High-Dimensional Sparse Linear Contextual Bandits with Heavy-tailed Rewards

Truong Vu

November 1, 2025

Abstract

We study high-dimensional linear contextual bandits under heavy-tailed reward distributions, where both robustness and sparsity are essential for reliable decision-making. In many real-world settings, such as personalized recommendations or adaptive clinical trials, the contextual dimension is large, yet only a small subset of features influences the expected reward, and the observed feedback may be contaminated by outliers or exhibit heavy tails. Classical algorithms relying on least-squares estimation are fragile under such conditions and fail to provide meaningful guarantees.

To address these challenges, we propose a unified framework for robust sparse contextual bandits based on Huber and Huber-Lasso estimators. Our approach integrates robust loss functions with ℓ_1-regularization and adaptive exploration to handle both model sparsity and non-Gaussian noise. A key technical contribution is the development of non-asymptotic deviation inequalities for Huber-type estimators when data are collected adaptively—beyond the traditional i.i.d. setting of robust regression. These results establish restricted strong convexity and concentration properties under adaptive sampling, enabling finite-sample analysis of robust estimators in bandit environments.

Building on these results, we design Huber Bandit and Huber-Lasso Bandit algorithms with provable high-probability regret bounds. Under standard sparsity and restricted eigenvalue conditions, the proposed methods achieve near-optimal regret growth rates while requiring only finite $(1+\delta)$-th reward moments. Our theory extends sparse linear bandit analyses to heavy-tailed regimes and demonstrates that robustness and sparsity can be effectively combined for stable and efficient learning in high-dimensional, noisy environments.

1 Introduction

Sequential decision-making under uncertainty is a central problem in modern data-driven systems. The *contextual bandit* framework provides a principled approach to learning personalized actions—such as recommending a product, prescribing a treatment, or allocating online resources—by sequentially observing contextual information and optimizing cumulative rewards. In the *linear contextual bandit model*, the expected reward of each arm is modeled as a linear function of the observed features, and the decision maker learns the unknown parameters through adaptive exploration.

While extensively studied in the literature, standard linear bandit algorithms typically rely on two critical assumptions: (i) a moderate feature dimension relative to the sample size, and (ii) light-tailed (sub-Gaussian) reward noise. These assumptions, however, rarely hold in practice. In high-dimensional environments—such as recommendation systems with hundreds of user and item attributes, or clinical decision-making with numerous patient biomarkers—the number of contextual features can vastly exceed the available samples. Moreover, real-world reward signals often exhibit *heavy tails* or outliers due to unmodeled user behavior, measurement errors, or sporadic shocks. In such settings, conventional least-squares-based methods (e.g., LinUCB or Lasso-bandit) can fail dramatically, leading to unreliable parameter estimates and unstable learning performance.

This paper develops a unified framework for **robust sparse contextual bandits** that simultaneously addresses *high-dimensional sparsity* and *heavy-tailed noise*. We consider a sparse linear contextual bandit model where each arm's parameter vector is high-dimensional but only a small subset of features contributes to the reward. To handle heavy-tailed or contaminated reward distributions, we employ **Huber-type robust estimators** that balance sensitivity and resistance to outliers. Specifically, we propose two algorithms:

(i) **Huber Bandit**, based on the classical Huber loss; and

(ii) **Huber-Lasso Bandit**, which integrates ℓ_1-regularization to promote sparsity in high-dimensional settings.

These algorithms combine ideas from robust statistics and online decision-making, bridging modern high-dimensional regression theory with adaptive learning. A key technical challenge is that the data collected by a bandit algorithm are *adaptively dependent*, violating the independence conditions assumed in most robust estimation analyses. To overcome this, we establish **non-asymptotic deviation inequalities** for the Huber and Huber-Lasso estimators under adaptive sampling, extending the results of **?** beyond the i.i.d. setting. Our analysis leverages restricted eigenvalue conditions, localized restricted strong convexity, and concentration arguments adapted to martingale sequences.

Building on these theoretical tools, we derive **high-probability bounds** on both estimation error and cumulative regret. In particular, we show that the proposed Huber and Huber-Lasso bandit algorithms achieve polylogarithmic dependence on the time horizon and maintain robustness against heavy-tailed rewards with only finite $(1+\delta)$-th moments. The results match or improve existing sparse linear bandit guarantees (e.g., Bastani and Bayati (2020)) under substantially weaker noise assumptions, demonstrating that robustness and sparsity can be jointly exploited for stable and efficient online learning.

From a broader perspective, our framework offers a statistically principled and computationally tractable approach for decision-making in noisy, high-dimensional environments. It applies naturally to applications such as personalized recommendations, adaptive experiments, and risk-aware resource allocation—domains where high-dimensional features and non-Gaussian noise are intrinsic. Beyond contextual bandits, our techniques contribute to the emerging intersection of robust high-dimensional inference and adaptive learning algorithms, and open new avenues for the design of provably reliable machine learning systems operating under distributional uncertainty.

Contributions and Organization. Our main contributions are threefold:

(a) **Robust high-dimensional estimation under adaptive sampling.** We establish new non-asymptotic deviation inequalities for the Huber and Huber-Lasso estimators when the feature vectors are collected adaptively. These results extend robust regression theory from the i.i.d. setting to the contextual bandit framework, where dependence across samples arises naturally from sequential decision-making.

(b) **Design of adaptive Huber and Huber-Lasso bandit algorithms.** We develop two robust bandit algorithms that integrate forced exploration with adaptive estimation using Huber-type losses. The algorithms are provably stable in the presence of heavy-tailed noise and efficiently recover sparse parameter structures in high-dimensional contexts.

(c) **Finite-sample regret guarantees.** We provide high-probability bounds on both estimation error and cumulative regret. Under standard sparsity and restricted eigenvalue conditions, our algorithms achieve regret bounds of order $O(\log T)$ for forced-sampling steps and $O(T^{1/2+\epsilon})$ for the overall process, depending only polynomially on the sparsity level and logarithmically on the ambient dimension. These guarantees match the best-known rates for sparse linear bandits under light-tailed noise while relaxing moment conditions to merely finite $(1+\delta)$-th moments.

The remainder of this paper is organized as follows. Section 2 introduces the problem formulation, model assumptions, and tail inequalities for the Huber and Huber-Lasso estimators with adaptively collected data. Section 2.3 presents the proposed algorithms and Sec 3 establishes the corresponding regret bounds. Appendices provide auxiliary lemmas and detailed proofs of the main results.

2 Problem Formulation

2.1 Sparse linear contextual bandit and heavy-tailed reward

Let T denote the unknown number of time steps (time horizon), d the feature dimension, and K the number of arms (decisions). Suppose that each arm $a \in [K]$ has an unknown parameter $\beta_a^* \in \mathbb{R}^d$. At each time step $t \in [T]$, the decision maker (DM) observes a d-dimensional context/feature vector X_t for the newly arrived user, pulls an arm (chooses an action) $a \in [K]$, which reveals a stochastic reward

$$Y_{a,t} = X_t^{\mathrm{T}} \beta_a^* + \varepsilon_{a,t},$$

where $\varepsilon_{a,t}$ are independent zero-mean random variables that are also independent of the context sequence $\{X_{t'}\}_{t' \geq 1}$. We follow the model setup considered in Bastani and Bayati (2020) and assume that the sequence of contexts $\{X_t\}_{t \geq 1}$ consists of independent random vectors drawn from some unknown distribution $\mathcal{P}_X$ on the feature space $\mathcal{X} \subseteq \mathbb{R}^d$.

For contextual bandit problems, the primary objective is to design a sequential decision-making policy, denoted as $\pi = (\pi_1, \pi_2, \ldots)$, which learns the underlying arm parameters $\{\beta_a^*\}_{a=1}^K$ over time with the aim of maximizing expected reward for each arm. At time $t \in [T]$, such a policy π chooses the arm $\pi_t \in [K]$, which is then compared to the optimal arm $\pi_t^* := \max_{1 \le a \le K} X_t^{\mathrm{T}} \beta_a^*$, selected by an oracle π^* that knows $\{\beta_a^*\}_{a=1}^K$. The *regret* incurred by π at time t is defined as

$$r_t = X_t^{\mathrm{T}} \beta_{\pi_t^*}^* - X_t^{\mathrm{T}} \beta_{\pi_t}^*.$$

The DM aims to seek a policy π that minimizes the cumulative regret $R_T = \sum_{t=1}^T r_t$ over T rounds. In this work, we consider the linear contextual bandits problem within the high-dimensional sparse regime, while addressing the additional challenge posed by heavy-tailed reward distributions.

Assumption 2.1 (Parameter Set and Stochastic Contexts)**.** The arm parameters $\{\beta_a^*\}_{a=1}^K$ satisfy (i) $\max_{1 \le a \le K} \|\beta_a^*\|_0 \le s_0 \le \min(T, d)$, and (ii) $\max_{1 \le a \le K} \|\beta_a^*\|_2 \le b_2$. The stochastic contexts $X_t \in \mathbb{R}^d$ are such that $\sup_{t \ge 1} \|X_t\|_\infty \le x_{\max}$ almost surely for some positive constant $x_{\max} \ge 1$.

Assumption 2.2 (Heavy-tailed Reward)**.** The noise variables $\varepsilon_{a,t}$ are independent and satisfy $\mathbb{E}(\varepsilon_{a,t} | X_t) = 0$, $\mathbb{E}(|\varepsilon_{a,t}|^{1+\delta} | X_t) \le \nu_\delta < \infty$ for some $\delta \in (0, 1]$.

The quantity b_2 in Assumption 2.1 can be viewed as dimension-free. To see this, assume that the noise variables $\varepsilon_{a,t}$ have finite variance. Moreover, suppose the feature vector X_t is *isotropic*, meaning $\mathbb{E}(X_t X_t^{\mathrm{T}}) = \mathrm{I}_d$. Then, the second moment of the stochastic reward $Y_{a,t}$ is $\mathbb{E}(Y_{a,t}^2) = \|\beta_a^*\|_2^2 + \mathrm{var}(\varepsilon_{a,t})$, implying that $\|\beta_a^*\|_2^2 \le \mathbb{E}(Y_{a,t}^2)$. On the other hand, since β_a^* is s_0-sparse, the worst-case bound for the ℓ_1-norm is $\|\beta_a^*\|_1 \le s_0^{1/2} \|\beta_a^*\|_2 = \mathcal{O}(s_0^{1/2})$.

