(1 - x + |x|^{1+\epsilon}\right) \le f'_1(x) \le \log\left(1 + x + |x|^{1+\epsilon}\right)$.
4. $f''_\tau(x) = \mathbb{1}\{|x| \le \tau\}$.

Recall that $z_t(\theta) := \frac{r_t - \langle \theta, X_t\rangle}{\sigma_t}$ is the normalized prediction error. Based on Property 1, the gradient of the loss (5) can be written as

$$\nabla \ell_t(\theta) = \begin{cases} -z_t(\theta)\dfrac{X_t}{\sigma_t}, & \text{if } |z_t(\theta)| \le \tau_t, \\ -\tau_t \dfrac{X_t}{\sigma_t}, & \text{if } z_t(\theta) > \tau_t, \\ \tau_t \dfrac{X_t}{\sigma_t}, & \text{if } z_t(\theta) < -\tau_t. \end{cases} \tag{15}$$

Moreover, the norms of the gradient and the Hessian are given by

$$\left\|\nabla \ell_t(\theta)\right\|_2 = \left\| \min\left\{|z_t(\theta)|, \tau_t\right\} \frac{X_t}{\sigma_t} \right\|_2, \tag{16}$$

$$\nabla^2 \ell_t(\theta) = \mathbb{1}\left\{|z_t(\theta)| \le \tau_t\right\} \frac{X_t X_t^\top}{\sigma_t^2}. \tag{17}$$

Clean Huber loss. We introduce the clean (noise-free and corruption-free) loss to separate the stochastic noise and adversarial corruption effects. We define

$$\tilde{z}_t(\theta) := \frac{X_t^\top(\theta^* - \theta)}{\sigma_t},$$

which we call the clean normalized prediction error. Based on $\tilde{z}_t(\theta)$, we define the corresponding clean Huber loss:

$$\tilde{\ell}_t(\theta) := \begin{cases} \frac{1}{2}\tilde{z}_t(\theta)^2, & \text{if } |\tilde{z}_t(\theta)| \le \tau_t, \\ \tau_t|\tilde{z}_t(\theta)| - \frac{\tau_t^2}{2}, & \text{if } |\tilde{z}_t(\theta)| > \tau_t. \end{cases}$$

Similarly to $\ell_t(\theta)$, the gradient and Hessian of $\tilde{\ell}_t(\theta)$ take the form

$$\nabla \tilde{\ell}_t(\theta) = \begin{cases} -\tilde{z}_t(\theta)\dfrac{X_t}{\sigma_t}, & \text{if } |\tilde{z}_t(\theta)| \le \tau_t, \\ -\tau_t \dfrac{X_t}{\sigma_t}, & \text{if } \tilde{z}_t(\theta) > \tau_t, \\ \tau_t \dfrac{X_t}{\sigma_t}, & \text{if } \tilde{z}_t(\theta) < -\tau_t, \end{cases} \tag{18}$$

$$\nabla^2 \tilde{\ell}_t(\theta) = \mathbb{1}\{|\tilde{z}_t(\theta)| \le \tau_t\} \frac{X_t X_t^\top}{\sigma_t^2}. \tag{19}$$

Huber threshold parameter. Recall that the threshold parameter τ_t is defined as

$$\tau_t := \tau_0 \frac{\sqrt{1+w_t^2}}{w_t} t^{\frac{1-\epsilon}{2(1+\epsilon)}}, \qquad \tau_0 := \frac{\sqrt{2\kappa}(\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}}{\left(\log \frac{2T^2}{\delta}\right)^{\frac{1}{1+\epsilon}}}. \tag{20}$$

Here, w_t is given by

$$w_t := \frac{1}{\sqrt{\alpha}} \left\| \frac{X_t}{\sigma_t} \right\|_{V_{t-1}^{-1}}. \tag{21}$$

Scale parameter. Recall that the scale parameter σ_t is defined as

$$\sigma_t := \max\left\{ \nu_t, \sigma_{\min}, \sqrt{\frac{2\beta_{t-1}}{\tau_0 \sqrt{\alpha} t^{\frac{1-\epsilon}{2(1+\epsilon)}}}} \|X_t\|_{V_{t-1}^{-1}}, \sqrt{C}\kappa^{-1/4} \|X_t\|_{V_{t-1}^{-1}}^{1/2} \right\}. \tag{22}$$

Good event. To state high-probability guarantees, we define the following confidence event

$$\mathcal{A}_t := \left\{ \forall s \in [t], \left\| \hat{\theta}_s - \theta^* \right\|_{V_{s-1}} \le \beta_{s-1} \right\}. \tag{23}$$

B.1 KEY LEMMA

In this section, we present three lemmas that are key to the proof of Lemma 4.4. The Lemma B.1, established by Lemma 3 of Wang et al. [2025], provides a decomposition of the estimation error into four components. Based on this decomposition, Lemma B.2 analyzes the stability term, while Lemma B.3 analyzes the generalization-gap term.

Lemma B.1 (Lemma 3 in Wang et al. [2025])**.** *Suppose the good event $\mathcal{A}_t$ holds, and let τ_t be chosen as in* (20)*. Assume that σ_t satisfies*

$$\sigma_t^2 \ge \frac{2\|X_t\|_{V_{t-1}^{-1}}^2 \beta_{t-1}}{\sqrt{\alpha}\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}}}. \tag{24}$$

Then the estimation error admits the decomposition

$$\begin{aligned} \|\hat{\theta}_{t+1} - \theta^*\|_{V_t}^2 \le & \underbrace{4\lambda S^2}_{\text{(i) regularization term}} + \underbrace{\sum_{s=1}^{t} \left\| \nabla \ell_s(\hat{\theta}_s) \right\|_{V_s^{-1}}^2}_{\text{(ii) stability term}} + \underbrace{2\sum_{s=1}^{t} \left\langle \nabla \tilde{\ell}_s(\hat{\theta}_s) - \nabla \ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^* \right\rangle}_{\text{(iii) generalization-gap term}} \\ & + \underbrace{\left(\frac{1}{\alpha} - 1\right) \sum_{s=1}^{t} \|\hat{\theta}_s - \theta^*\|_{\frac{X_s X_s^\top}{\sigma_s^2}}}_{\text{(iv) negative term}}. \end{aligned} \tag{25}$$

The proof follows the same argument as in Wang et al. [2025]. Although our model additionally allows adversarial corruption, the Lemma B.1 depends only on the good event $\mathcal{A}_t$, condition (24), and the construction of τ_t, all of which are independent of the adversarial corruption. Hence, the same argument applies.

The four terms respectively correspond to (i) the regularization term reflecting prior knowledge, (ii) the stability term, which quantifies how stably the iterates evolve, (iii) the generalization-gap term, which captures the effect of adversarial corruption and heavy-tailed noise, and (iv) a negative term characteristic of OMD analyses, involving the estimation error.

Lemma B.2 (Stability Lemma)**.** *Let τ_t be defined in* (20) *and choose σ_t as*

$$\sigma_t = \max\left\{ \nu_t, \sigma_{\min}, 2\sqrt{2}\|X_t\|_{V_{t-1}^{-1}}, \sqrt{C}\kappa^{-1/4} \|X_t\|_{V_{t-1}^{-1}}^{1/2} \right\}, \tag{26}$$

where $\kappa = d \log\left(1 + \frac{L^2 T}{\sigma_{\min}^2 \lambda \alpha d}\right)$. Let $\alpha > 0$ be a constant stepsize parameter in the OMD update. Then with probability at least $1-\delta$, for all $t \ge 1$,

$$\sum_{s=1}^{t} \left\| \nabla \ell_s(\hat{\theta}_s) \right\|_{V_s^{-1}}^2 \le 18\alpha \left[t^{\frac{1-\epsilon}{2(1+\epsilon)}} \sqrt{2\kappa} (\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}} \left(\log(2T^2/\delta)\right)^{\frac{\epsilon}{1+\epsilon}} \right]^2 + \frac{3}{8} \sum_{s=1}^{t} \left\| \theta^* - \hat{\theta}_s \right\|_{\frac{X_s X_s^\top}{\sigma_s^2}}^2 + 3\kappa.$$

Lemma B.3 (Generalization-gap Lemma)**.** *Let τ_t be defined in* (20)*. We choose σ_t as*

$$\sigma_t = \max\left\{\nu_t, \sigma_{\min}, 2\sqrt{2}\|X_t\|_{V_{t-1}^{-1}}, \sqrt{\frac{2\beta_{t-1}}{\sqrt{\alpha}\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}}}}\|X_t\|_{V_{t-1}^{-1}}, \sqrt{C}\kappa^{-1/4}\|X_t\|_{V_{t-1}^{-1}}^{1/2}\right\}. \tag{27}$$

The event $\mathcal{A}_t$ is defined as in (23). Then, with probability at least $1-3\delta$, for any $t \geq 1$,

$$\begin{aligned}
&\sum_{s=1}^{t}\left\langle \nabla\tilde{\ell}_s(\hat{\theta}_s) - \nabla\ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^*\right\rangle \mathbb{1}_{\mathcal{A}_s} \\
&\leq 65\sqrt{\alpha+\tfrac{1}{8}}\log\frac{2T^2}{\delta}\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}}\max_{u\in[t+1]}\beta_{u-1} + 12\sqrt{1+\tfrac{1}{8\alpha}}\sqrt{\kappa}t^{\frac{1-\epsilon}{2(1+\epsilon)}}\max_{u\in[t+1]}\beta_{u-1} + \frac{1}{4}\left(\lambda\alpha + \sum_{s=1}^{t}\left\langle\frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^*\right\rangle^2\right) \\
&\quad + \left(8t^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{2\kappa}(\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}\left(\log\frac{2T^2}{\delta}\right)^{\frac{\epsilon}{1+\epsilon}}\right)^2 + \sqrt{\kappa}\max_{u\in[t+1]}\beta_{u-1}.
\end{aligned}$$