Definition 2.1 (Restricted Eigenvalue (RE) condition)**.** Let $\mathcal{I} \subseteq [d]$ be a subset of indices. Given some positive constant $\phi > 0$, we say that a $d \times d$ positive semidefinite (PSD) matrix A satisfies the $\mathrm{RE}(\mathcal{I}, \phi)$ condition with respect to $\mathcal{I}$ if

$$\delta^{\mathrm{T}} A \delta \ge \phi \|\delta\|_2^2 \quad \text{for all } \delta \in \mathbb{R}^d \text{ s.t. } \|\delta_{\mathcal{I}^c}\|_1 \le 3 \|\delta_{\mathcal{I}}\|_1.$$

Let $\mathfrak{C}(\mathcal{I}, \phi)$ denote the collection of all PSD matrices that satisfy the $\mathrm{RE}(\mathcal{I}, \phi)$ condition.

Assumption 2.3 (Arm optimality)**.** We assume the the arm set can be divided into two sets, optimal arms and sub-optimal arms: $[K] = K_{\mathrm{opt}} \cup K_{\mathrm{sub}}$ where $K_{\mathrm{opt}} \cap K_{\mathrm{sub}} = \emptyset$, such that there exists $h > 0$ satisfying

(i) For every $i \in K_{\mathrm{sub}}$ we have $X^{\mathrm{T}} \beta_i < \max_{j \ne i} X^{\mathrm{T}} \beta_i - h$.

(ii) For each i in K_{opt}, we define

$$U_i = \{X \in \mathbb{R}^d \,|\, X^{\mathrm{T}} \beta_i > \max_{j \ne i} X^{\mathrm{T}} \beta_i + h\}, \tag{2.1}$$

then there exists $p_* > 0$ such that $\min_{i \in K_{\mathrm{opt}}} \mathbb{P}(X \in U_i) \ge p_*$.

Assumption 2.4. We assume that the expected Gram matrix of contexts in U_i is positive definite, that is, for each i in K_{opt} as in Assumption 2.3, there exists $\gamma > 0$ such that $\lambda_{\min}(\mathbb{E}[X X^{\mathrm{T}} | X \in U_i]) \ge \gamma$.

2.2 Tail inequalities for Huber and Huber-Lasso estimators with adaptively collected data

In this section, we establish general non-asymptotic deviation bounds for the Huber-Lasso and Huber estimators with adaptively collected data. We assume that the observed sample $\{(Y_t, X_t)\}_{t=1}^n$ follows a linear model $Y_t = X_t^{\mathrm{T}}\beta^* + \varepsilon_t$, where $\beta^* \in \mathbb{R}^p$ is the unknown parameter of interest. The noise variables ε_t are independent and satisfy $\mathbb{E}(\varepsilon_t|X_t) = 0$, $\mathbb{E}(|\varepsilon_t|^{1+\delta}|X_t) \le \nu_\delta$ for some $\delta \in (0,1]$. The feature sequence $\{X_t\}_{t=1}^n$ is collected adaptively, meaning that at time t, X_t may depend on the past features and the resulting responses $\{(X_{t'}, Y_{t'})\}_{t'=1}^{t-1}$.

Given a *robustification* parameter $\tau > 0$ and a *regularization* parameter $\lambda > 0$, the *Huber* and *Huber-Lasso* estimators are defined as

$$\widehat{\beta}_n(\tau) = \operatorname*{argmin}_{\beta \in \mathbb{R}^d} \underbrace{\frac{1}{n}\sum_{t=1}^n \ell_\tau(Y_t - X_t^{\mathrm{T}}\beta)}_{=:\, L_{n,\tau}(\beta)} \quad \text{and} \quad \widehat{\beta}_n(\tau,\lambda) \in \operatorname*{argmin}_{\beta\in\mathbb{R}^d} \big\{L_{n,\tau}(\beta) + \lambda\|\beta\|_1\big\},$$

respectively, where

$$\ell_\tau(u) = (u^2/2)\mathbb{1}(|u| \le \tau) + (\tau|u| - \tau^2/2)\mathbb{1}(|u| > \tau),\ u \in \mathbb{R}$$

denotes the Huber loss (Huber, 1973), which is continuously differentiable with the derivative $\ell_\tau'(u) = \mathrm{sign}(u)\min\{|u|, \tau\}$.

Let $\mathcal{S} = \mathrm{supp}(\beta^*) \subseteq [d]$ be the support of β^* satisfying $|\mathcal{S}| \le s_0 \le d$. Given $r, \phi > 0$, define the event

$$\Omega_n(\phi, r) = \left\{\langle \Delta_n(\delta), \delta\rangle \ge \phi\|\delta\|_2^2 \ \ \forall \delta \in \mathbb{R}^d \text{ s.t. } \|\delta_{\mathcal{S}^c}\|_1 \le 3\|\delta_{\mathcal{S}}\|_1,\ \|\delta\|_2 \le r\right\}, \tag{2.2}$$

where $\Delta_n(\delta) := \nabla L_{n,\tau}(\beta^* + \delta) - \nabla L_{n,\tau}(\beta^*)$ for $\delta \in \mathbb{R}^d$. In the low-dimensional setting where $\mathcal{S} = [d]$ and $s_0 = d$, the event $\Omega_n(\phi, r)$ is reduced to $\Omega_n(\phi, r) = \{\langle \Delta_n(\delta), \delta\rangle \ge \phi\|\delta\|_2^2 \ \ \forall \delta \in \mathbb{R}^d \text{ s.t. } \|\delta\|_2 \le r\}$.

Proposition 2.1. Assume we observe a sample $\{(Y_t, X_t)\}_{t=1}^n$ of size n from a linear model, with adaptively collected random feature vectors $X_t \in \mathbb{R}^d$, independent heavy-tailed errors ε_t with bounded $(1+\delta)$-th absolute moments for some $\delta \in (0,1]$, and the true parameter $\beta^* \in \mathbb{R}^d$ being s_0-sparse. Moreover, assume $\|X_t\|_\infty \le x_{\max}$ (almost surely) and that ε_t is dependent of $\mathcal{F}_{t-1}$, where $\mathcal{F}_t = \sigma(\{(X_{t'}, Y_{t'})\}_{t'=1}^t)$. For any $z > 0$, set $\tau = \tau_0(n/z)^{1/(1+\delta)}$ with $\tau_0 \ge \nu_\delta^{1/(1+\delta)}$.

(i) (Huber-Lasso) Assume $\lambda \ge 4x_{\max}\tau_0(z/n)^{\delta/(1+\delta)}$. Then, for any $\phi > 0$ and $r \ge 1.5 s_0^{1/2}\lambda/\phi$, the corresponding Huber-Lasso estimator satisfies, with probability at least $1 - 2e^{-z+\log d} - \mathbb{P}\{\Omega_n(\phi,r)^{\mathrm{c}}\}$ that

$$\|\widehat{\beta}_n(\tau,\lambda) - \beta^*\|_2 \le 3\sqrt{s_0}\lambda/(2\phi) \quad \text{and} \quad \|\widehat{\beta}_n(\tau,\lambda) - \beta^*\|_1 \le 6s_0\lambda/\phi.$$

(ii) (Huber) For any $\phi > 0$ and $r \ge 2\phi^{-1}x_{\max}\tau_0 d^{1/2}(z/n)^{\delta/(1+\delta)}$, the Huber estimator $\widehat{\beta}_n(\tau)$ satisfies, with probability at least $1 - 2e^{-z+\log d} - \mathbb{P}\{\Omega_n(\phi,r)^{\mathrm{c}}\}$ that

$$\|\widehat{\beta}_n(\tau) - \beta^*\|_2 \le 2\frac{x_{\max}d^{1/2}\tau_0}{\phi}\left(\frac{z}{n}\right)^{\delta/(1+\delta)} \le r.$$

2.3 Adaptive Huber/Huber-Lasso bandit algorithms

We prescribe a set of times when we forced sample arm i:

$$\mathcal{T}_i := \{(2^n - 1)Kq + j \mid n \in \{0, 1, 2, \dots\} \tag{2.3}$$

$$\text{and} \quad j \in \{q(i-1)+1, q(i-1)+2, \dots, qi\}, q \in \mathbb{N}\}. \tag{2.4}$$

Let $\mathcal{T}_{i,t} := \mathcal{T}_i \cap [t]$ be the set of forced arm i up to time t. The maximum regret at forced sampling steps is $O(K \log T)$ because $|\mathcal{T}_{i,t}| = O(\log T)$. We denote $\mathcal{S}_{i,t} := \{r \mid a(r) = i,\, r \leq t\}$ be the set of time steps where arm i is chosen until time t, either forcedly or adaptively. Observe that $\mathcal{T}_{i,t} \subset \mathcal{S}_{i,t}$.

Algorithm 1 Huber bandit

Input: q, h, ν_δ, α.