B.2 PROOF OF LEMMA B.2

Proof. For any $t \geq 1$, the squared V_t^{-1}-norm of the gradient can be written as

$$\begin{aligned}
&\left\|\nabla\ell_t(\hat{\theta}_t)\right\|_{V_t^{-1}}^2 \\
&= \left\|\min\left\{\left|\frac{r_t - X_t^\top\hat{\theta}_t}{\sigma_t}\right|, \tau_t\right\}\frac{X_t}{\sigma_t}\right\|_{V_t^{-1}}^2 \\
&= \left\|\min\left\{\left|\frac{X_t^\top\theta^* + \eta_t + c_t - X_t^\top\hat{\theta}_t}{\sigma_t}\right|, \tau_t\right\}\frac{X_t}{\sigma_t}\right\|_{V_t^{-1}}^2 \\
&= \left\|\min\left\{\left|\frac{\eta_t}{\sigma_t} + \frac{c_t}{\sigma_t} + \frac{X_t^\top(\theta^* - \hat{\theta}_t)}{\sigma_t}\right|, \tau_t\right\}\frac{X_t}{\sigma_t}\right\|_{V_t^{-1}}^2 \\
&\leq \left(\min\left\{\left|\frac{\eta_t}{\sigma_t}\right| + \left|\frac{X_t^\top(\theta^* - \hat{\theta}_t)}{\sigma_t}\right| + \left|\frac{c_t}{\sigma_t}\right|, \tau_t\right\}\right)^2\left\|\frac{X_t}{\sigma_t}\right\|_{V_t^{-1}}^2 \\
&\leq \left(\min\left\{\left|\frac{\eta_t}{\sigma_t}\right|, \tau_t\right\} + \min\left\{\left|\frac{X_t^\top(\theta^* - \hat{\theta}_t)}{\sigma_t}\right|, \tau_t\right\} + \min\left\{\left|\frac{c_t}{\sigma_t}\right|, \tau_t\right\}\right)^2\left\|\frac{X_t}{\sigma_t}\right\|_{V_t^{-1}}^2 \\
&\leq 3\left(\min\left\{\left|\frac{\eta_t}{\sigma_t}\right|, \tau_t\right\}\right)^2\left\|\frac{X_t}{\sigma_t}\right\|_{V_t^{-1}}^2 + 3\left(\min\left\{\left|\frac{X_t^\top(\theta^* - \hat{\theta}_t)}{\sigma_t}\right|, \tau_t\right\}\right)^2\left\|\frac{X_t}{\sigma_t}\right\|_{V_t^{-1}}^2 \\
&\quad + 3\left(\min\left\{\left|\frac{c_t}{\sigma_t}\right|, \tau_t\right\}\right)^2\left\|\frac{X_t}{\sigma_t}\right\|_{V_t^{-1}}^2 \\
&\leq 3\left(\min\left\{\left|\frac{\eta_t}{\sigma_t}\right|, \tau_t\right\}\right)^2\left\|\frac{X_t}{\sigma_t}\right\|_{V_t^{-1}}^2 + 3\left(\frac{X_t^\top(\theta^* - \hat{\theta}_t)}{\sigma_t}\right)^2\left\|\frac{X_t}{\sigma_t}\right\|_{V_t^{-1}}^2 + 3\left(\frac{c_t}{\sigma_t}\right)^2\left\|\frac{X_t}{\sigma_t}\right\|_{V_t^{-1}}^2.
\end{aligned} \tag{28}$$

The first equality uses the definition of the gradient in (16), the second equality uses the observed reward $r_t = X_t^\top\theta^* + \eta_t + c_t$ in (2). The first inequality follows from the triangle inequality. The second uses $\min\{a+b, c\} \leq \min\{a, c\} + \min\{b, c\}$ for all $a, b, c \geq 0$. The third uses $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$ for all $a, b, c \in \mathbb{R}$, and the last uses $\min\{a, b\} \leq a$ for all $a, b \in \mathbb{R}$.

To simplify the expression of the weighted norm $\left\|\frac{X_t}{\sigma_t}\right\|_{V_t^{-1}}$, define

$$\tilde{V}_t := \lambda\alpha I_d + \sum_{s=1}^{t}\frac{X_s X_s^\top}{\sigma_s^2} = \alpha V_t. \tag{29}$$

Using $\tilde{V}_t = \alpha V_t$, w_t defined in (21) can be rewritten as

$$w_t = \frac{1}{\sqrt{\alpha}} \left\| \frac{X_t}{\sigma_t} \right\|_{V_{t-1}^{-1}} = \left\| \frac{X_t}{\sigma_t} \right\|_{\tilde{V}_{t-1}^{-1}}. \tag{30}$$

Since $\tilde{V}_t = \tilde{V}_{t-1} + X_t X_t^\top / \sigma_t^2$, applying the Sherman–Morrison–Woodbury formula [Horn and Johnson, 2012] gives

$$\tilde{V}_t^{-1} = \left(\tilde{V}_{t-1} + \frac{X_t X_t^\top}{\sigma_t^2} \right)^{-1} = \tilde{V}_{t-1}^{-1} - \tilde{V}_{t-1}^{-1} \frac{X_t}{\sigma_t} \left(1 + \frac{X_t^\top}{\sigma_t} \tilde{V}_{t-1}^{-1} \frac{X_t}{\sigma_t} \right)^{-1} \frac{X_t^\top}{\sigma_t} \tilde{V}_{t-1}^{-1} = \tilde{V}_{t-1}^{-1} - \frac{\tilde{V}_{t-1}^{-1} X_t X_t^\top \tilde{V}_{t-1}^{-1}}{\sigma_t^2 (1 + w_t^2)}.$$

Then,

$$\left\| \frac{X_t}{\sigma_t} \right\|_{\tilde{V}_t^{-1}}^2 = \frac{X_t^\top \tilde{V}_t^{-1} X_t}{\sigma_t^2} = \frac{1}{\sigma_t^2} X_t^\top \left(\tilde{V}_{t-1}^{-1} - \frac{\tilde{V}_{t-1}^{-1} X_t X_t^\top \tilde{V}_{t-1}^{-1}}{\sigma_t^2 (1 + w_t^2)} \right) X_t = w_t^2 - \frac{w_t^4}{1 + w_t^2} = \frac{w_t^2}{1 + w_t^2}.$$

Recalling that (30), we obtain

$$\left\| \frac{X_t}{\sigma_t} \right\|_{V_t^{-1}}^2 = \alpha \frac{w_t^2}{1 + w_t^2}. \tag{31}$$

Plugging (31) into (28) and summing over $s = 1, \ldots, t$, we can obtain

$$\begin{aligned} \sum_{s=1}^{t} \left\| \nabla \ell_s(\hat{\theta}_s) \right\|_{V_s^{-1}}^2 \leq & \underbrace{3\alpha \sum_{s=1}^{t} \left(\min \left\{ \left| \frac{\eta_s}{\sigma_s} \right|, \tau_s \right\} \right)^2 \frac{w_s^2}{1 + w_s^2}}_{\text{TERM (A.1)}} \\ & + \underbrace{3\alpha \sum_{s=1}^{t} \left(\frac{X_s^\top (\theta^* - \hat{\theta}_s)}{\sigma_s} \right)^2 \frac{w_s^2}{1 + w_s^2}}_{\text{TERM (A.2)}} + \underbrace{3\alpha \sum_{s=1}^{t} \left(\frac{c_s}{\sigma_s} \right)^2 \frac{w_s^2}{1 + w_s^2}}_{\text{TERM (A.3)}}. \end{aligned}$$

TERM (A.1) is the stochastic term corresponding to the Huber gradient induced by the noise η_s, TERM (A.2) is the deterministic term corresponding to the estimation error $\theta^* - \hat{\theta}_s$, and TERM (A.3) is the term induced by the corruption c_s. We now control each of the above terms separately.

Bound on TERM (A.1). This term can be bounded by the same argument as in Wang et al. [2025]. Applying Lemma D.2 under our choice of τ_t in (20) and $b := \max_{t \in [T]} \frac{\nu_t}{\sigma_t} \leq 1$, we obtain the following bound that holds with probability at least $1 - \delta$ for all $t \geq 1$:

$$\sum_{s=1}^{t} \left(\min \left\{ \left| \frac{\eta_s}{\sigma_s} \right|, \tau_s \right\} \right)^2 \frac{w_s^2}{1 + w_s^2} \leq t^{\frac{1-\epsilon}{1+\epsilon}} \left(\sqrt{\tau_0^{1-\epsilon} (2\kappa)^{\frac{1+\epsilon}{2}} (\log 3T)^{\frac{1-\epsilon}{2}}} + \tau_0 \sqrt{2 \log(2T^2/\delta)} \right)^2. \tag{32}$$

With the choice of τ_0 in (20), we observe that $\sqrt{\tau_0^{1-\epsilon} (2\kappa)^{\frac{1+\epsilon}{2}} (\log 3T)^{\frac{1-\epsilon}{2}}} = \tau_0 \sqrt{\log(2T^2/\delta)}$. Therefore, both terms inside the parentheses of (32) coincide, and hence

$$\begin{aligned} 3 \sum_{s=1}^{t} \left(\min \left\{ \left| \frac{\eta_s}{\sigma_s} \right|, \tau_s \right\} \right)^2 \frac{w_s^2}{1 + w_s^2} &\leq 3 t^{\frac{1-\epsilon}{1+\epsilon}} (1 + \sqrt{2})^2 \left(\sqrt{2\kappa} (\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}} (\log(2T^2/\delta))^{\frac{\epsilon}{1+\epsilon}} \right)^2 \\ &\leq 18 t^{\frac{1-\epsilon}{1+\epsilon}} \left(\sqrt{2\kappa} (\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}} (\log(2T^2/\delta))^{\frac{\epsilon}{1+\epsilon}} \right)^2. \end{aligned} \tag{33}$$

Bound on TERM (A.2). This term can be bounded in the same way as in Wang et al. [2025]. By the choice of σ_t in (26), we have $\sigma_t \geq 2\sqrt{2} \|X_t\|_{V_{t-1}^{-1}}$ for all t, and hence

$$\alpha w_t^2 = \left\| \frac{X_t}{\sigma_t} \right\|_{V_{t-1}^{-1}}^2 = \|X_t\|_{V_{t-1}^{-1}}^2 \frac{1}{\sigma_t^2} \leq \|X_t\|_{V_{t-1}^{-1}}^2 \frac{1}{\left(2\sqrt{2} \|X_t\|_{V_{t-1}^{-1}} \right)^2} = \frac{1}{8}. \tag{34}$$

Since $\frac{1}{1+w_s^2} \le 1$ and $\alpha w_s^2 \le \frac{1}{8}$ for all s, we obtain

$$3\alpha \sum_{s=1}^{t} \left(\frac{X_s^\top (\theta^* - \hat{\theta}_s)}{\sigma_s} \right)^2 \frac{w_s^2}{1+w_s^2} \le \frac{3}{8} \sum_{s=1}^{t} \left(\frac{X_s^\top (\theta^* - \hat{\theta}_s)}{\sigma_s} \right)^2 = \frac{3}{8} \sum_{s=1}^{t} \left\| \theta^* - \hat{\theta}_s \right\|^2_{\frac{X_s X_s^\top}{\sigma_s^2}}. \tag{35}$$

Bound on TERM (A.3). Using the definition of σ_s in (26) and the fact that $\frac{1}{1+w_s^2} \le 1$, we obtain

$$\begin{aligned} 3\alpha \sum_{s=1}^{t} \left(\frac{c_s}{\sigma_s} \right)^2 \frac{w_s^2}{1+w_s^2} &\le 3\alpha \sum_{s=1}^{t} \frac{c_s^2}{\sigma_s^2} w_s^2 \\ &= 3\alpha \sum_{s=1}^{t} c_s^2 \frac{1}{\sigma_s^4} \left(\sigma_s^2 w_s^2 \right) \\ &\le 3\alpha \sum_{s=1}^{t} c_s^2 \frac{\kappa}{C^2 \|X_s\|^2_{V_{s-1}^{-1}}} \left(\sigma_s^2 w_s^2 \right) \\ &= 3 \sum_{s=1}^{t} c_s^2 \frac{\kappa}{C^2} \frac{1}{\| \frac{X_s}{\sigma_s} \|^2_{V_{s-1}^{-1}}} w_s^2 \\ &= 3 \frac{\kappa}{C^2} \sum_{s=1}^{t} c_s^2, \end{aligned} \tag{36}$$

where the first inequality uses $1/(1+w_s^2) \le 1$, the second inequality follows from the definition of σ_s in (26), which ensures that $\sigma_s \ge \sqrt{C} \kappa^{-1/4} \|X_s\|^{1/2}_{V_{s-1}^{-1}}$, and the last equality uses $\|X_s/\sigma_s\|^2_{V_{s-1}^{-1}} = \alpha^{-1} w_s^2$, which follows from the definition of w_s in (30).

Since the corruption level C satisfies

$$C^2 = \left(\sum_{s=1}^{T} |c_s| \right)^2 \ge \left(\sum_{s=1}^{t} |c_s| \right)^2 \ge \sum_{s=1}^{t} c_s^2,$$

we conclude from (36) that

$$3\alpha \sum_{s=1}^{t} \left(\frac{c_s}{\sigma_s} \right)^2 \frac{w_s^2}{1+w_s^2} \le 3\kappa. \tag{37}$$

Combining (33), (35), and (37), we obtain that, with probability at least $1-\delta$, for all $t \ge 1$,

$$\sum_{s=1}^{t} \left\| \nabla \ell_s(\hat{\theta}_s) \right\|^2_{V_s^{-1}} \le 18\alpha \left[t^{\frac{1-\epsilon}{2(1+\epsilon)}} \sqrt{2\kappa} (\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}} \left(\log(2T^2/\delta) \right)^{\frac{\epsilon}{1+\epsilon}} \right]^2 + \frac{3}{8} \sum_{s=1}^{t} \left\| \theta^* - \hat{\theta}_s \right\|^2_{\frac{X_s X_s^\top}{\sigma_s^2}} + 3\kappa.$$

□

B.3 PROOF OF LEMMA B.3

Proof. Following Wang et al. [2025], we decompose the generalization-gap term as

$$\begin{aligned} &\left\langle \nabla \tilde{\ell}_t(\hat{\theta}_t) - \nabla \ell_t(\hat{\theta}_t), \hat{\theta}_t - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_t} \\ &= \left\langle \nabla \tilde{\ell}_t(\hat{\theta}_t) + \nabla \ell_t(\theta^*) - \nabla \ell_t(\theta^*) - \nabla \ell_t(\hat{\theta}_t), \hat{\theta}_t - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_t} \\ &= \underbrace{\left\langle \nabla \tilde{\ell}_t(\hat{\theta}_t) + \nabla \ell_t(\theta^*) - \nabla \ell_t(\hat{\theta}_t), \hat{\theta}_t - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_t}}_{\text{TERM (B.1)}} + \underbrace{\left\langle -\nabla \ell_t(\theta^*), \hat{\theta}_t - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_t}}_{\text{TERM (B.2)}}. \end{aligned}$$

Since $\nabla\tilde{\ell}_t(\theta^*) = 0$, TERM (B.1) compares the gradient increments between θ^* and $\hat{\theta}_t$, namely $\nabla\tilde{\ell}_t(\hat{\theta}_t) - \nabla\tilde{\ell}_t(\theta^*)$ and $\nabla\ell_t(\hat{\theta}_t) - \nabla\ell_t(\theta^*)$. Hence, TERM (B.1) measures how adversarial corruption and heavy-tailed noise distort these gradient increments. TERM (B.2) corresponds to the gradient evaluated at θ^*, capturing the direct effect of adversarial corruption and heavy-tailed noise. Since adversarial corruption enters both terms through the gradient $\nabla\ell_t(\cdot)$, we further decompose each term. We now analyze TERM (B.1) and TERM (B.2) separately.

Analysis of TERM (B.1). To isolate the nonlinearity of the Huber loss, we introduce its derivative $h_t(\cdot)$, defined for any $z \in \mathbb{R}$ as

$$h_t(z) = \begin{cases} z, & |z| \le \tau_t, \\ \tau_t, & z > \tau_t, \\ -\tau_t, & z < -\tau_t. \end{cases} \tag{38}$$

With this notation, the gradients of the loss (15) and the clean loss (18) can be written as

$$\nabla\ell_t(\theta) = -\frac{h_t(z_t(\theta))X_t}{\sigma_t}, \qquad \nabla\tilde{\ell}_t(\theta) = -\frac{h_t(z_t(\theta) - z_t(\theta^*))X_t}{\sigma_t}. \tag{39}$$

Substituting these expressions into TERM (B.1), we obtain

$$\begin{aligned} &\left\langle \nabla\tilde{\ell}_t(\hat{\theta}_t) + \nabla\ell_t(\theta^*) - \nabla\ell_t(\hat{\theta}_t), \hat{\theta}_t - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_t} \\ &= \left[h_t(z_t(\hat{\theta}_t)) - h_t(z_t(\theta^*)) - h_t(z_t(\hat{\theta}_t) - z_t(\theta^*))\right] \frac{\langle X_t, \hat{\theta}_t - \theta^* \rangle}{\sigma_t} \mathbb{1}_{\mathcal{A}_t}. \end{aligned} \tag{40}$$

The linear factor $\langle X_t, \hat{\theta}_t - \theta^* \rangle / \sigma_t$ can be controlled using the condition of $\mathcal{A}_t$. Hence, we focus on analyzing the inner difference term $h_t(z_t(\hat{\theta}_t)) - h_t(z_t(\theta^*)) - h_t(z_t(\hat{\theta}_t) - z_t(\theta^*))$.

Under the event $\mathcal{A}_t$, we have

$$|z_t(\hat{\theta}_t) - z_t(\theta^*)| = \left| \frac{X_t^\top(\hat{\theta}_t - \theta^*)}{\sigma_t} \right| \le \left\| \frac{X_t}{\sigma_t} \right\|_{V_{t-1}^{-1}} \left\| \hat{\theta}_t - \theta^* \right\|_{V_{t-1}} \le \frac{\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}}}{2 w_t \beta_{t-1}} \beta_{t-1} \le \frac{1}{2}\tau_0 \frac{\sqrt{1+w_t^2}}{w_t} t^{\frac{1-\epsilon}{2(1+\epsilon)}} = \frac{\tau_t}{2},$$

which follows from Cauchy–Schwarz and the definitions of σ_t and τ_t. Hence, $h_t(z_t(\hat{\theta}_t) - z_t(\theta^*)) = z_t(\hat{\theta}_t) - z_t(\theta^*)$. Since the value of

$$h_t(z_t(\hat{\theta}_t)) - h_t(z_t(\theta^*)) - (z_t(\hat{\theta}_t) - z_t(\theta^*))$$

depends on the magnitude of $z_t(\theta^*)$, we distinguish two cases.

When $|z_t(\theta^*)| \le \tau_t/2$, we have

$$|z_t(\hat{\theta}_t)| \le |z_t(\hat{\theta}_t) - z_t(\theta^*)| + |z_t(\theta^*)| \le \tau_t,$$

and also $|z_t(\hat{\theta}_t) - z_t(\theta^*)| \le \tau_t/2 < \tau_t$. Hence all the arguments of h_t lie in its linear region, i.e., $h_t(u) = u$ whenever $|u| \le \tau_t$. Therefore, the difference equals zero:

$$h_t(z_t(\hat{\theta}_t)) - h_t(z_t(\theta^*)) - h_t(z_t(\hat{\theta}_t) - z_t(\theta^*)) = 0. \tag{41}$$

In general, regardless of the values of $z_t(\theta^*)$, the uniform bound $|h_t(u)| \le \tau_t$ for all $u \in \mathbb{R}$ implies

$$\begin{aligned} &\left| h_t(z_t(\hat{\theta}_t)) - h_t(z_t(\theta^*)) - h_t(z_t(\hat{\theta}_t) - z_t(\theta^*)) \right| \\ &\le \left| h_t(z_t(\hat{\theta}_t)) \right| + |h_t(z_t(\theta^*))| + \left| h_t(z_t(\hat{\theta}_t) - z_t(\theta^*)) \right| \le 3\tau_t. \end{aligned} \tag{42}$$

Applying (41) and (42) to (40), we obtain the following bound for TERM (B.1):

$$
\begin{aligned}
&\left\langle \nabla\tilde{\ell}_t(\hat{\theta}_t) + \nabla\ell_t(\theta^*) - \nabla\ell_t(\hat{\theta}_t), \hat{\theta}_t - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_t} \\
&= \left[h_t(z_t(\hat{\theta}_t)) - h_t(z_t(\theta^*)) - h_t(z_t(\hat{\theta}_t) - z_t(\theta^*))\right] \frac{\langle X_t, \hat{\theta}_t - \theta^* \rangle}{\sigma_t} \mathbb{1}_{\mathcal{A}_t} \\
&\le \left|h_t(z_t(\hat{\theta}_t)) - h_t(z_t(\theta^*)) - h_t(z_t(\hat{\theta}_t) - z_t(\theta^*))\right| \left\| \frac{X_t}{\sigma_t} \right\|_{V_{t-1}^{-1}} \left\| \hat{\theta}_t - \theta^* \right\|_{V_{t-1}} \mathbb{1}_{\mathcal{A}_t} \\
&\le \mathbb{1}\left\{ |z_t(\theta^*)| \le \frac{\tau_t}{2} \right\} \cdot 0 + \mathbb{1}\left\{ |z_t(\theta^*)| > \frac{\tau_t}{2} \right\} 3\tau_t \left\| \frac{X_t}{\sigma_t} \right\|_{V_{t-1}^{-1}} \left\| \hat{\theta}_t - \theta^* \right\|_{V_{t-1}} \\
&= \mathbb{1}\left\{ |z_t(\theta^*)| > \frac{\tau_t}{2} \right\} 3\tau_0 \frac{\sqrt{1+w_t^2}}{w_t} t^{\frac{1-\epsilon}{2(1+\epsilon)}} \sqrt{\alpha} w_t \left\| \hat{\theta}_t - \theta^* \right\|_{V_{t-1}} \\
&\le \mathbb{1}\left\{ |z_t(\theta^*)| > \frac{\tau_t}{2} \right\} 3\tau_0 \sqrt{\alpha + \alpha w_t^2} t^{\frac{1-\epsilon}{2(1+\epsilon)}} \beta_{t-1} \\
&\le \mathbb{1}\left\{ |z_t(\theta^*)| > \frac{\tau_t}{2} \right\} 3\sqrt{\alpha + \tfrac{1}{8}} \tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} \beta_{t-1}.
\end{aligned}
$$

The first inequality follows from Cauchy–Schwarz, the second from (41) and (42), and the third from the condition of $\mathcal{A}_t$, and the last from (34).