Initialize $\mathcal{T}_{i,t}$ and $\mathcal{S}_{i,t}$ by the empty set, and $\widehat{\beta}_{\mathcal{T}_{i,0}}(\tau, \lambda)$ and $\widehat{\beta}_{\mathcal{S}_{i,0}}(\tau, \lambda)$ by 0 in $\mathbb{R}^d$ for all $i \in [K]$.

for $t \in [T]$ **do**

  Observe $X_t \sim P_X$

  if $t \in T_i$ **then**

    $a(t) = i$

    else

      $\mathcal{D} = \{i \in [K] \mid \max_{j \in [K]} X_t^{\mathrm{T}} \widehat{\beta}_{\mathcal{T}_{j,t-1}}(\tau, \lambda) - X_t^{\mathrm{T}} \widehat{\beta}_{\mathcal{T}_{i,t-1}}(\tau, \lambda) \leq \frac{h}{2}\}$

      $a(t) = \mathrm{argmax}_{i \in \mathcal{D}} X_t^{\mathrm{T}} \widehat{\beta}_{\mathcal{S}_{i,t-1}}(\tau, \lambda)$

    end

  end

  Update $\mathcal{S}_{a(t),t} = \mathcal{S}_{a(t),t-1} \cup \{t\}$

  Observe reward $y_t = X_t^{\mathrm{T}} \beta_{a(t)} + \varepsilon_{a(t),t}$

  if $t \in T_i$ **then**

    $\tau(\mathcal{T}_{i,t}) = \nu_\delta \Big(\frac{|\mathcal{T}_{i,t}|}{\log(\frac{t^2(2d+1)}{\alpha})}\Big)^{\frac{1}{1+\delta}}$

    $\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau, \lambda) = \mathrm{argmin}_{\beta \in \mathbb{R}^d} \frac{1}{|\mathcal{T}_{i,t}|} \sum_{r \in \mathcal{T}_{i,t}} \ell_\tau(y_r - X_r^{\mathrm{T}} \beta)$

    else

      $\tau(\mathcal{S}_{i,t}) = \nu_\delta \Big(\frac{|\mathcal{S}_{i,t}|}{\log(\frac{t^2(2d+1)}{\alpha})}\Big)^{\frac{1}{1+\delta}}$

      $\widehat{\beta}_{\mathcal{S}_{i,t}}(\tau, \lambda) = \mathrm{argmin}_{\beta \in \mathbb{R}^d} \frac{1}{|\mathcal{S}_{i,t}|} \sum_{r \in \mathcal{S}_{i,t}} \ell_\tau(y_r - X_r^{\mathrm{T}} \beta)$

    end

  end

end

---

3 Regret Analysis

For any subset $\mathcal{A} \subseteq [n]$, we define the Huber-Lasso and Huber estimators trained on samples indexed by $\mathcal{A}$ as

$$\widehat{\beta}_{\mathcal{A}}(\tau, \lambda) \in \operatorname*{argmin}_{\beta \in \mathbb{R}^d} \big\{ L_{\mathcal{A},\tau}(\beta) + \lambda \|\beta\|_1 \big\} \quad \text{and} \quad \widehat{\beta}_{\mathcal{A}}(\tau) \in \operatorname*{argmin}_{\beta \in \mathbb{R}^d} L_{\mathcal{A},\tau}(\beta), \tag{3.1}$$

where $L_{\mathcal{A},\tau}(\beta) = |\mathcal{A}|^{-1}\sum_{t\in\mathcal{A}}\ell_\tau(Y_t - X_t^{\mathrm{T}}\beta)$, and $Y_t = X_t^{\mathrm{T}}\beta^* + \varepsilon_t$. As before, we assume β^* is s_0-sparse for some $1 \le s_0 \le d$, and ε_t's are independent zero-mean random errors with bounded $(1+\delta)$-th moments, that is, $\mathbb{E}(|\varepsilon_t|^{1+\delta}|X_t) \le \nu_\delta$ for some $\delta \in (0,1]$.

Since the feature vectors X_t's are adaptively collected, the samples X_t, $t \in \mathcal{A}$ are not necessarily i.i.d. Instead, we assume the existence of an unknown subset $\mathcal{A}' \subseteq \mathcal{A}$, comparable in size to $\mathcal{A}$, which comprises i.i.d. samples $\{X_t, t \in \mathcal{A}'\}$ from some sub-exponential distribution $\mathbb{P}_X$ satisfying Assumption 2.1. Letting $\Sigma = \mathbb{E}_{X\sim\mathbb{P}_X}(XX^{\mathrm{T}}) \in \mathbb{R}^{d\times d}$, we further assume that $\Sigma \in \mathfrak{C}(\mathcal{S},\phi)$ for some positive constant ϕ, where $\mathcal{S} = \mathrm{supp}(\beta^*)$ and $\mathfrak{C}(\mathcal{S},\phi)$ is given in Definition 2.1.

Proposition 3.1. Assume $|\mathcal{A}'|/|\mathcal{A}| \ge p/2$ for a positive constant p. For any $\alpha \in (0,1)$ and $z \ge 1$, set $\tau = \tau_0(|\mathcal{A}|/z)^{1/(1+\delta)}$ with $\tau_0 \ge \nu_\delta^{1/(1+\delta)}$.

(i) (Huber-Lasso) Assume $|\mathcal{A}| \ge C_1 p^{-1}\max\{(x_{\max}^2 s_0/\phi)^2\log(2d/\alpha), (x_{\max}^2 s_0/\phi)z\}$ for a sufficiently large absolute constant $C_1 > 1$, and let λ satisfy

$$4x_{\max}\tau_0\left(\frac{z}{|\mathcal{A}|}\right)^{\delta/(1+\delta)} \le \lambda \le \frac{p\phi}{60x_{\max}s_0}\tau_0\left(\frac{|\mathcal{A}|}{\tau}\right)^{1/(1+\delta)}. \tag{3.2}$$

Then, with probability at least $1-\alpha-2e^{-z+\log d}$,

$$\|\widehat{\beta}_{\mathcal{A}}(\tau,\lambda) - \beta^*\|_2 \le 12\frac{\sqrt{s_0}\lambda}{p\phi} \quad \text{and} \quad \|\widehat{\beta}_{\mathcal{A}}(\tau,\lambda) - \beta^*\|_1 \le 48\frac{s_0\lambda}{p\phi}.$$

(ii) (Huber) Assume $|\mathcal{A}| \ge C_2 p^{-1}\max\{(x_{\max}^2 d/\phi)^2\log(e/\alpha), (x_{\max}^2 d/\phi)z\}$ for a sufficiently large absolute constant $C_2 > 1$. Then, for any

$$\frac{16x_{\max}d^{1/2}\tau_0}{p\phi}\left(\frac{z}{|\mathcal{A}|}\right)^{\delta/(1+\delta)} \le r \le \frac{\tau_0}{1.5x_{\max}d^{1/2}}\left(\frac{|\mathcal{A}|}{\tau}\right)^{1/(1+\delta)}, \tag{3.3}$$

the Huber estimator $\widehat{\beta}_{\mathcal{A}}(\tau)$ satisfies $\mathbb{P}\{\|\widehat{\beta}_{\mathcal{A}}(\tau) - \beta^*\|_2 \ge r\} \le \alpha + 2e^{-z+\log d}$.

Proposition 3.2. (i) (Huber-Lasso) For all $i \in [K]$ and $z \ge 1$, set $\tau = \tau_0(|\mathcal{T}_{i,t}|/z)^{1/(1+\delta)}$ with $\tau_0 \ge \nu_\delta^{1/(1+\delta)}$, $\lambda = \frac{hp\phi}{192x_{\max}s_0}$ satisfying

$$4x_{\max}\tau_0\left(\frac{z}{|\mathcal{T}_{i,t}|}\right)^{\delta/(1+\delta)} \le \lambda \le \frac{p\phi}{60x_{\max}s_0}\tau_0\left(\frac{|\mathcal{T}_{i,t}|}{\tau}\right)^{1/(1+\delta)}, \tag{3.4}$$

$t \ge (Kq)^2$, $q \ge C_1 p^{-1}\max\{(x_{\max}^2 s_0/\phi)^2\log(2d/\alpha), (x_{\max}^2 s_0/\phi)z\}$, then the forced sample estimator $\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau,\lambda)$ satisfying

$$\Pr[\|\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau,\lambda) - \beta_i\|_1 > \frac{h}{4x_{\max}}] \le \frac{5}{t^4}. \tag{3.5}$$

(ii) (Huber) For $\alpha \in (0,1)$, $i \in [K]$ and $t \ge (Kq)^2$, we have

$$\mathbb{P}\{\|\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau) - \beta_i\|_2 \ge \frac{h}{4x_{\max}}\} \le \frac{5}{t^4}. \tag{3.6}$$

Now, we give a tail inequality for the optimal arms K_{opt} all-sample estimator. The problem lies in the fact that the online decisions made by the algorithm determine the all-sample sets $\mathcal{S}_{i,t}$. More specifically, the algorithm uses X_t and prior observations $\{X_{t'}\}_{1\leq t'\leq t-1}$ (which are utilized to estimate β_i) to select arm i at time t. Consequently, there might be a correlation between the variables $\{X_t | t\in\mathcal{S}_{i,t}\}$.Furthermore, we do not have a guarantee that a constant fraction of the all-sample sets $\mathcal{S}_{i,t}$ are i.i.d., in contrast to the forced-sample estimator.Specifically, we will show that $|\mathcal{S}_{i,t}| = \mathcal{O}(T)$ for optimal arms $i \in K_{opt}$ with high probability, but only the $|\mathcal{T}_{i,t}| = \mathcal{O}(\log T)$ forced samples are guaranteed to be i.i.d. Therefore, we are unable to use $\mathcal{A} = \mathcal{S}_{i,t}$ and $\mathcal{A}' = \mathcal{T}'_{i,t}$ as before in Proposition 3.1. In order to address this, we first demonstrate that our algorithm utilizes the forced-sample estimator $\mathcal{O}(T)$ times with a high probability, and then a constant fraction of the samples where we use the forced-sample estimator are i.i.d. from the regions U_i. We then invoke Proposition 3.1 with a modified $\mathcal{A}'$ such that $|\mathcal{A}'| = \mathcal{O}(T)$. In particular, we define the event

$$A_t := \{\|\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau,\lambda) - \beta_i\|_1 \leq \frac{h}{4x_{\max}}, \forall i \in [K]\}. \tag{3.7}$$

$$B_t := \{\|\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau) - \beta_i\|_2 \leq \frac{h}{4x_{\max}}, \forall i \in [K]\}. \tag{3.8}$$

Since the event A_t depends only on forced-samples, the random variables $\{X_t \,|\, A_t \text{ occurs}\}$ are i.i.d. with distribution $\mathcal{P}_X$. Furthermore, if we let

$$\mathcal{S}'_{i,t} := \{t' \in [t] \,|\, A_{t'-1} \text{ holds} \quad X_{t'} \in U_i, \text{ and } t' \notin \cup_{j\in[K]}\mathcal{T}_{j,t}\}, \tag{3.9}$$

$$\mathcal{S}''_{i,t} := \{t' \in [t] \,|\, B_{t'-1} \text{ holds} \quad X_{t'} \in U_i, \text{ and } t' \notin \cup_{j\in[K]}\mathcal{T}_{j,t}\}, \tag{3.10}$$

then the random variables $\{X_{t'} \,|\, t' \in \mathcal{S}'_{i,t}\}$ are i.i.d. with distribution $\mathcal{P}_{X|X\in U_i}$ because the event $\{X_{t'} \,|\, t' \in U_i\}$ is independent of the event $A_{t'-1}$ while $\{t' \notin \mathcal{T}_{j,t}\}$ is deterministic. Finally, we will show that for $i \in K_{opt}$, the event $A_{t'-1}$ ensures that Huber-LASSO Bandit chooses arm i at time t' when $X_{t'} \in U_i$ so $\mathcal{S}'_{i,t} \subset \mathcal{S}_{i,t}$. Then, we will use Proposition 3.2 to show that events $A_{t'-1}$ occur frequently enough so that $|\mathcal{S}'_{i,t}|$ is sufficiently large. Then, we can use Proposition 3.1 with $\mathcal{A} = \mathcal{S}_{i,t}$ and $\mathcal{A}' = \mathcal{S}'_{i,t}$ to prove Proposition **??**.

A Technical Lemmas

Consider a linear model $Y = \mathbf{X}\beta^* + \varepsilon$, where $Y = (Y_1,\ldots,Y_n)^{\mathrm{T}} \in \mathbb{R}^n$ is the response vector, $\mathbf{X} = (X_1,\ldots,X_n)^{\mathrm{T}} \in \mathbb{R}^{n\times d}$ is the design matrix, and the noise vector $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_n)^{\mathrm{T}}$ comprises independent random variables satisfying $\mathbb{E}(\varepsilon_t|X_t) = 0$ and $\mathbb{E}(|\varepsilon_t|^{1+\delta}|X_t) \leq \nu_\delta$ for some $\delta \in (0,1]$. Here the sequence $\{X_t\}_{t=1}^n$ is assumed to form an adapted sequence of features, that is, X_t is allowed to depend on $\{(X_{t'},Y_{t'})\}_{t'=1}^{t-1}$.

Recall that given a robustification parameter $\tau > 0$ and a regularization parameter $\lambda > 0$, the Huber-Lasso estimator is defined as

$$\widehat{\beta}_n(\tau,\lambda) \in \operatorname*{argmin}_{\beta\in\mathbb{R}^d} \big\{L_{n,\tau}(\beta) + \lambda\|\beta\|_1\big\}, \tag{A.1}$$

where $L_{n,\tau}(\beta) = (1/n)\sum_{t=1}^{n} \ell_\tau(Y_t - X_t^{\mathrm{T}}\beta)$ is the empirical Huber loss.

The two key steps toward establishing upper bounds on the estimation error $\|\widehat{\beta}_n(\tau, \lambda) - \beta^*\|_l$ $(l = 1, 2)$ are to (i) bound the ℓ_∞-norm of the gradient vector

$$\nabla L_{n,\tau}(\beta^*) = -\frac{1}{n}\sum_{t=1}^{n} \ell'_\tau(\varepsilon_t)X_t \quad \text{with} \quad \ell'_\tau(\varepsilon_t) = \mathrm{sign}(\varepsilon_t)\min\{|\varepsilon_t|, \tau\},$$

and (ii) to show that the empirical loss satisfies some local *restricted strong convexity* (RSC) property; see event $\Omega_n(\phi, r)$ in Definition 2.2. To bound $\|\nabla L_{n,\tau}(\beta^*)\|_\infty$ with adaptively sampled data, we will extend the argument used in the proof of Theorem 1 by Sun, Zhou and Fan (2020) within the i.i.d. setting.

For the local RSC property, the independence assumption is crucial to show that event $\Omega_n(\phi, r)$, with properly chosen ϕ and r, holds with high probability (Sun, Zhou and Fan, 2020). However, for adaptively sampled data, particularly X_t, the existing result cannot be directly applied. The key ingredient that ensures RSC in the absence of independence is the existence of some subset $\mathcal{A} \subseteq [n]$, the size of which is proportional to n, such that $\{X_t, t \in \mathcal{A}\}$ comprises i.i.d. samples from some distribution $\mathbb{P}_X$. To see that, from the convexity of the Huber loss we have

$$\{\ell'_\tau(u) - \ell'_\tau(v)\}(u - v) \geq 0, \quad \forall u, v \in \mathbb{R}.$$

It follows that, for any $\delta \in \mathbb{R}^d$,

$$\begin{aligned} \langle \delta, \Delta_n(\delta)\rangle &= \frac{1}{n}\sum_{t=1}^{n}\{\ell'_\tau(\varepsilon_t) - \ell'_\tau(\varepsilon_t - X_t^{\mathrm{T}}\delta)\}X_t^{\mathrm{T}}\delta \\ &\geq \frac{1}{n}\sum_{t\in\mathcal{A}}\{\ell'_\tau(\varepsilon_t) - \ell'_\tau(\varepsilon_t - X_t^{\mathrm{T}}\delta)\}X_t^{\mathrm{T}}\delta = \frac{|\mathcal{A}|}{n}\langle \delta, \Delta_{\mathcal{A}}(\delta)\rangle, \end{aligned}$$

where $\Delta_{\mathcal{A}}(\delta) = \nabla L_{\mathcal{A},\tau}(\beta^* + \delta) - \nabla L_{\mathcal{A},\tau}(\beta^*)$ and $L_{\mathcal{A},\tau}(\beta) = |\mathcal{A}|^{-1}\sum_{t\in\mathcal{A}} \ell_\tau(Y_t - X_t^{\mathrm{T}}\beta)$.

In light of the above analysis, we define the event

$$\Omega_{\mathcal{A}}(\phi, r) = \Big\{\langle \Delta_{\mathcal{A}}(\delta), \delta\rangle \geq \phi\|\delta\|_2^2 \;\; \forall \delta \in \mathbb{R}^d \text{ s.t. } \|\delta_{\mathcal{S}^c}\|_1 \leq 3\|\delta_{\mathcal{S}}\|_1, \, \|\delta\|_2 \leq r\Big\}, \tag{A.2}$$

where r represents a local radius parameter and $\phi > 0$ denotes a curvature parameter. The following lemma provides conditions on the cardinality of $\mathcal{A}$, denoted $|\mathcal{A}|$, the robustification parameter τ and the local radius r under which event $\Omega_{\mathcal{A}}(\phi, r)$ holds with high probability for some $\phi > 0$.

Lemma A.1. Assume that $\mathcal{A} \subseteq [n]$ determines a subset of i.i.d. samples $\{X_t, t \in \mathcal{A}\} \sim \mathbb{P}_X$, satisfying $\|X_t\|_\infty \leq x_{\max}$ and $\Sigma = \mathbb{E}_{X\sim\mathbb{P}_X}(XX^{\mathrm{T}}) \in \mathfrak{C}(\mathcal{S}, \underline{\phi})$ for some $x_{\max}, \underline{\phi} > 0$.

(i) (Huber-Lasso) Assume $\beta^* \in \mathbb{R}^d$ is s_0-sparse. For any $\alpha \in (0, 1)$, the event $\Omega_{\mathcal{A}}(\underline{\phi}/4, r)$ occurs with probability at least $1 - \alpha$ as long as $\tau \geq 5\max(\nu_\delta^{1/(1+\delta)}, x_{\max}s_0^{1/2}r)$ and

$$16x_{\max}^2 s_0 \frac{8\sqrt{\log(2d)} + \sqrt{\log(1/\alpha)}}{\underline{\phi}\sqrt{2|\mathcal{A}|}} + \nu_\delta\left(\frac{5}{\tau}\right)^{1+\delta} \leq \frac{3}{4}. \tag{A.3}$$

(ii) (Huber) For any $\alpha \in (0,1)$, the event $\Omega_{\mathcal{A}}(\underline{\phi}/4, r)$ occurs with probability at least $1-\alpha$ if $\tau \geq 3\max\{\nu_\delta^{1/(1+\delta)}, x_{\max} d^{1/2} r/2\}$ and

$$x_{\max}^2 d \frac{4\sqrt{2} + \sqrt{\log(1/\alpha)}}{\underline{\phi}\sqrt{2|\mathcal{A}|}} + \nu_\delta \left(\frac{3}{\tau}\right)^{1+\delta} \leq \frac{3}{4}. \tag{A.4}$$

Proof of Lemma A.1. PROOF OF PART (I). Given $r > 0$, define the local parameter set $\Theta_r = \{\delta \in \mathbb{R}^d : \|\delta_{\mathcal{S}^c}\|_1 \leq 3\|\delta_{\mathcal{S}}\|_1, \|\delta\|_2 \leq r\}$. It suffices to show that with high probability, $\langle \delta, \Delta_{\mathcal{A}}(\delta)\rangle \geq \phi\|\delta\|_2^2$ holds uniformly for $\delta \in \Theta_r$, where ϕ and r are appropriately chosen. Note that for any $\delta \in \Theta_r$,

$$\|\delta\|_1 = \|\delta_{\mathcal{S}}\|_1 + \|\delta_{\mathcal{S}^c}\|_1 \leq 4\|\delta_{\mathcal{S}}\|_1 \leq 4 s_0^{1/2}\|\delta\|_2,$$

which, under the condition $\tau \geq 5 x_{\max} s_0^{1/2} r$, implies

$$|X_t^{\mathrm{T}}\delta| \leq \underbrace{4x_{\max}s_0^{1/2}}_{=:L}\|\delta\|_2 \leq \frac{4\tau}{5}, \quad \forall \delta \in \Theta_r.$$