Summing over $s = 1, \dots, t$ and using the monotonicity of $s^{\frac{1-\epsilon}{2(1+\epsilon)}}$ together with the fact that $\max_{u\in[t]} \beta_{u-1} \le \max_{u\in[t+1]} \beta_{u-1}$, we obtain

$$
\begin{aligned}
\sum_{s=1}^{t} \left\langle \nabla\tilde{\ell}_s(\hat{\theta}_s) + \nabla\ell_s(\theta^*) - \nabla\ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_s} &\le \sum_{s=1}^{t} \mathbb{1}\left\{ |z_s(\theta^*)| > \frac{\tau_s}{2} \right\} 3\sqrt{\alpha + \tfrac{1}{8}} \tau_0 s^{\frac{1-\epsilon}{2(1+\epsilon)}} \beta_{s-1} \\
&\le 3\sqrt{\alpha + \tfrac{1}{8}} \tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} \max_{u\in[t+1]} \beta_{u-1} \sum_{s=1}^{t} \mathbb{1}\left\{ |z_s(\theta^*)| > \frac{\tau_s}{2} \right\}.
\end{aligned}
$$

Since $z_s(\theta^*) = (\eta_s + c_s)/\sigma_s$ contains both the heavy-tailed noise and the corruption term, we further decompose the indicator. Using the union bound, namely that for any $x, y \in \mathbb{R}$ and $z > 0$, $\mathbb{1}\{|x+y| > z\} \le \mathbb{1}\{|x| > z/2\} + \mathbb{1}\{|y| > z/2\}$, we obtain

$$
\begin{aligned}
\sum_{s=1}^{t} \mathbb{1}\left\{ |z_s(\theta^*)| > \frac{\tau_s}{2} \right\} &= \sum_{s=1}^{t} \mathbb{1}\left\{ \left| \frac{\eta_s}{\sigma_s} + \frac{c_s}{\sigma_s} \right| > \frac{\tau_s}{2} \right\} \\
&\le \underbrace{\sum_{s=1}^{t} \mathbb{1}\left\{ \left| \frac{\eta_s}{\sigma_s} \right| > \frac{\tau_s}{4} \right\}}_{\text{TERM (B.1.1)}} + \underbrace{\sum_{s=1}^{t} \mathbb{1}\left\{ \left| \frac{c_s}{\sigma_s} \right| > \frac{\tau_s}{4} \right\}}_{\text{TERM (B.1.2)}}.
\end{aligned}
$$

TERM (B.1.1) represents the stochastic contribution from the heavy-tailed noise η_s, whereas TERM (B.1.2) represents the adversarial contribution due to the corruption term c_s.

Bound on TERM (B.1.1). By applying Lemma D.3 with the threshold $\tau_s/4$ in place of $\tau_s/2$, we obtain that, with probability at least $1-\delta$, for all $t \ge 1$,

$$
\sum_{s=1}^{t} \mathbb{1}\left\{ \left| \frac{\eta_s}{\sigma_s} \right| > \frac{\tau_s}{4} \right\} \le \frac{65}{3} \log \frac{2T^2}{\delta}. \tag{43}
$$

Bound on TERM (B.1.2). We bound the exceedance count directly:

$$\begin{aligned}
\sum_{s=1}^{t} \mathbb{1}\left\{\left|\frac{c_s}{\sigma_s}\right| > \frac{\tau_s}{4}\right\} &= \sum_{s=1}^{t} \mathbb{1}\left\{\frac{4}{\tau_s}\left|\frac{c_s}{\sigma_s}\right| > 1\right\} \\
&\le \sum_{s=1}^{t} \frac{4}{\tau_s}\frac{|c_s|}{\sigma_s} \\
&= 4\sum_{s=1}^{t} \frac{1}{\tau_0}\frac{w_s}{\sqrt{1+w_s^2}} s^{-\frac{1-\epsilon}{2(1+\epsilon)}} \frac{|c_s|}{\sigma_s^2}\sigma_s \\
&\le 4\sqrt{\kappa}\frac{1}{\sqrt{\alpha}}\frac{1}{\tau_0}\sum_{s=1}^{t} |c_s|\frac{1}{C} w_s \frac{\sqrt{\alpha}}{\|X_s/\sigma_s\|_{V_{s-1}^{-1}}} \\
&= 4\frac{1}{\sqrt{\alpha}}\frac{1}{\tau_0}\sum_{s=1}^{t} |c_s|\frac{\sqrt{\kappa}}{C} \\
&\le 4\frac{1}{\sqrt{\alpha}}\frac{\sqrt{\kappa}}{\tau_0}. \qquad (44)
\end{aligned}$$

The first inequality uses the elementary fact that $\mathbb{1}\{u > 1\} \le u$ for any $u \ge 0$. The first equality that follows comes from the definition of τ_s in (20). For the second inequality, we use $s^{-(1-\epsilon)/(2(1+\epsilon))} \le 1$ and $\frac{1}{\sqrt{1+w_s^2}} \le 1$, together with the lower bound $\sigma_s \ge \sqrt{C}\kappa^{-1/4}\|X_s\|_{V_{s-1}^{-1}}^{1/2}$ from (27).

Combining the bounds (43) and (44), we conclude that, with probability at least $1-\delta$, for all $t \in [T]$,

$$\begin{aligned}
&\sum_{s=1}^{t} \left\langle \nabla\tilde{\ell}_s(\hat{\theta}_s) + \nabla\ell_s(\theta^*) - \nabla\ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_s} \\
&\le 3\sqrt{\alpha + \tfrac{1}{8}}\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} \max_{u\in[t+1]} \beta_{u-1}\left(\frac{65}{3}\log\frac{2T^2}{\delta} + 4\frac{1}{\sqrt{\alpha}}\frac{\sqrt{\kappa}}{\tau_0}\right) \\
&= 65\sqrt{\alpha+\tfrac{1}{8}}\left(\log\frac{2T^2}{\delta}\right)\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} \max_{u\in[t+1]}\beta_{u-1} + 12\sqrt{1+\tfrac{1}{8\alpha}}\sqrt{\kappa} t^{\frac{1-\epsilon}{2(1+\epsilon)}} \max_{u\in[t+1]}\beta_{u-1}. \qquad (45)
\end{aligned}$$

Analysis of TERM (B.2). Recall from (39), the self-regularization term satisfies the following inequality:

$$\begin{aligned}
\sum_{s=1}^{t} \left\langle -\nabla\ell_s(\theta^*), \hat{\theta}_s - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_s} &= \sum_{s=1}^{t} h_s(z_s(\theta^*)) \left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_s} \\
&\le \left| \sum_{s=1}^{t} h_s(z_s(\theta^*)) \left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^* \right\rangle \right|. \qquad (46)
\end{aligned}$$

To separate the corruption term c_s from the stochastic noise η_s inside $h_s(z_s(\theta^*))$, we exploit the Lipschitz continuity of h_s. Since $h_s(\cdot)$ is 1-Lipschitz, it satisfies $|h_s(x) - h_s(y)| \le |x-y|$ for all $x, y \in \mathbb{R}$. Applying this property to $x = (\eta_s + c_s)/\sigma_s$ and $y = \eta_s/\sigma_s$ yields

$$\left| h_s\left(\frac{\eta_s + c_s}{\sigma_s}\right) - h_s\left(\frac{\eta_s}{\sigma_s}\right)\right| \le \left|\frac{c_s}{\sigma_s}\right|. \qquad (47)$$

Moreover, we decompose

$$h_s(z_s(\theta^*)) = h_s\left(\frac{\eta_s + c_s}{\sigma_s}\right) = h_s\left(\frac{\eta_s}{\sigma_s}\right) + \left\{h_s\left(\frac{\eta_s + c_s}{\sigma_s}\right) - h_s\left(\frac{\eta_s}{\sigma_s}\right)\right\}. \qquad (48)$$

Substituting (47) into (46) gives

$$\begin{aligned}
\sum_{s=1}^{t}\left\langle -\nabla\ell_s(\theta^*), \hat{\theta}_s - \theta^*\right\rangle \mathbb{1}_{\mathcal{A}_s} &= \sum_{s=1}^{t} h_s\left(\frac{\eta_s + c_s}{\sigma_s}\right)\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^*\right\rangle \mathbb{1}_{\mathcal{A}_s} \\
&= \sum_{s=1}^{t}\left\{h_s\left(\frac{\eta_s}{\sigma_s}\right) + h_s\left(\frac{\eta_s + c_s}{\sigma_s}\right) - h_s\left(\frac{\eta_s}{\sigma_s}\right)\right\}\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^*\right\rangle \mathbb{1}_{\mathcal{A}_s} \\
&\le \left|\sum_{s=1}^{t} h_s\left(\frac{\eta_s}{\sigma_s}\right)\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^*\right\rangle \mathbb{1}_{\mathcal{A}_s}\right| \\
&\quad + \sum_{s=1}^{t}\left|h_s\left(\frac{\eta_s + c_s}{\sigma_s}\right) - h_s\left(\frac{\eta_s}{\sigma_s}\right)\right|\left|\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^*\right\rangle\right| \mathbb{1}_{\mathcal{A}_s} \\
&\le \left|\sum_{s=1}^{t} h_s\left(\frac{\eta_s}{\sigma_s}\right)\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^*\right\rangle \mathbb{1}_{\mathcal{A}_s}\right| + \sum_{s=1}^{t}\left|\frac{c_s}{\sigma_s}\right|\left|\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^*\right\rangle \mathbb{1}_{\mathcal{A}_s}\right| \\
&\le \underbrace{\left|\sum_{s=1}^{t} h_s\left(\frac{\eta_s}{\sigma_s}\right)\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^*\right\rangle\right|}_{\text{TERM (B.2.1)}} + \underbrace{\sum_{s=1}^{t}\left|\frac{c_s}{\sigma_s}\right|\left|\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^*\right\rangle\right|}_{\text{TERM (B.2.2)}},
\end{aligned} \tag{49}$$

where the second equality follows from (48), the first inequality from the triangle inequality, the second inequality from (47), and third inequality from $\mathbb{1}_{\mathcal{A}_s} \le 1$. TERM (B.2.1) corresponds to the stochastic noise component, while TERM (B.2.2) captures the bias term induced by c_s. We analyze them separately.

Bound on TERM (B.2.1). This term can be handled in the same way as in Wang et al. [2025]. Applying Lemma D.1 under our choice of τ_t in (20) and $b := \max_{t\in[T]} \frac{\nu_t}{\sigma_t} \le 1$, we obtain that, with probability at least $1 - 2\delta$, the following holds for all $t \ge 1$:

$$\left\|\sum_{s=1}^{t} h_s\left(\frac{\eta_s}{\sigma_s}\right)\frac{X_s}{\sigma_s}\right\|_{\tilde{V}_t^{-1}} \le 8t^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{2\kappa}(\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}\left(\log\tfrac{2T^2}{\delta}\right)^{\frac{\epsilon}{1+\epsilon}}, \tag{50}$$

where $\tilde{V}_t$ is defined in (29). As in Wang et al. [2025], this vector bound reduces to the scalar form corresponding to $Z_s = \langle X_s, \hat{\theta}_s - \theta^*\rangle$, yielding

$$\left|\sum_{s=1}^{t} h_s\left(\frac{\eta_s}{\sigma_s}\right)\frac{Z_s}{\sigma_s}\right| \le \frac{1}{4}\left(\lambda\alpha + \sum_{s=1}^{t}\frac{Z_s^2}{\sigma_s^2}\right) + \left(8t^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{2\kappa}(\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}\left(\log\tfrac{2T^2}{\delta}\right)^{\frac{\epsilon}{1+\epsilon}}\right)^2. \tag{51}$$

Bound on TERM (B.2.2). By triangle inequality and Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}
\left|\sum_{s=1}^{t}\left|\frac{c_s}{\sigma_s}\right|\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^*\right\rangle\right| &\le \sum_{s=1}^{t}\frac{|c_s|}{\sigma_s}\left\|\frac{X_s}{\sigma_s}\right\|_{V_{s-1}^{-1}}\left\|\hat{\theta}_s - \theta^*\right\|_{V_{s-1}} \\
&\le \sum_{s=1}^{t}\frac{|c_s|}{\sigma_s^2}\|X_s\|_{V_{s-1}^{-1}}\beta_{s-1} \\
&\le \sqrt{\kappa}\sum_{s=1}^{t}\frac{|c_s|}{C}\beta_{s-1} \\
&\le \sqrt{\kappa}\max_{u\in[t+1]}\beta_{u-1},
\end{aligned} \tag{52}$$

where we used the definition of $\mathcal{A}_s$ in the second inequality, and the third inequality follows from $\sigma_s \ge \sqrt{C}\kappa^{-1/4}\|X_s\|_{V_{s-1}^{-1}}^{1/2}$ in (27), and the last follows from $\beta_{s-1} \le \max_{u\in[t+1]}\beta_{u-1}$ for all $s \le t$ and $C = \sum_{s=1}^{T}|c_s| \ge \sum_{s=1}^{t}|c_s|$.