Therefore, we have

$$\begin{aligned}
&\langle \delta, \Delta_{\mathcal{A}}(\delta)\rangle \geq \frac{1}{|\mathcal{A}|}\sum_{t\in\mathcal{A}}\{\ell'_\tau(\varepsilon_t) - \ell'_\tau(\varepsilon_t - X_t^{\mathrm{T}}\delta)\}X_t^{\mathrm{T}}\delta \\
&\geq \frac{1}{|\mathcal{A}|}\sum_{t\in\mathcal{A}}\{\ell'_\tau(\varepsilon_t) - \ell'_\tau(\varepsilon_t - X_t^{\mathrm{T}}\delta)\}X_t^{\mathrm{T}}\delta \cdot \mathbb{1}(|\varepsilon_t| \leq \tau/5) \geq \frac{1}{|\mathcal{A}|}\sum_{t\in\mathcal{A}}(X_t^{\mathrm{T}}\delta)^2\mathbb{1}(|\varepsilon_t| \leq \tau/5).
\end{aligned}$$

Define

$$\Psi_{\mathcal{A},r} = \sup_{\delta\in\Theta_r}\frac{1}{|\mathcal{A}|}\sum_{t\in\mathcal{A}}\{-f_\delta(X_t,\varepsilon_t) + \mathbb{E}f_\delta(X_t,\varepsilon_t)\},$$

where $f_\delta(X_t,\varepsilon_t) = \mathbb{1}(|\varepsilon_t| \leq \tau/5)\cdot(X_t^{\mathrm{T}}\delta)^2/\|\delta\|_2^2$ and

$$\mathbb{E}f_\delta(X_t,\varepsilon_t) = \frac{\|\delta\|_\Sigma^2 - \mathbb{E}\{\mathbb{1}(|\varepsilon_t| > \tau/5)\cdot(X_t^{\mathrm{T}}\delta)^2\}}{\|\delta\|_2^2} \geq \{1 - \nu_\delta(5/\tau)^{1+\delta}\}\frac{\|\delta\|_\Sigma^2}{\|\delta\|_2^2}.$$

Provided $\tau \geq 5\nu_\delta^{1/(1+\delta)}$, it follows that

$$\langle \delta, \Delta_{\mathcal{A}}(\delta)\rangle \geq \left\{1 - \nu_\delta(5/\tau)^{1+\delta} - \Psi_{\mathcal{A},r}/\underline{\phi}\right\}\cdot\underline{\phi}\|\delta\|_2^2. \tag{A.5}$$

To bound the supremum $\Psi_{\mathcal{A},r}$, we use a similar but simplified argument as that presented in the proof of Lemma C.3 in Sun, Zhou and Fan (2020). Note that $0 \leq f_\delta(X_t,\varepsilon_t) \leq L^2$ for all $\delta \in \Theta_r$, where $L = 4x_{\max}s_0^{1/2}$. Since $(X_t,\varepsilon_t), t \in \mathcal{A}$ are independent, it follows from McDiarmid's inequality that with probability at least $1-\alpha$,

$$\Psi_{\mathcal{A},r} \leq \mathbb{E}(\Psi_{\mathcal{A},r}) + L^2\sqrt{\frac{\log(1/\alpha)}{2|\mathcal{A}|}} = \mathbb{E}(\Psi_{\mathcal{A},r}) + 16x_{\max}^2 s_0\sqrt{\frac{\log(1/\alpha)}{2|\mathcal{A}|}}. \tag{A.6}$$

To bound the expectation $\mathbb{E}(\Psi_{\mathcal{A},r})$, by Rademacher symmetrization we have

$$\mathbb{E}(\Psi_{\mathcal{A},r}) \leq 2\mathbb{E}\left\{\sup_{\delta\in\Theta(r)}\frac{1}{|\mathcal{A}|}\sum_{t\in\mathcal{A}}e_t f_\delta(X_t,\varepsilon_t)\right\},$$

where $e_t, t \in \mathcal{A}$ are independent Rademacher random variables. Since $f_\delta(X_t, \varepsilon_t)$ is $(2L)$-Lipschitz function in $X_t^{\mathrm{T}}\delta/\|\delta\|_2$ for $\delta \in \Theta_r$, that is, for any $\delta, \delta' \in \Theta_r$,

$$|f_\delta(X_t, \varepsilon_t) - f_{\delta'}(X_t, \varepsilon_t)| \le 2L \cdot |X_t^{\mathrm{T}}\delta/\|\delta\|_2 - X_t^{\mathrm{T}}\delta'/\|\delta'\|_2|.$$

Then, by using Talagrand's contraction principle (see, e.g., (4.20) in **?**), we obtain

$$\begin{aligned}
\mathbb{E}(\Psi_{\mathcal{A},r}) &\le 2\mathbb{E}\bigg\{ \sup_{\delta \in \Theta_r} \frac{1}{|\mathcal{A}|} \sum_{t \in \mathcal{A}} e_t f_\delta(X_t, \varepsilon_t) \bigg\} \\
&\le 4L \cdot \mathbb{E}\bigg(\sup_{\delta \in \Theta_r} \frac{1}{|\mathcal{A}|} \sum_{t \in \mathcal{A}} e_t X_t^{\mathrm{T}}\delta/\|\delta\|_2 \bigg) \\
&\le 4L \cdot \mathbb{E}\bigg\| \frac{1}{|\mathcal{A}|} \sum_{t \in \mathcal{A}} e_t X_t \bigg\|_\infty \cdot \sup_{\delta \in \Theta_r} \frac{\|\delta\|_1}{\|\delta\|_2} \\
&\le 64 x_{\max} s_0 \cdot \mathbb{E}\bigg\| \frac{1}{|\mathcal{A}|} \sum_{t \in \mathcal{A}} e_t X_t \bigg\|_\infty. \qquad \text{(A.7)}
\end{aligned}$$

Write $S_j = \sum_{t \in \mathcal{A}} e_t X_{tj}$ for $j \in [d]$, and note that

$$\mathbb{E}(e^{\lambda \sum_{t \in \mathcal{A}} e_t X_{tj}} | \{X_t\}_{t \in \mathcal{A}}) = \prod_{t \in \mathcal{A}} \frac{1}{2}(e^{-\lambda X_{tj}} + e^{\lambda X_{tj}}) \le \prod_{t \in \mathcal{A}} e^{\lambda^2 X_{tj}^2/2} = e^{\lambda^2 \sum_{t \in \mathcal{A}} X_{tj}^2/2}, \quad \forall \lambda > 0.$$

Then, applying Theorem 2.5 in Boucheron$_L$*ugosi*$_M$*assart* (2013) yields

$$\mathbb{E}\bigg(\max_{1 \le j \le d} |S_j| \bigg| \{X_t\}_{t \in \mathcal{A}} \bigg) \le \sqrt{2 \sum_{t \in \mathcal{A}} X_{tj}^2 \log(2d)} \le x_{\max} \sqrt{2|\mathcal{A}| \log(2d)} \quad \text{almost surely.}$$

Combining this with (A.7) and (A.6), we obtain that with probability at least $1 - \alpha$,

$$\Psi_{\mathcal{A},r} \le 64 x_{\max}^2 s_0 \sqrt{\frac{2\log(2d)}{|\mathcal{A}|}} + 16 x_{\max}^2 s_0 \sqrt{\frac{\log(1/\alpha)}{2|\mathcal{A}|}}.$$

Combining this with (A.3) and (A.5) proves the claim.

Proof of Part (ii). With slight abuse of notation, we write $\Theta_r = \{\delta \in \mathbb{R}^d : \|\delta\|_2 \le r\} = \mathbb{B}_2(r)$. As long as $\tau \ge 1.5 x_{\max} d^{1/2} r$, we have

$$|X_t^{\mathrm{T}}\delta| \le \|X_t\|_2 \|\delta\|_2 \le x_{\max} d^{1/2} r \le \frac{2}{3}\tau, \quad \forall \delta \in \Theta_r.$$

Moreoever, define $\Psi_{\mathcal{A},r} = \sup_{\delta \in \Theta_r} |\mathcal{A}|^{-1} \sum_{t \in \mathcal{A}} \{-f_\delta(X_t, \varepsilon_t) + \mathbb{E} f_\delta(X_t, \varepsilon_t)\}$, where $f_\delta(X_t, \varepsilon_t) = \mathbb{1}(|\varepsilon_t| \le \tau/3) \cdot (X_t^{\mathrm{T}}\delta)^2/\|\delta\|_2^2$ and

$$\mathbb{E} f_\delta(X_t, \varepsilon_t) = \frac{\|\delta\|_\Sigma^2 - \mathbb{E}\{\mathbb{1}(|\varepsilon_t| > \tau/3) \cdot (X_t^{\mathrm{T}}\delta)^2\}}{\|\delta\|_2^2} \ge \{1 - \nu_\delta (3/\tau)^{1+\delta}\} \frac{\|\delta\|_\Sigma^2}{\|\delta\|_2^2}.$$

Provided $\tau \ge 3\nu_\delta^{1/(1+\delta)}$, it follows that

$$\langle \delta, \Delta_{\mathcal{A}}(\delta) \rangle \ge \frac{1}{|\mathcal{A}|} \sum_{t \in \mathcal{A}} (X_t^{\mathrm{T}}\delta)^2 \mathbb{1}(|\varepsilon_t| \le \tau/3) \ge \{1 - \nu_\delta (3/\tau)^{1+\delta} - \Psi_{\mathcal{A},r}/\underline{\phi}\} \cdot \underline{\phi} \|\delta\|_2^2. \qquad \text{(A.8)}$$