From (51) and (52), with probability at least $1-2\delta$, the following holds for all $t \in [T]$,

$$\sum_{s=1}^{t} \left\langle -\nabla \ell_s(\theta^*), \hat{\theta}_s - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_s}$$
$$\leq \frac{1}{4}\left(\lambda\alpha + \sum_{s=1}^{t}\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^* \right\rangle^2\right) + \left(8t^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{2\kappa}(\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}\left(\log\frac{2T^2}{\delta}\right)^{\frac{\epsilon}{1+\epsilon}}\right)^2 + \sqrt{\kappa}\max_{u\in[t+1]}\beta_{u-1}. \tag{53}$$

Combining (45) and (53) and applying the union bound, we obtain that, with probability at least $1-3\delta$, for all $t \geq 1$,

$$\sum_{s=1}^{t} \left\langle \nabla \tilde{\ell}_s(\hat{\theta}_s) - \nabla \ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_s}$$
$$\leq 65\sqrt{\alpha+\tfrac{1}{8}}\left(\log\frac{2T^2}{\delta}\right)\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}}\max_{u\in[t+1]}\beta_{u-1} + 12\sqrt{1+\tfrac{1}{8\alpha}}\sqrt{\kappa}t^{\frac{1-\epsilon}{2(1+\epsilon)}}\max_{u\in[t+1]}\beta_{u-1}$$
$$+\frac{1}{4}\left(\lambda\alpha + \sum_{s=1}^{t}\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^* \right\rangle^2\right) + \left(8t^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{2\kappa}(\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}\left(\log\frac{2T^2}{\delta}\right)^{\frac{\epsilon}{1+\epsilon}}\right)^2 + \sqrt{\kappa}\max_{u\in[t+1]}\beta_{u-1},$$

which completes the proof of Lemma B.3. □

B.4 PROOF OF LEMMA 4.4

Proof. The proof proceeds in four steps.

Step 1: Collecting the two high-probability bounds. To upper bound the right-hand side of Lemma B.1, we consider an expression of a similar form and apply Lemma B.2 and B.3. Applying a union bound, we obtain that, with probability at least $1-4\delta$, for all $t \in [T]$,

$$4\lambda S^2 + \sum_{s=1}^{t}\|\nabla\ell_s(\hat{\theta}_s)\|^2_{V_s^{-1}} + 2\sum_{s=1}^{t}\left\langle \nabla\tilde{\ell}_s(\hat{\theta}_s) - \nabla\ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^*\right\rangle \mathbb{1}_{\mathcal{A}_s} + \left(\tfrac{1}{\alpha}-1\right)\sum_{s=1}^{t}\left\|\hat{\theta}_s - \theta^*\right\|^2_{\frac{X_sX_s^\top}{\sigma_s^2}}$$
$$\leq 4\lambda S^2 + 18\alpha\left[t^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{2\kappa}(\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}\left(\log\frac{2T^2}{\delta}\right)^{\frac{\epsilon}{1+\epsilon}}\right]^2 + \frac{3}{8}\sum_{s=1}^{t}\left\|\theta^* - \hat{\theta}_s\right\|^2_{\frac{X_sX_s^\top}{\sigma_s^2}} + 3\kappa$$
$$+ 130\sqrt{\alpha+\tfrac{1}{8}}\left(\log\tfrac{2T^2}{\delta}\right)\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}}\max_{u\in[t+1]}\beta_{u-1} + 24\sqrt{1+\tfrac{1}{8\alpha}}\sqrt{\kappa}t^{\frac{1-\epsilon}{2(1+\epsilon)}}\max_{u\in[t+1]}\beta_{u-1}$$
$$+\frac{1}{2}\left(\lambda\alpha + \sum_{s=1}^{t}\left\langle \frac{X_s}{\sigma_s}, \hat{\theta}_s - \theta^*\right\rangle^2\right) + 2\left(8t^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{2\kappa}(\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}\left(\log\frac{2T^2}{\delta}\right)^{\frac{\epsilon}{1+\epsilon}}\right)^2$$
$$+ 2\sqrt{\kappa}\max_{u\in[t+1]}\beta_{u-1} + \left(\tfrac{1}{\alpha}-1\right)\sum_{s=1}^{t}\left\|\hat{\theta}_s - \theta^*\right\|^2_{\frac{X_sX_s^\top}{\sigma_s^2}}.$$

Step 2: Simplifying the coefficients. For $t \in [T]$, our parameter choices ensure that

$$\sqrt{2}t^{\frac{1-\epsilon}{2(1+\epsilon)}}(\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}\left(\log\tfrac{2T^2}{\delta}\right)^{\frac{\epsilon}{1+\epsilon}} \geq 1.$$

Hence, we multiply the terms of 3κ or $2\sqrt{\kappa}\max_{u\in[t+1]}\beta_{u-1}$ by this factor to match the scale of the other terms. Recall that

$$(\log\tfrac{2T^2}{\delta})\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} = t^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{2\kappa}(\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}(\log\tfrac{2T^2}{\delta})^{\frac{\epsilon}{1+\epsilon}},$$

so that all terms on the right-hand side of the above form can be rewritten in terms of $(\log\frac{2T^2}{\delta})\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}}$.

Substituting $\alpha = 8$ into the bound obtained in Step 1 and collecting terms of the same form, we obtain that, with probability at least $1 - 4\delta$, for all $t \in [T]$,

$$\begin{aligned}
&4\lambda S^2 + \sum_{s=1}^{t} \|\nabla \ell_s(\hat{\theta}_s)\|^2_{V_s^{-1}} + 2\sum_{s=1}^{t} \left\langle \nabla \tilde{\ell}_s(\hat{\theta}_s) - \nabla \ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_s} \\
&\quad + \left(\tfrac{1}{\alpha} - 1\right) \sum_{s=1}^{t} \left\| \hat{\theta}_s - \theta^* \right\|^2_{\frac{X_s X_s^\top}{\sigma_s^2}} \\
&\le (4S^2 + 2)\lambda + (144 + 128 + 3)\Big((\log \tfrac{2T^2}{\delta})\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}}\Big)^2 \\
&\quad + \Big(130\sqrt{8 + \tfrac{1}{8}} + 24\sqrt{1 + \tfrac{1}{64}} + 2\Big)(\log \tfrac{2T^2}{\delta})\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} \max_{u \in [t+1]} \beta_{u-1} \\
&\le (2 + 4S^2)\lambda + 275\Big((\log \tfrac{2T^2}{\delta})\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}}\Big)^2 \\
&\quad + 397(\log \tfrac{2T^2}{\delta})\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} \max_{u \in [t+1]} \beta_{u-1}.
\end{aligned}$$

Here, the term $\sum_{s=1}^{t} \|\hat{\theta}_s - \theta^*\|_{\frac{X_s X_s^\top}{\sigma_s}}$ cancels out because of the choice $\alpha = 8$.

To match the standard confidence bound form $\|\hat{\theta}_t - \theta^*\|_{V_t} \le \beta_t$, it suffices to choose β_t so that, for all $t \in [T]$,

$$\beta_t^2 \ge (2 + 4S^2)\lambda + 275\Big((\log \tfrac{2T^2}{\delta})\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}}\Big)^2 + 397(\log \tfrac{2T^2}{\delta})\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} \max_{u \in [t+1]} \beta_{u-1}.$$

To simplify the notation, define

$$b = 397 t^{\frac{1-\epsilon}{2(1+\epsilon)}}, c = (4S^2 + 2)\lambda + 275\Big((\log \tfrac{2T^2}{\delta})\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}}\Big)^2.$$

Then the above condition can be written as

$$b \max_{u \in [t+1]} \beta_{u-1} + c \le \beta_t^2.$$

Step 3: Choosing the confidence radius. By choosing the sequence $\{\beta_t\}_{t=1}^{T}$ to be non-decreasing, we have $\max_{u \le t+1} \beta_{u-1} = \beta_t$. Therefore, it suffices to require

$$b\beta_t + c \le \beta_t^2 \qquad \text{for all } t \in [T].$$

The quadratic inequality holds whenever

$$\beta_t \ge \frac{b}{2} + \sqrt{\Big(\frac{b}{2}\Big)^2 + c}.$$

Using the inequality $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ for $a, b > 0$, we obtain the simpler sufficient condition

$$\beta_t \ge b + \sqrt{c}.$$

Recalling the explicit form of b and c, and slightly enlarging the numerical constant for simplicity, we define

$$\beta_t := 409(\log \tfrac{2T^2}{\delta})\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} + \sqrt{(2 + 4S^2)\lambda}.$$

With this choice, with probability at least $1 - 4\delta$, the following inequality holds for all $t \in [T]$:

$$\beta_t^2 \ge 4\lambda S^2 + \sum_{s=1}^{t} \|\nabla \ell_s(\hat{\theta}_s)\|^2_{V_s^{-1}} + 2\sum_{s=1}^{t} \left\langle \nabla \tilde{\ell}_s(\hat{\theta}_s) - \nabla \ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^* \right\rangle \mathbb{1}_{\mathcal{A}_s} + \left(\tfrac{1}{\alpha} - 1\right) \sum_{s=1}^{t} \left\| \hat{\theta}_s - \theta^* \right\|^2_{\frac{X_s X_s^\top}{\sigma_s^2}}. \quad (54)$$

Step 4: Verifying the confidence radius via induction. Let $\mathcal{B}$ denote the event that the conditions in (54) hold for all $t \geq 1$. By construction, $\mathbb{P}(\mathcal{B}) \geq 1 - 4\delta$. We define the event

$$\mathcal{C} := \bigcap_{t \geq 1} \Big\{ \|\hat{\theta}_t - \theta^*\|_{V_{t-1}} \leq \beta_{t-1} \Big\} = \bigcap_{t \geq 1} \mathcal{A}_t.$$

We show that $\mathcal{B} \subseteq \mathcal{C}$ by mathematical induction, which immediately implies $\mathbb{P}(\mathcal{C}) \geq \mathbb{P}(\mathcal{B}) \geq 1 - 4\delta$. By definition, $\mathcal{A}_1$ holds since $\|\hat{\theta}_1 - \theta^*\|_{V_0} \leq \sqrt{4\lambda S^2} \leq \sqrt{\lambda(2 + 4S^2)} = \beta_0$.

Assume that at iteration $t \geq 1$, $\mathcal{A}_s$ holds for all $s \in [t]$. We now show that $\mathcal{A}_{t+1}$ also holds. By Lemma B.1, we have

$$\begin{aligned}
&\|\hat{\theta}_{t+1} - \theta^*\|_{V_t}^2 \\
&\leq 4\lambda S^2 + \sum_{s=1}^{t} \|\nabla \ell_s(\hat{\theta}_s)\|_{V_s^{-1}}^2 + 2\sum_{s=1}^{t} \big\langle \nabla \tilde{\ell}_s(\hat{\theta}_s) - \nabla \ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^* \big\rangle + \Big(\frac{1}{\alpha} - 1\Big) \sum_{s=1}^{t} \|\hat{\theta}_s - \theta^*\|^2_{\frac{X_s X_s^\top}{\sigma_s^2}} \\
&= 4\lambda S^2 + \sum_{s=1}^{t} \|\nabla \ell_s(\hat{\theta}_s)\|_{V_s^{-1}}^2 + 2\sum_{s=1}^{t} \big\langle \nabla \tilde{\ell}_s(\hat{\theta}_s) - \nabla \ell_s(\hat{\theta}_s), \hat{\theta}_s - \theta^* \big\rangle \mathbf{1}_{\mathcal{A}_s} + \Big(\frac{1}{\alpha} - 1\Big) \sum_{s=1}^{t} \|\hat{\theta}_s - \theta^*\|^2_{\frac{X_s X_s^\top}{\sigma_s^2}} \\
&\leq \beta_t^2.
\end{aligned}$$

Here, the first inequality follows from Lemma B.1, the first equality uses the induction hypothesis that $\mathcal{A}_s$ holds for all $s \in [t]$, and the second inequality follows from (54). Consequently, $\mathcal{A}_{t+1}$ holds. Since $t \geq 1$ was arbitrary, we conclude that $\mathcal{A}_t$ holds for all $t \geq 1$, which implies that $\mathbb{P}(\mathcal{C}) \geq 1 - 4\delta$.

Moreover, we can verify that

$$\sqrt{\frac{2\beta_{t-1}}{\tau_0 \sqrt{\alpha} t^{\frac{1-\epsilon}{2(1+\epsilon)}}}} \|X_t\|_{V_{t-1}^{-1}} = \sqrt{\frac{409 \left(\log \frac{2T^2}{\delta}\right) \tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} + \sqrt{\lambda(2 + 4S^2)}}{\sqrt{2} \tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}}}} \|X_t\|_{V_{t-1}^{-1}} \geq \sqrt{\frac{409}{\sqrt{2}}} \|X_t\|_{V_{t-1}^{-1}} \geq 2\sqrt{2} \|X_t\|_{V_{t-1}^{-1}}.$$

Hence, our choice of σ_t satisfies the lower bound condition required in Lemma B.2.

Consequently, under the choice of σ_t and τ_t,

$$\sigma_t := \max \left\{ \nu_t, \frac{1}{\sqrt{T}}, \sqrt{\frac{2\beta_{t-1}}{\tau_0 \sqrt{\alpha} t^{\frac{1-\epsilon}{2(1+\epsilon)}}}} \|X_t\|_{V_{t-1}^{-1}}, \sqrt{C} \kappa^{-1/4} \|X_t\|_{V_{t-1}^{-1}}^{1/2} \right\}, \qquad \tau_t = \tau_0 \frac{\sqrt{1 + w_t^2}}{w_t} t^{\frac{1-\epsilon}{2(1+\epsilon)}},$$

we obtain that for any $\delta \in (0, \frac{1}{4})$, with probability at least $1 - 4\delta$, the following holds for all $t \geq 1$:

$$\|\hat{\theta}_{t+1} - \theta_*\|_{V_t} \leq 409 \log \frac{2T^2}{\delta} \tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} + \sqrt{\lambda(2 + 4S^2)}.$$

□

C REGRET ANALYSIS

C.1 PROOF OF THEOREM 4.5

Proof. We define $X_t^* = \arg\max_{\mathbf{x} \in \mathcal{X}_t} \mathbf{x}^\top \theta^*$. By Lemma 4.4 and the fact that $X_t^*, X_t \in \mathcal{X}_t$, we have, with probability at least $1 - 4\delta$, for all $t \in [T]$,

$$\langle X_t^*, \theta^* - \hat{\theta}_t \rangle \leq \|\theta^* - \hat{\theta}_t\|_{V_{t-1}} \|X_t^*\|_{V_{t-1}^{-1}} \leq \beta_{t-1} \|X_t^*\|_{V_{t-1}^{-1}}, \tag{55}$$

which can be rewritten as

$$X_t^{*\top} \theta^* \leq X_t^{*\top} \hat{\theta}_t + \beta_{t-1} \|X_t^*\|_{V_{t-1}^{-1}}. \tag{56}$$

Similarly, by Lemma 4.4, with probability at least $1 - 4\delta$, for all $t \in [T]$,

$$\langle X_t, \theta^* - \hat{\theta}_t \rangle \geq -\beta_{t-1} \|X_t\|_{V_{t-1}^{-1}}, \tag{57}$$

which implies

$$X_t^\top \theta^* \geq X_t^\top \hat{\theta}_t - \beta_{t-1}\|X_t\|_{V_{t-1}^{-1}}. \tag{58}$$

By applying the union bound to (56) and (58), we obtain that, with probability at least $1 - 8\delta$, the following holds for all $t \in [T]$:

$$\begin{aligned} X_t^{*\top}\theta^* - X_t^\top \theta^* &\leq \big(X_t^{*\top}\hat{\theta}_t + \beta_{t-1}\|X_t^*\|_{V_{t-1}^{-1}}\big) - \big(X_t^\top \hat{\theta}_t - \beta_{t-1}\|X_t\|_{V_{t-1}^{-1}}\big) \\ &= X_t^{*\top}\hat{\theta}_t - X_t^\top \hat{\theta}_t + \beta_{t-1}\big(\|X_t^*\|_{V_{t-1}^{-1}} + \|X_t\|_{V_{t-1}^{-1}}\big). \end{aligned} \tag{59}$$

On the other hand, by the arm selection rule (12),

$$X_t^{*\top}\hat{\theta}_t + \beta_{t-1}\|X_t^*\|_{V_{t-1}^{-1}} \leq X_t^\top \hat{\theta}_t + \beta_{t-1}\|X_t\|_{V_{t-1}^{-1}},$$

which implies

$$X_t^{*\top}\hat{\theta}_t - X_t^\top \hat{\theta}_t \leq \beta_{t-1}\big(\|X_t\|_{V_{t-1}^{-1}} - \|X_t^*\|_{V_{t-1}^{-1}}\big). \tag{60}$$

Substituting (60) into (59) yields

$$X_t^{*\top}\theta^* - X_t^\top \theta^* \leq \beta_{t-1}\big(\|X_t\|_{V_{t-1}^{-1}} - \|X_t^*\|_{V_{t-1}^{-1}}\big) + \beta_{t-1}\big(\|X_t^*\|_{V_{t-1}^{-1}} + \|X_t\|_{V_{t-1}^{-1}}\big) = 2\beta_{t-1}\|X_t\|_{V_{t-1}^{-1}}.$$

Therefore, the regret satisfies

$$\mathrm{Reg}(T) = \sum_{t=1}^{T}\big(X_t^{*\top}\theta^* - X_t^\top \theta^*\big) \leq 2\sum_{t=1}^{T}\beta_{t-1}\|X_t\|_{V_{t-1}^{-1}} \leq 2\beta_T \sum_{t=1}^{T}\|X_t\|_{V_{t-1}^{-1}} = 2\sqrt{\alpha}\beta_T \sum_{t=1}^{T}\sigma_t w_t,$$

where $\beta_t = 409\big(\log\frac{2T^2}{\delta}\big)\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} + \sqrt{(4S^2+2)\lambda}$.

Next we bound the sum of bonuses $\sum_{t=1}^{T}\sigma_t w_t$ by splitting the time indices $[T]$ into three cases depending on which the candidate term attains the value of σ_t. If the maximum is attained by multiple candidate terms, we assign the index t to an arbitrary one of them. We define

$$\begin{aligned} \mathcal{J}_1 &:= \Big\{t \in [T] : \sigma_t \in \{\nu_t, \sigma_{\min}\}\Big\}, \\ \mathcal{J}_2 &:= \left\{t \in [T] : \sigma_t = \sqrt{\frac{2\beta_{t-1}}{\tau_0\sqrt{\alpha}t^{\frac{1-\epsilon}{2(1+\epsilon)}}}}\|X_t\|_{V_{t-1}^{-1}}\right\}, \\ \mathcal{J}_3 &:= \Big\{t \in [T] : \sigma_t = \sqrt{C}\kappa^{-1/4}\|X_t\|_{V_{t-1}^{-1}}^{1/2}\Big\}. \end{aligned}$$

We first bound the contribution from the indices in $\mathcal{J}_1$. Recalling that $\sigma_{\min} = 1/\sqrt{T}$, by the Cauchy–Schwarz inequality and Lemma D.4 we obtain

$$\begin{aligned} \sum_{t\in\mathcal{J}_1}\sigma_t w_t = \sum_{t\in\mathcal{J}_1}\max\{\nu_t, \sigma_{\min}\}w_t &\leq \sum_{t=1}^{T}\max\{\nu_t, \sigma_{\min}\}w_t \\ &\leq \sqrt{\sum_{t=1}^{T}(\nu_t^2 + \sigma_{\min}^2)}\sqrt{\sum_{t=1}^{T}w_t^2} \leq \sqrt{2\kappa}\sqrt{1 + \sum_{t=1}^{T}\nu_t^2}, \end{aligned} \tag{61}$$

where in the last inequality we used $\sum_{t=1}^{T}\sigma_{\min}^2 = 1$ and Lemma D.4 to bound $\sum_{t=1}^{T}w_t^2 \leq 2\kappa$.

For the indices in $\mathcal{J}_2$, we use the fact that Lemma 4.4 holds, and hence the confidence radius β_t takes the form

$$\beta_t = 409\left(\log\frac{2T^2}{\delta}\right)\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} + \sqrt{\lambda(2+4S^2)}.$$

Therefore,

$$w_t^{-2} = \frac{2\beta_{t-1}}{\tau_0\sqrt{\alpha}t^{\frac{1-\epsilon}{2(1+\epsilon)}}} = \frac{2\left(409\left(\log\frac{2T^2}{\delta}\right)\tau_0 t^{\frac{1-\epsilon}{2(1+\epsilon)}} + \sqrt{\lambda(2+4S^2)}\right)}{\tau_0\sqrt{\alpha}t^{\frac{1-\epsilon}{2(1+\epsilon)}}} \leq \frac{\sqrt{\lambda(2+4S^2)}}{\tau_0} + 409\log\frac{2T^2}{\delta}, \quad (62)$$

and letting

$$c_0 := \frac{\sqrt{\lambda(2+4S^2)}}{\tau_0} + 409\log\frac{2T^2}{\delta},$$

we obtain

$$\sum_{t\in\mathcal{J}_2}\sigma_t w_t = \sum_{t\in\mathcal{J}_2}\frac{1}{w_t^2}\|X_t\|_{\tilde{V}_{t-1}^{-1}} w_t^2 \leq \frac{c_0 L}{\sqrt{\alpha\lambda}}\sum_{t\in\mathcal{J}_2} w_t^2 \leq \frac{2c_0\kappa L}{\sqrt{\alpha\lambda}}, \quad (63)$$

where the first inequality follows from $\tilde{V}_{t-1} \succeq \lambda\alpha I_d$, which implies $\|X_t\|_{\tilde{V}_{t-1}^{-1}} \leq L/\sqrt{\alpha\lambda}$, and the second inequality follows from Lemma D.4, which gives $\sum_{t=1}^T w_t^2 \leq 2\kappa$.