Since $0 \leq f_\delta(X_t, \varepsilon_t) \leq \|X_t\|_2^2 \leq x_{\max}^2 d$ (almost surely), by applying McDiarmid's inequality once again, we deduce that with probability at least $1-\alpha$, $\Psi_{\mathcal{A},r} \leq \mathbb{E}(\Psi_{\mathcal{A},r}) \leq x_{\max}^2 d\sqrt{\log(1/\alpha)/(2|\mathcal{A}|)}$. Additionally, it follows from Rademacher symmetrization and Talagrand's contraction principle that

$$\begin{aligned}
\mathbb{E}(\Psi_{\mathcal{A},r}) &\leq 2\mathbb{E}\bigg\{ \sup_{\delta \in \Theta_r} \frac{1}{|\mathcal{A}|} \sum_{t \in \mathcal{A}} e_t f_\delta(X_t, \varepsilon_t) \bigg\} \\
&\leq 4x_{\max} d^{1/2} \cdot \mathbb{E}\bigg\{ \sup_{\delta \in \Theta_r} \frac{1}{|\mathcal{A}|} \sum_{t \in \mathcal{A}} e_t X_t^{\mathrm{T}} \delta / \|\delta\|_2 \bigg\} \\
&\leq 4x_{\max} d^{1/2} \cdot \mathbb{E}\bigg\| \frac{1}{|\mathcal{A}|} \sum_{t \in \mathcal{A}} e_t X_t \bigg\|_2 \\
&\leq 4x_{\max} d^{1/2} \sqrt{\mathbb{E}\bigg\| \frac{1}{|\mathcal{A}|} \sum_{t \in \mathcal{A}} e_t X_t \bigg\|_2^2} \\
&\leq 4x_{\max} d^{1/2} \sqrt{\frac{\mathrm{tr}(\Sigma)}{|\mathcal{A}|}} \leq 4x_{\max}^2 \frac{d}{\sqrt{|\mathcal{A}|}}.
\end{aligned}$$

Putting together the pieces, we conclude that with probability at least $1-\alpha$,

$$\Psi_{\mathcal{A},r} \leq 4x_{\max}^2 d \sqrt{\frac{1}{|\mathcal{A}|}} + x_{\max}^2 d \sqrt{\frac{\log(1/\alpha)}{2|\mathcal{A}|}}.$$

Combining this with (A.4) and (A.8) proves the claim. □

B Proofs of Main Results

B.1 Proof of Proposition 2.1

Proof of Part (i). The proof follows a similar argument to that of Theorem B.2 in Sun, Zhou and Fan (2020), with some modifications to account for the adaptively collected data. For simplicity, we write $\widehat{\beta} = \widehat{\beta}_n(\tau, \lambda)$. Given $r \geq 1.5 s_0^{1/2} \lambda / \phi$ and $\phi > 0$, we define $\gamma = \sup\{u \in [0,1] : (1-u)\beta + u\widehat{\beta} \in \mathbb{B}_2(r)\}$. We observe that $\gamma = 1$ if $\widehat{\beta} \in \beta + \mathbb{B}_2(r)$, and $\gamma \in (0,1)$ otherwise. Define the intermediate estimator $\widetilde{\beta}$ as $\gamma\widehat{\beta} + (1-\gamma)\beta^*$, which satisfies that (i) $\widetilde{\beta} \in \beta^* + \mathbb{B}_2(r)$, (ii) $\widetilde{\beta}$ lies on the boundary of $\beta^* + \mathbb{B}_2(r)$ with $\gamma \in (0,1)$, and (iii) $\widetilde{\beta} = \widehat{\beta}$ with $\gamma = 1$ if $\widehat{\beta} \in \beta^* + \mathbb{B}_2(r)$. Applying Lemma C.1 in Sun, Zhou and Fan (2020) to the loss function $L_{n,\tau}(\beta)$, we have

$$\langle \nabla L_{n,\tau}(\widetilde{\beta}) - \nabla L_{n,\tau}(\beta^*), \widetilde{\beta} - \beta^* \rangle \leq \gamma \langle \nabla L_{n,\tau}(\widehat{\beta}) - \nabla L_{n,\tau}(\beta^*), \widehat{\beta} - \beta^* \rangle. \tag{B.1}$$

Recall that $\mathcal{S} = \mathrm{supp}(\beta^*) \subseteq [d]$ is the support of β^*. Following a similar argument to the proof of Theorem B.2 in Sun, Zhou and Fan (2020), we conclude that conditioned on the event $\{\lambda > 2\|\nabla L_{n,\tau}(\beta^*)\|_\infty\}$, $\widehat{\beta}$ satisfies

$$\|(\widehat{\beta} - \beta^*)_{\mathcal{S}^c}\|_1 \leq 3\|(\widehat{\beta} - \beta^*)_{\mathcal{S}}\|_1. \tag{B.2}$$

The above inequality also applies to $\widetilde{\beta}$ based on its construction.

Since $\widehat{\beta}$ is the minimizer of $\beta \mapsto L_{n,\tau}(\beta) + \lambda\|\beta\|_1$, there exists a subgradient $\widehat{z} \in \partial\|\widehat{\beta}\|_1$ such that $\langle \widehat{z}, \beta^* - \widehat{\beta}\rangle \le \|\beta^*\|_1 - \|\widehat{\beta}\|_1$ and $\langle \nabla L_{n,\tau}(\widehat{\beta}) + \lambda\widehat{z}, \widehat{\beta} - \beta^*\rangle \le 0$. Conditioned on the event $\{\lambda > 2\|\nabla L_{n,\tau}(\beta^*)\|_\infty\}$, applying Hölder's inequality yields

$$\begin{aligned}\langle \nabla L_{n,\tau}(\widehat{\beta}) - \nabla L_{n,\tau}(\beta^*), \widehat{\beta} - \beta^*\rangle &< \lambda(\|\beta^*\|_1 - \|\widehat{\beta}\|_1) + \frac{\lambda}{2}\|\widehat{\beta} - \beta^*\|_1 \\ &\le \lambda\{\|(\widehat{\beta} - \beta^*)_{\mathcal{S}}\|_1 - \|\widehat{\beta}_{\mathcal{S}^c}\|_1\} + \frac{\lambda}{2}\|\widehat{\beta} - \beta^*\|_1 \le \frac{3\lambda}{2}\|(\widehat{\beta} - \beta^*)_{\mathcal{S}}\|_1.\end{aligned}$$

Combining this with (B.1), we have

$$\langle \nabla L_{n,\tau}(\widetilde{\beta}) - \nabla L_{n,\tau}(\beta^*), \widetilde{\beta} - \beta^*\rangle < \frac{3\lambda}{2}\|(\widetilde{\beta} - \beta^*)_{\mathcal{S}}\|_1.$$

For the left-hand side, it holds conditioned on $\Omega_n(\phi, r)$ that $\langle \nabla L_{n,\tau}(\widetilde{\beta}) - \nabla L_{n,\tau}(\beta^*), \widetilde{\beta} - \beta^*\rangle \ge \phi\|\widetilde{\beta} - \beta^*\|_2^2$, which further implies

$$\phi\|\widetilde{\beta} - \beta^*\|_2^2 < \frac{3\lambda}{2}\|(\widetilde{\beta} - \beta^*)_{\mathcal{S}}\|_1 \le \frac{3\lambda}{2}s_0^{1/2}\|\widetilde{\beta} - \beta^*\|_2.$$

Canceling out $\|\widetilde{\beta} - \beta^*\|_2$, we find that $\|\widetilde{\beta} - \beta^*\|_2 < 3s_0^{1/2}\lambda/(2\phi) \le r$. This indicates that $\widetilde{\beta}$ falls within the interior of the neighborhood $\beta^* + \mathbb{B}_2(r)$. Via proof by contradiction, we must have $\gamma = 1$, and consequently, $\widetilde{\beta} = \widehat{\beta}$. It further follows from (B.2) that

$$\|\widehat{\beta} - \beta^*\|_1 \le \|(\widehat{\beta} - \beta^*)_{\mathcal{S}}\|_1 + \|(\widehat{\beta} - \beta^*)_{\mathcal{S}^c}\|_1 \le 4\|(\widehat{\beta} - \beta^*)_{\mathcal{S}}\|_1 \le 6s_0\lambda/\phi.$$

It remains to establish a lower bound for λ so that the event $\{\lambda > \|\nabla L_{n,\tau}(\beta^*)\|_\infty\}$ occurs with high probability. Denote $Z_t = X_t/x_{\max} = (Z_{t1}, \ldots, Z_{td})^{\mathrm{T}} \in \mathbb{R}^d$ satisfying $\|Z_t\|_{\max} \le 1$, and let $\mathcal{F}_t = \sigma(\{(X_{t'}, Y_{t'})\}_{t'=1}^t)$ be the σ-field generated by $\{(X_{t'}, Y_{t'})\}_{t'=1}^t$. To bound

$$\|\nabla L_{n,\tau}(\beta^*)\|_\infty = x_{\max}\max_{1\le j\le d}\left|\frac{1}{n}\sum_{t=1}^n \ell'_\tau(\varepsilon_t)Z_{tj}\right|,$$

we follow and extend the argument used in the proof of Theorem 1 by Sun, Zhou and Fan (2020) under the i.i.d. setting. To this end, define

$$\psi(u) = \mathrm{sign}(u)\min\{|u|, 1\} \quad \text{and} \quad S_{n,j} = \sum_{t=1}^n \psi(\varepsilon_t/\tau)Z_{tj},$$

so that $\|\nabla L_{n,\tau}(\beta^*)\|_\infty = (x_{\max}\tau/n)\max_{1\le j\le d}|S_{n,j}|$. By Markov's inequality, it holds for any $z \ge 0$ that

$$\mathbb{P}(S_{n,j} > 2z) \le e^{-2z}\mathbb{E}(e^{S_{n,j}}) = e^{-\nu_\delta z}\mathbb{E}\{e^{S_{n-1,j}}\mathbb{E}(e^{\psi_1(\varepsilon_n/\tau)Z_{nj}}|\mathcal{F}_{n-1})\}.$$

Using the property that $-\log(1 - u + |u|^{1+\delta}) \le \psi(u) \le \log(1 + u + |u|^{1+\delta})$ for any $u \in \mathbb{R}$ and $\delta > 0$, we have $e^{\psi_1(\varepsilon_n/\tau)Z_{nj}} \le 1 + Z_{nj}\varepsilon_n/\tau + |\varepsilon_n/\tau|^{1+\delta}$, which in turn implies

$$\mathbb{E}(e^{\psi_1(\varepsilon_n/\tau)Z_{nj}}|\mathcal{F}_{n-1}) \le \mathbb{E}(1 + Z_{nj}\varepsilon_n/\tau + |\varepsilon_n/\tau|^{1+\delta}|\mathcal{F}_{n-1}) \le 1 + \nu_\delta/\tau^{1+\delta} \le e^{\nu_\delta/\tau^{1+\delta}}.$$

Applying this bound repeatedly, and taking $\tau = \tau_0(n/z)^{1/(1+\delta)}$ with $\tau_0 \ge \nu_\delta^{1/(1+\delta)}$, it follows that

$$\mathbb{P}(S_{n,j} > 2z) \le e^{-2z + n\nu_\delta/\tau^{1+\delta}} = e^{-z}.$$

Through the same argument, we also have $\mathbb{P}(-S_{n,j} > 2z) \le e^{-z}$. Furthermore, taking the union bound over $j = 1, \ldots, d$, we conclude that with probability at least $1 - 2de^{-z}$,

$$\|\nabla L_{n,\tau}(\beta^*)\|_\infty = \frac{x_{\max}\tau}{n} \max_{1\le j\le d} |S_{n,j}| < \frac{2x_{\max}\tau z}{n} = 2x_{\max}\tau_0 \left(\frac{z}{n}\right)^{\delta/(1+\delta)}.$$

Consequently, we let $\lambda \ge 4x_{\max}\tau_0(z/n)^{\delta/(1+\delta)}$ so that the event $\{\lambda > 2\|\nabla L_{n,\tau}(\beta^*)\|_\infty\}$ holds with probability at least $1 - 2de^{-z}$.

Putting the pieces together, we conclude that the error bounds $\|\widehat{\beta} - \beta^*\|_2 \le 3s_0^{1/2}\lambda/(2\phi)$ and $\|\widehat{\beta} - \beta^*\|_1 \le s_0\lambda/\phi$ hold true with probability at least $1 - 2de^{-z} - \mathbb{P}\{\Omega_n(\phi, r)^{\rm c}\}$, as claimed.

Proof of Part (ii). The proof is similar to that of part (i). Note that the Huber estimator $\widehat{\beta} = \widehat{\beta}_n(\tau) \in \mathrm{argmin}_{\beta\in\mathbb{R}^d} L_{n,\tau}(\beta)$ satisfies the first-order condition $\nabla L_{n,\tau}(\widehat{\beta}) = 0$. We thus have

$$\langle \nabla L_{n,\tau}(\widetilde{\beta}) - \nabla L_{n,\tau}(\beta^*), \widetilde{\beta} - \beta^*\rangle = -\langle \nabla L_{n,\tau}(\beta^*), \widetilde{\beta} - \beta^*\rangle \le \|\nabla L_{n,\tau}(\beta^*)\|_2 \cdot \|\widetilde{\beta} - \beta^*\|_2.$$

To bound $\|\nabla L_{n,\tau}(\beta^*)\|_2$ with $\tau = \tau_0(n/z)^{1/(1+\delta)}$, following the same argument as above, it holds that with probability at least $1 - 2de^{-z}$,

$$\|\nabla L_{n,\tau}(\beta^*)\|_2 \le d^{1/2}\|\nabla L_{n,\tau}(\beta^*)\|_\infty < 2x_{\max}\tau_0 d^{1/2}(z/n)^{\delta/(1+\delta)}.$$

Conditioned on event $\Omega_n(\phi, r)$, the left-hand side of (B.1) is bounded from below by $\phi\|\widetilde{\beta} - \beta^*\|_2^2$. On the other hand, with probability at least $1-2de^{-z}$, the right-hand side is bounded from above by

$$2x_{\max}\tau_0 d^{1/2}\left(\frac{z}{n}\right)^{\delta/(1+\delta)} \cdot \|\widetilde{\beta} - \beta^*\|_2.$$

Provided $r \ge 2x_{\max}\tau_0 d^{1/2}(z/n)^{\delta/(1+\delta)}/\phi$, $\widetilde{\beta}$ falls within the interior of the neighborhood $\beta^* + \mathbb{B}_2(r)$. Using proof by contradiction, we must have $\widetilde{\beta} = \widehat{\beta}$, and thus the claimed bound follows immediately. □

B.2 Proof of Proposition 3.1

Proof of Part (i). Recall that $\Delta_{\mathcal{A}}(\delta) = \nabla L_{\mathcal{A},\tau}(\beta^* + \delta) - \nabla L_{\mathcal{A},\tau}(\beta^*)$ satisfies

$$\langle \delta, \Delta_{\mathcal{A}}(\delta)\rangle \ge \frac{|\mathcal{A}'|}{|\mathcal{A}|}\langle \delta, \Delta_{\mathcal{A}'}(\delta)\rangle \ge \frac{p}{2}\langle \delta, \Delta_{\mathcal{A}'}(\delta)\rangle, \quad \forall \delta \in \mathbb{R}^d.$$

By Lemma A.1, there exists an absolute constant $C_1 > 1$ such that for any $\alpha \in (0, 1)$, the event $\Omega_{\mathcal{A}}(p\underline{\phi}/8, \tau/(5x_{\max}s_0^{1/2}))$ holds with probability at least $1 - \alpha$ as long as $|\mathcal{A}| \ge C_1 p^{-1}(x_{\max}^2 s_0/\underline{\phi})^2 \log(2d/\alpha)$. Moreover, let λ satisfy

$$r := \frac{3s_0^{1/2}\lambda}{2\phi} \le \frac{\tau}{5x_{\max}s_0^{1/2}} \quad \text{with} \quad \phi = \frac{p\underline{\phi}}{8},$$

so that that $\Omega_{\mathcal{A}}(\phi, \tau/(5x_{\max}s_0^{1/2})) \subseteq \Omega_{\mathcal{A}}(\phi, r)$. Combining this with Proposition 2.1, we further obtain that as long as

$$4x_{\max}\tau_0\left(\frac{z}{|\mathcal{A}|}\right)^{\delta/(1+\delta)} \le \lambda \le \frac{p\underline{\phi}}{60x_{\max}s_0}\tau = \frac{p\underline{\phi}}{60x_{\max}s_0}\tau_0\left(\frac{|\mathcal{A}|}{z}\right)^{1/(1+\delta)},$$

it holds with probability at least $1-\alpha-2e^{-z+\log d}$ that

$$\|\widehat{\beta}_{\mathcal{A}}(\tau,\lambda)-\beta^*\|_2 \le 12(p\phi)^{-1}s_0^{1/2}\lambda \quad \text{and} \quad \|\widehat{\beta}_{\mathcal{A}}(\tau,\lambda)-\beta^*\|_1 \le 48(p\phi)^{-1}s_0\lambda,$$

as claimed.

Proof of Part (ii). Following the proof of part (i) and employing the part (ii) of Lemma A.1 instead, there exists an absolute constant $C_2>1$ such that for any $\alpha\in(0,1)$, the event $\Omega_{\mathcal{A}}(p\phi/8,\tau/(1.5x_{\max}d^{1/2}))$ occurs with probability at least $1-\alpha$ as long as $|\mathcal{A}|\ge C_2p^{-1}(x_{\max}^2d/\phi)^2\log(e/\alpha)$. Furthermore, let r satisfy

$$\frac{16x_{\max}\tau_0d^{1/2}}{p\phi}\left(\frac{z}{|\mathcal{A}|}\right)^{\delta/(1+\delta)} \le r \le \frac{\tau_0}{1.5x_{\max}d^{1/2}}\left(\frac{|\mathcal{A}|}{z}\right)^{1/(1+\delta)}$$

so that $\Omega_{\mathcal{A}}(p\phi/8,\tau/(1.5x_{\max}d^{1/2}))\subseteq\Omega_{\mathcal{A}}(p\phi/8,r)$. By combining this with Proposition 2.1-(ii) we conclude that with probability at least $1-\alpha-2e^{-z+\log d}$, $\|\widehat{\beta}_{\mathcal{A}}(\tau)-\beta^*\|_2\le r$, thus completing the proof. □

B.3 Proof of Proposition 3.2

Recall that at each $t\in\mathcal{T}_{i,t}$, we draw X_t randomly, sampled i.i.d from $\mathcal{P}_X$, and play arm i. Using Proposition 3.1, we derive a tail inequality for the forced sample estimators $\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau,\lambda)$ and $\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau)$. Moreover, we assumed that $\Sigma_i\in\Omega_n(\phi,r)$ where $\Sigma_i=\mathbb{E}_{X\sim\mathcal{P}_{X|X\in U_i}}[XX^{\mathrm{T}}]$ and $\mathbb{P}[X_t\in U_i]\ge p_*$. We will use the following results, which the proofs can be found from Bastani and Bayati (2020).

Lemma B.1. If $t\ge(Kq)^2$, then $|\mathcal{T}_{i,t}|\in[\frac{1}{2}q\log t, 6q\log t]$.

Lemma B.2. Let $\mathcal{T}'_{i,t}$ be the set of all r in $\mathcal{T}_{i,t}$ such that X_r belongs to U_i. Then, with probability at least p_* that given $r\in\mathcal{T}_{i,t}$ we have $r\in\mathcal{T}'_{i,t}$. Furthermore, $\{X_r\}_{r\in\mathcal{T}'_{i,t}}$ are i.i.d. from $\mathcal{P}_{X|X\in U_i}$.

Lemma B.3. Let $\mathcal{T}'_{i,t}\subset\mathcal{T}_{i,t}$ be define as in Lemma B.2. Then, for all $t\ge(Kq)^2$, the event $\frac{|\mathcal{T}'_{i,t}|}{|\mathcal{T}_{i,t}|}\ge\frac{p_*}{2}$ occurs with probability at least $1-\frac{2}{t^4}$.