Finally, we bound the contribution from the indices in $\mathcal{J}_3$:

$$\sum_{t\in\mathcal{J}_3}\sigma_t w_t = \sum_{t\in\mathcal{J}_3}\frac{\sqrt{\alpha}C}{\sqrt{\kappa}} w_t^2 \leq 2\sqrt{\alpha\kappa}C, \quad (64)$$

where the inequality follows from Lemma D.4.

Choosing $\delta = 1/(8T)$ and $\lambda = d$, we obtain $\kappa = \tilde{\mathcal{O}}(d)$ and $\beta_T = \tilde{\mathcal{O}}\left(\sqrt{d}T^{\frac{1-\epsilon}{2(1+\epsilon)}}\right)$. Plugging these orders into the bounds (61), (63), and (64), and summing them up, with probability at least $1 - \frac{1}{T}$, we have

$$\begin{aligned}
\mathrm{Reg}(T) &\leq 2\sqrt{\alpha}\beta_T\left(\sqrt{2\kappa}\sqrt{1+\sum_{t=1}^T\nu_t^2} + \frac{2c_0\kappa L}{\sqrt{\alpha\lambda}} + \sqrt{\alpha\kappa}C\right) \\
&= \tilde{\mathcal{O}}\left(\sqrt{d}T^{\frac{1-\epsilon}{2(1+\epsilon)}}\left(\sqrt{d\sum_{t=1}^T\nu_t^2} + \sqrt{d}\right) + \sqrt{d}C\right) \\
&= \tilde{\mathcal{O}}\left(dT^{\frac{1-\epsilon}{2(1+\epsilon)}}\sqrt{\sum_{t=1}^T\nu_t^2} + dT^{\frac{1-\epsilon}{2(1+\epsilon)}} + dT^{\frac{1-\epsilon}{2(1+\epsilon)}}C\right).
\end{aligned}$$

□

C.2 PROOF OF COROLLARY 4.7

Proof. The argument follows the same structure as in the proof of Theorem 4.5. The only difference is that we replace C with $\bar{C}$ in the definition of $\mathcal{J}_3$, i.e.,

$$\mathcal{J}_3 := \{t \in [T] : \sigma_t = \sqrt{\bar{C}}\kappa^{-1/4}\|X_t\|_{V_{t-1}^{-1}}^{1/2}\}.$$

With this definition of $\mathcal{J}_3$, we obtain the bound on the contribution of indices in $\mathcal{J}_3$:

$$\sum_{t\in\mathcal{J}_3}\sigma_t w_t \leq 2\sqrt{\alpha\kappa}\bar{C}.$$

The bounds for $\mathcal{J}_1$ and $\mathcal{J}_2$ remain identical to those in Theorem 4.5. Under the same choice of δ and λ as in Theorem 4.5,

with probability at least $1 - \frac{1}{T}$, we have

$$\begin{aligned}\mathrm{Reg}(T) &\le 2\sqrt{\alpha}\beta_T \left(\sqrt{2\kappa}\sqrt{\sum_{t=1}^{T} \nu_t^2 + 1} + \frac{2c_0 L\kappa}{\sqrt{\alpha\lambda}} + 2\sqrt{\alpha\kappa}\bar{C} \right) \\ &= \tilde{\mathcal{O}}\left(dT^{\frac{1-\varepsilon}{2(1+\varepsilon)}} \sqrt{\sum_{t=1}^{T} \nu_t^2} + dT^{\frac{1-\varepsilon}{2(1+\varepsilon)}} + dT^{\frac{1-\varepsilon}{2(1+\varepsilon)}} \bar{C} \right).\end{aligned}$$

□

C.3 PROOF OF COROLLARY 4.8

Proof. The argument follows the same structure as in the proof of Theorem 4.5. The only difference is that ν_t is replaced by ν in the definition of $\mathcal{J}_1$, namely,

$$\mathcal{J}_1 := \{t \in [T] : \sigma_t \in \{\nu, \sigma_{\min}\}\}.$$

With this definition of $\mathcal{J}_1$, we obtain the bound on the contribution of indices in $\mathcal{J}_1$:

$$\begin{aligned}\sum_{t\in\mathcal{J}_1} \sigma_t w_t &= \sum_{t\in\mathcal{J}_1} \max\{\nu, \sigma_{\min}\} w_t \\ &\le \sqrt{\sum_{t\in\mathcal{J}_1} (\nu^2 + \sigma_{\min}^2)} \sqrt{\sum_{t\in\mathcal{J}_1} w_t^2} \le \sqrt{2\kappa}\nu\sqrt{T+1}.\end{aligned}$$

The bounds for $\mathcal{J}_2$ and $\mathcal{J}_3$ remain identical to those in Theorem 4.5. Under the same choice of δ and λ as in Theorem 4.5, with probability at least $1 - \frac{1}{T}$, we have

$$\begin{aligned}\mathrm{Reg}(T) &\le 2\sqrt{\alpha}\beta_T \sum_{t=1}^{T} \sigma_t w_t \\ &\le 4\beta_T \left(\sqrt{2\kappa}\nu\sqrt{T+1} + \frac{2c_0 L\kappa}{\sqrt{\alpha\lambda}} + 2\sqrt{\alpha\kappa}C \right) \\ &= \tilde{\mathcal{O}}\left(dT^{\frac{1}{1+\varepsilon}} + dT^{\frac{1-\varepsilon}{2(1+\varepsilon)}} + dT^{\frac{1-\varepsilon}{2(1+\varepsilon)}} C \right).\end{aligned}$$

□

C.4 DISCUSSION ON THE DEPENDENCE ON CORRUPTION LEVEL

For adversarially corrupted linear bandits with sub-Gaussian noise, Bogunovic et al. [2021] showed that the linear dependence dC on the corruption level is optimal. Our regret bound (Theorem 4.5) recovers the same dependence in the finite-variance case ($\epsilon = 1$) under a constant moment bound. For $\epsilon < 1$, however, our regret bound contains an additional T-dependent factor in the corruption-dependent term. This should not be interpreted as evidence that a stronger dependence on the corruption level is fundamentally necessary under bounded $(1+\epsilon)$-moment noise, since this noise class includes sub-Gaussian noise as a special case. Rather, the additional factor arises from our current analysis. In particular, as shown in Appendix C.1, the corruption-dependent term in the regret decomposition is multiplied by the confidence radius β_t. Under bounded $(1+\epsilon)$-moment noise, the confidence radius scales as $t^{\frac{1-\epsilon}{2(1+\epsilon)}}$ and therefore depends explicitly on t. Consequently, this time dependence propagates to the corruption-dependent term, leading to the additional T-dependent factor in our regret bound for $\epsilon < 1$.

D AUXILIARY LEMMAS

We collect several auxiliary lemmas adapted from Huang et al. [2023] and Wang et al. [2025]. For clarity, let $\mathcal{F}_t = \sigma(X_{1:t}, \eta_{1:t-1}, c_{1:t-1})$ denote the filtration associated with our process. Compared to the filtrations used in the above works, ours additionally includes the corruption sequence $c_{1:t-1}$. This difference is immaterial, since the lemmas in Huang et al. [2023] and Wang et al. [2025] only require a conditional moment bound with respect to the underlying filtration.

Lemma D.1 (Partial result of Lemma C.2 in Huang et al. [2023])**.** *Let $\{\mathcal{F}_t\}_{t=1}^{\infty}$ be a filtration and $\{\eta_t\}_{t=1}^{\infty}$ be a sequence of random variables such that for all $t \geq 1$,*

$$\mathbb{E}\left[\frac{\eta_t}{\sigma_t}\middle|\mathcal{F}_t\right] = 0, \qquad \mathbb{E}\left[\left|\frac{\eta_t}{\sigma_t}\right|^{1+\epsilon}\middle|\mathcal{F}_t\right] \leq b^{1+\epsilon}.$$

For each $t \in \mathbb{N}$, let X_t be an $\mathbb{R}^d$-valued random variable satisfying $\|X_t\|_2 \leq L$. Define $U_0 = \lambda I_d$ and $U_t = U_{t-1} + \frac{X_t X_t^\top}{\sigma_t^2}$ for $t \geq 1$, where $\lambda > 0$. Moreover, for $t \geq 1$ set

$$\tau_t = \tau_0 \frac{\sqrt{1+w_t^2}}{w_t} t^{\frac{1-\epsilon}{2(1+\epsilon)}}, \qquad \tau_0 = \frac{\sqrt{2\kappa} b (\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}}{\left(\log \frac{2T^2}{\delta}\right)^{\frac{1}{1+\epsilon}}}.$$

Then, with probability at least $1-2\delta$, for all $t \geq 1$,

$$\left\|\sum_{s=1}^{t} f'_{\tau_s}\left(\frac{\eta_s}{\sigma_s}\right)\frac{X_s}{\sigma_s}\right\|_{U_t^{-1}} \leq 8 t^{\frac{1-\epsilon}{2(1+\epsilon)}} \sqrt{2\kappa} (\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}} \left(\log \frac{2T^2}{\delta}\right)^{\frac{\epsilon}{1+\epsilon}}.$$

Lemma D.2 (Lemma C.5 in Huang et al. [2023])**.** *Let $\{\mathcal{F}_t\}_{t=1}^{\infty}$ be a filtration and $\{\eta_t\}_{t=1}^{\infty}$ be a sequence of random variables such that for all $t \geq 1$,*

$$\mathbb{E}\left[\frac{\eta_t}{\sigma_t}\middle|\mathcal{F}_t\right] = 0, \qquad \mathbb{E}\left[\left|\frac{\eta_t}{\sigma_t}\right|^{1+\epsilon}\middle|\mathcal{F}_t\right] \leq b^{1+\epsilon}.$$

Moreover, for $t \geq 1$ set

$$\tau_t = \tau_0 \frac{\sqrt{1+w_t^2}}{w_t} t^{\frac{1-\epsilon}{2(1+\epsilon)}}, \qquad \tau_0 = \frac{\sqrt{2\kappa} b (\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}}{\left(\log(2T^2/\delta)\right)^{\frac{1}{1+\epsilon}}}.$$

Then, with probability at least $1-\delta$, for all $t \geq 1$,

$$\sum_{s=1}^{t} {f'_{\tau_s}}^2\left(\frac{\eta_s}{\sigma_s}\right)\frac{w_s^2}{1+w_s^2} \leq t^{\frac{1-\epsilon}{1+\epsilon}} \left(\sqrt{\tau_0^{1-\epsilon}(\sqrt{2\kappa}b)^{1+\epsilon}(\log 3t)^{\frac{1-\epsilon}{2}}} + \tau_0\sqrt{2\log\frac{2t^2}{\delta}}\right)^2.$$

Lemma D.3 (Eq. (C.12) in Huang et al. [2023])**.** *Let $\{\mathcal{F}_t\}_{t=1}^{\infty}$ be a filtration and $\{\eta_t\}_{t=1}^{\infty}$ be a sequence of random variables such that for all $t \geq 1$,*

$$\mathbb{E}\left[\frac{\eta_t}{\sigma_t}\middle|\mathcal{F}_t\right] = 0, \qquad \mathbb{E}\left[\left|\frac{\eta_t}{\sigma_t}\right|^{1+\epsilon}\middle|\mathcal{F}_t\right] \leq b^{1+\epsilon}.$$

Moreover, for $t \geq 1$ set

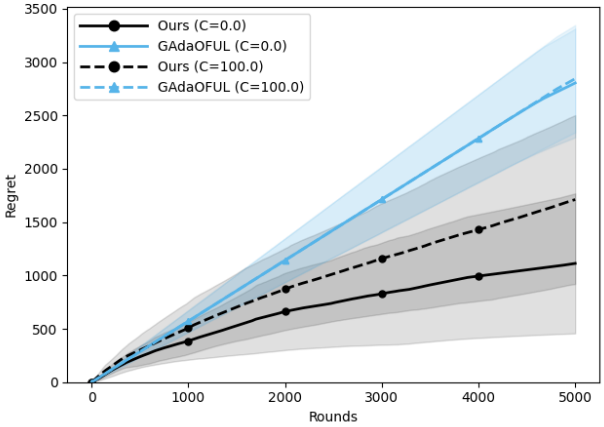
$$\tau_t = \tau_0 \frac{\sqrt{1+w_t^2}}{w_t} t^{\frac{1-\epsilon}{2(1+\epsilon)}}, \qquad \tau_0 = \frac{\sqrt{2\kappa} b (\log 3T)^{\frac{1-\epsilon}{2(1+\epsilon)}}}{\left(\log(2T^2/\delta)\right)^{\frac{1}{1+\epsilon}}}.$$

Then, with probability at least $1-\delta$, for all $t \geq 1$,

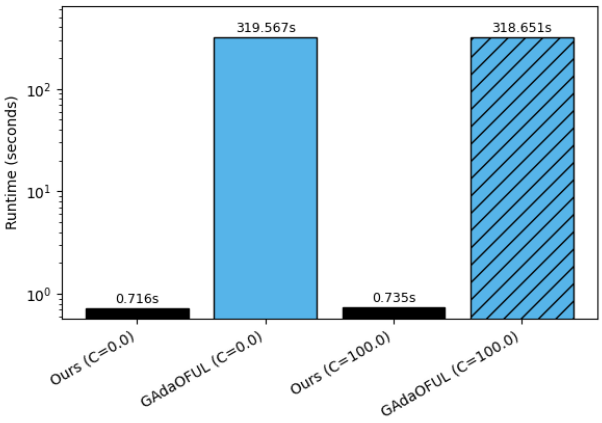
$$\sum_{s=1}^{t} \mathbb{1}\left\{\left|\frac{\eta_s}{\sigma_s}\right| > \frac{\tau_s}{2}\right\} \leq \frac{23}{3}\log\frac{2T^2}{\delta}.$$

Lemma D.4 (Elliptical potential lemma, Lemma 7 in Wang et al. [2025])**.** *For each $t \in \mathbb{N}$, let X_t be an $\mathbb{R}^d$-valued random variable satisfying $\|X_t\|_2 \leq L$. Let $U_0 = \lambda I_d$ and $U_t = U_{t-1} + X_t X_t^\top$ for $t \geq 1$, where $\lambda > 0$. Assume that for all $t \geq 1$, $\|U_{t-1}^{-1/2} X_t\|_2^2 \leq c_{\max}$. Then,*

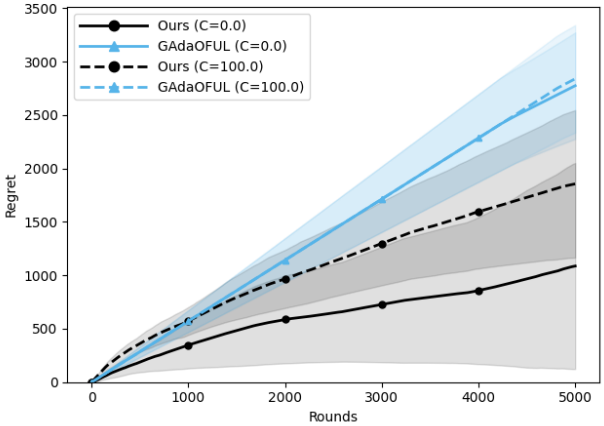
$$\sum_{t=1}^{T} \|U_{t-1}^{-1/2} X_t\|_2^2 \leq 2\max\{1, c_{\max}\} d \log\left(1 + \frac{L^2 T}{\lambda d}\right).$$



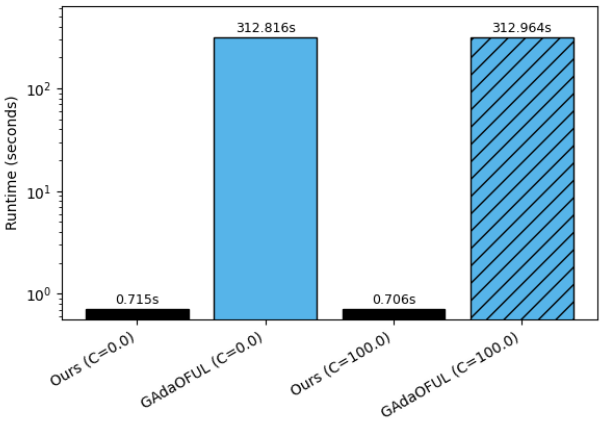
(a) Regret (Fisk distribution)



(b) Runtime (Fisk distribution)



(c) Regret (Lomax distribution)



(d) Runtime (Lomax distribution)

Figure 2: Comparison of Ours and GAdaOFUL

E ADDITIONAL EXPERIMENTS

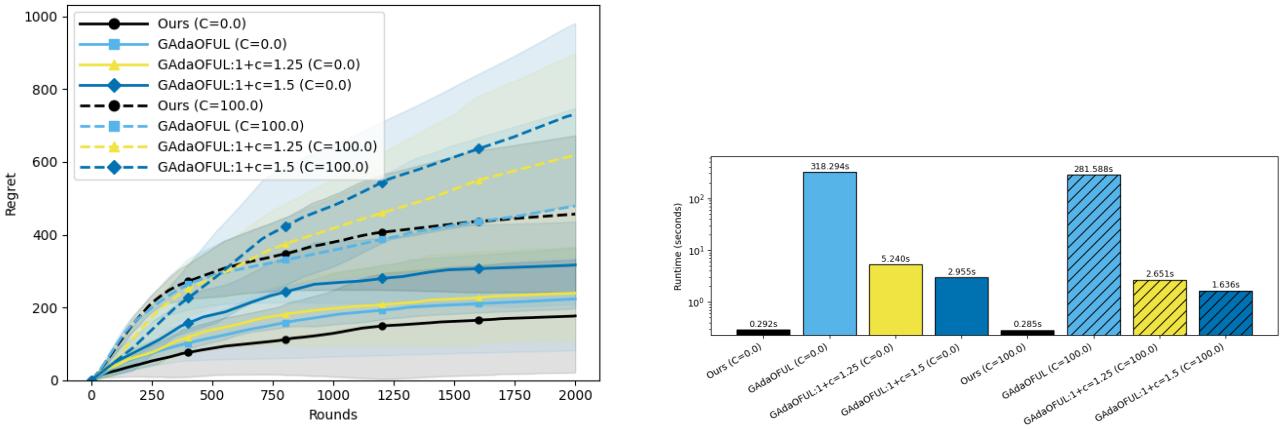
E.1 ADDITIONAL HEAVY-TAILED NOISE MODELS

We further examine whether the empirical robustness of our algorithm persists under different heavy-tailed noise distributions. To this end, we conduct additional experiments with Fisk and Lomax noise distributions, both with shape parameter 1.5, as representative heavy-tailed distributions. Except for the heavy-tailed noise distribution, we use the same experimental setup as in Section 6, including the problem instance, corruption levels, horizon, and number of independent runs. We compare our method with GAdaOFUL.

Figure 2 shows the results for the Fisk and Lomax noise distributions. Across both noise distributions, our algorithm exhibits sub-linear regret behavior under both the corruption-free case ($C = 0$) and the corrupted case ($C = 100$). In contrast, GAdaOFUL shows an approximately linear growth of regret in both cases, indicating that it is less robust to these heavy-tailed noise distributions. The runtime results further highlight the computational advantage of our method: our algorithm requires less than one second per run, whereas GAdaOFUL takes more than 300 seconds per run. Thus, our method is several hundred times faster than GAdaOFUL while maintaining robust regret performance under heavy-tailed noise and adversarial corruption.

E.2 COMPARISON WITH RARELY SWITCHING BASELINES

Abbasi-Yadkori et al. [2011] proposed the rarely switching OFUL algorithm, which avoids recomputing the estimator at



(a) Regret ($\epsilon = 1$)

(b) Runtime ($\epsilon = 1$)

Figure 3: Comparison of Ours, GAdaOFUL, and rarely switching variants of GAdaOFUL

every round. Instead, it recomputes the estimator only when the determinant of the regularized design matrix V_t increases by a factor of $1 + c$ for some constant $c \geq 0$. This reduces the total number of estimator recomputations up to time T to essentially $\mathcal{O}(\log T)$. The switching parameter c controls the trade-off between computational efficiency and the theoretical regret guarantee: larger values of c reduce the recomputation frequency but generally lead to larger regret. Motivated by this idea, we apply the rarely switching strategy to the optimization-based GAdaOFUL to investigate how much its total computational cost can be reduced in practice. We then compare this rarely switching variant with Ours in terms of both cumulative regret and running time.

We use the same setting as the finite-variance experiment in Section 6, except that the time horizon is set to $T = 2000$. Each experiment is averaged over 10 independent runs. As in the previous experiments, each line represents the mean, and the shaded region indicates one standard deviation from the mean.

Figure 3a shows the cumulative regret. As observed in Section 6, Ours and the original GAdaOFUL, which corresponds to the case $c = 0$, achieve comparable regret in both the corruption-free case ($C = 0$) and the corrupted case ($C = 100$). When $C = 0$, increasing the switching parameter from $c = 0.25$ to $c = 0.5$ gradually worsens the regret performance. This degradation becomes more pronounced under adversarial corruption: when $C = 100$, both rarely switching variants suffer substantially larger regret than the original GAdaOFUL, and the variant with $c = 0.5$ performs worse than that with $c = 0.25$. These results may suggest that the performance degradation caused by rarely switching is more pronounced under adversarial corruption.

Figure 3b shows the running time. As expected, increasing c reduces the running time because the number of estimator recomputations decreases. However, even with rarely switching updates, GAdaOFUL remains slower than Ours. These results suggest that the rarely switching strategy reduces the running time of GAdaOFUL, but appears to degrade its regret performance as the switching parameter increases, especially under adversarial corruption.