Proof of Proposition 3.2

Proof of Part (i)

Observe that

$$\|X^{\mathrm{T}}\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau,\lambda)-X^{\mathrm{T}}\beta^*\|_1\le x_{\max}\sqrt{d}\|\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau,\lambda)-\beta^*\|_2. \tag{B.3}$$

Hence, it is enough to show that the event $\|\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau,\lambda)-\beta^*\|_1\le\frac{h}{4x_{\max}}$ happens with high probability. We use Lemma B.2, Lemma B.3 and then apply Proposition 3.1 with $\mathcal{A}=\mathcal{T}_{i,t}$, $\mathcal{A}'=\mathcal{T}'_{i,t}$

Observe that

$$|\mathcal{T}_{i,t}|\ge\frac{1}{2}q\log t\ge\left(\frac{hp\phi}{768x_{\max}^2s_0\tau_0}\right)^{-\frac{1+\delta}{\delta}}5\log t \tag{B.4}$$

provided

$$q \geq 2C_1\, p^{-1}\Big((x_{\max}^2 s_0/\phi)^2 \log(2d/\alpha) + (x_{\max}^2 s_0/\phi) z\Big) + 10\left(\frac{hp\phi}{768 x_{\max}^2 s_0 \tau_0}\right)^{-\frac{1+\delta}{\delta}}. \tag{B.5}$$

Using Lemma B.2, Lemma B.3 and Proposition 3.1 part (i) , with $\alpha = \frac{1}{t^4}$, and $z = \left(\frac{hp\phi}{768 x_{\max}^2 s_0 \tau_0}\right)^{\frac{1+\delta}{\delta}} |\mathcal{T}_{i,t}|$, $\tau = \tau_0(|\mathcal{T}_{i,t}|/z)^{1/(1+\delta)}$ with $\tau_0 \geq \nu_\delta^{1/(1+\delta)}$, $\frac{h}{4x_{max}} = 48\frac{s_0\lambda}{p\phi}$ with $\lambda = \frac{hp\phi}{192 x_{\max} s_0}$ satisfying

$$4x_{\max}\tau_0\left(\frac{z}{|\mathcal{T}_{i,t}|}\right)^{\delta/(1+\delta)} \leq \lambda \leq \frac{p\phi}{60 x_{\max} s_0}\tau_0\left(\frac{|\mathcal{T}_{i,t}|}{\tau}\right)^{1/(1+\delta)}, \tag{B.6}$$

we obtain

$$\mathbb{P}[\|\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau,\lambda) - \beta^*\|_1 > \frac{h}{4x_{\max}}] \leq \frac{1}{t^4} + 2e^{-\left(\frac{hp\phi}{768 x_{\max}^2 s_0 \tau_0}\right)^{\frac{1+\delta}{\delta}} |\mathcal{T}_{i,t}| + \log d} + \frac{2}{t^4} \tag{B.7}$$

$$\leq \frac{5}{t^4} \tag{B.8}$$

provided $t \geq d$.

Proof of Part (ii). By Theorem 1 in Tropp (2012), we have

$$\mathbb{P}\Big[\lambda_{\min}\Big(\widehat{\Sigma}(\mathcal{T}_{i,t})\Big) \leq \frac{\gamma p}{2}\Big] \leq d e^{\frac{-|\mathcal{T}_{i,t}|\gamma p}{8\sqrt{d} x_{\max}}} \tag{B.9}$$

For $q \geq \frac{80\sqrt{d} x_{\max}}{\gamma p}$ and $t \geq \frac{d}{\alpha}$, we have

$$|\mathcal{T}_{i,t}| \geq \frac{1}{2} q \log t \geq \frac{40\sqrt{d} x_{\max}}{\gamma p} \log t \geq \frac{8\sqrt{d} x_{\max}}{\gamma p} \log\left(\frac{t^4 d}{\alpha}\right). \tag{B.10}$$

Therefore, the event $\lambda_{\min}\Big(\widehat{\Sigma}(\mathcal{T}_{i,t})\Big) \geq \frac{\gamma p}{2}$ occurs with probability at least $1 - \frac{\alpha}{t^4}$. When $q \geq \frac{192}{\gamma p} d^{\frac{1}{2}}$ and $t > \frac{2d+1}{\alpha}$, $|\mathcal{T}_{i,t}| > 32\lambda_{\min}^{-1}\Big(\widehat{\Sigma}(\mathcal{T}_{i,t})\Big) d^{1/2} \log(\frac{2d+1}{\alpha} t^2) \geq \frac{C_2}{p}\left(x_{\max}^4 \frac{d^2}{\phi^2} \log(\frac{e}{\alpha}) + x_{\max}^2 \frac{d}{\phi} z\right)$ Choosing

$$q > \frac{2C_2 d}{p}\left(x_{\max}^4 \frac{d}{\phi^2} \log(\frac{e}{\alpha}) + \frac{x_{\max}^2}{\phi} z\right) + 10\left(\frac{hp\phi}{768 d^{1/2} x_{\max}^2 \tau_0}\right)^{-\frac{1+\delta}{\delta}}$$

and $t \geq d$, then we can use Lemma B.2, Lemma B.3 and Proposition 3.1 part (ii) , with $\alpha = \frac{1}{t^4}$, $z = \left(\frac{hp\phi}{768 d^{1/2} x_{\max}^2 \tau_0}\right)^{\frac{1+\delta}{\delta}} |\mathcal{T}_{i,t}|$, $\tau = \tau_0(|\mathcal{T}_{i,t}|/z)^{1/(1+\delta)}$ with $\tau_0 \geq \nu_\delta^{1/(1+\delta)}$, and $r = \frac{h}{4x_{\max}}$, then the event $\|\widehat{\beta}_{\mathcal{T}_{i,t}}(\tau) - \beta^*\|_2 \geq \frac{h}{4x_{\max}}$ occurs with probability at most $\frac{5}{t^4}$.

$\square$

B.4 Proof of Huber-Lasso/Huber tail inequalities for all-sample estimator

In this section, we derive the tail inequality for the all-sample estimator $\widehat{\beta}_{\mathcal{S}_{i,t}}(\tau, \lambda_{2,t})$ and $\widehat{\beta}_{\mathcal{S}_{i,t}}(\tau)$ for arms in K_{opt}.

Lemma B.4. Given $i \in [K]$, if A_{t-1} and B_{t-1} hold and the events X_t belongs to U_i, and $t \notin \cup_{j\in[K]}\mathcal{T}_{i,t}$.

1. The Huber-Lasso bandit uses the forced-sample estimator $\widehat{\beta}_{\mathcal{T}_{i,t-1}}(\tau, \lambda_1)$ arrive at $\widehat{\mathcal{K}} = \{i\}$, implying that it plays the optimal arm at time t.
2. The Huber bandit uses the forced-sample estimator $\widehat{\beta}_{\mathcal{T}_{i,t-1}}(\tau)$ arrive at $\widehat{\mathcal{K}} = \{i\}$, implying that it plays the optimal arm at time t.

Proof. 1. Using the fact that $X_t \in U_i$, we have

$$X_t^{\mathrm{T}}\beta_i \geq h + \max_{j\neq i} X_t^{\mathrm{T}}\beta_j. \tag{B.11}$$

Therefore, for each $j \in [K] \setminus \{i\}$, since A_{t-1} holds, we get

$$X_t^{\mathrm{T}}[\widehat{\beta}_{\mathcal{T}_{i,t-1}}(\tau,\lambda) - \widehat{\beta}_{\mathcal{T}_{j,t-1}}(\tau,\lambda)] = X_t^{\mathrm{T}}[\widehat{\beta}_{\mathcal{T}_{i,t-1}}(\tau,\lambda) - \beta_i] - X_t^{\mathrm{T}}[\widehat{\beta}_{\mathcal{T}_{j,t-1}}(\tau,\lambda) - \beta_j] + X_t^{\mathrm{T}}(\beta_i - \beta_j) \tag{B.12}$$

$$\geq -x_{\max}\frac{h}{4x_{\max}} - x_{\max}\frac{h}{4x_{\max}} + h \tag{B.13}$$

$$\geq \frac{h}{2}. \tag{B.14}$$

The proof is similar to that of part 1. □

Lemma B.5. For all t with $t \geq (Kq)^2$, the event A_t holds with probability at least $1 - \frac{C}{t^4}$.

Lemma B.6. For all t with $t \geq (Kq)^2$, the event B_t holds with probability at least $1 - \frac{C}{t^4}$.

Proof. □

Lemma B.7. Let $i \in [K]$ and let $\mathcal{S}_{i,t}^{'} \subset [t]$ be the set of all time periods r such that the events $X_r \in U_i$ and A_{r-1} hold and we are not forced-sampling any arm $j \in [K]$. Then the following properties are satisfied

1. $\{X_r \mid r \in \mathcal{S}_{i,t}^{'}\}$ are i.i.d. from distribution $\mathcal{P}_{X|X\in U_i}$.
2. For each $r \in [t] \setminus \cup_{j\in[K]}\mathcal{T}_{j,t}$, with probability at least $\frac{p_*}{2}$ we have $r \in \mathcal{S}_{i,t}^{'}$ provided $t \geq (Kq)^2$.
3. $\mathcal{S}_{i,t}^{'} \subset \mathcal{S}_{i,t}$.

Lemma B.8. Let $i \in [K]$ and let $\mathcal{S}_{i,t}^{''} \subset [t]$ be the set of all time periods r such that the events $X_r \in U_i$ and B_{r-1} hold and we are not forced-sampling any arm $j \in [K]$. Then the following properties are satisfied

1. $\{X_r \mid r \in \mathcal{S}''_{i,t}\}$ are i.i.d. from distribution $\mathcal{P}_{X|X\in U_i}$.

2. For each $r \in [t] \setminus \cup_{j\in[K]}\mathcal{T}_{j,t}$, with probability at least $\frac{p_*}{2}$ we have $r \in \mathcal{S}''_{i,t}$ provided $t \geq (Kq)^2$.

3. $\mathcal{S}''_{i,t} \subset \mathcal{S}_{i,t}$.

Lemma B.9 (All sample Huber estimator bound)**.** For an optimal arm $i \in K_{\text{opt}}$ and time horizon $t > 4(Kq)^2$, with robustification parameter $\tau = \tau_0(|\mathcal{S}_{i,t}|/z)^{1/(1+\delta)}$, where $\tau_0 \geq \nu_\delta^{1/(1+\delta)}$, and regularization parameter $\lambda = \frac{hp\phi}{192x_{\max}s_0}$, the all-sample estimators satisfy:

(i) $\|\widehat{\beta}_{\mathcal{S}_{i,t}}(\tau,\lambda) - \beta_i\|_1 \leq \frac{h}{4x_{\max}}$ with probability at least $1 - \frac{5}{t^4} - 2e^{-ctp_*^2}$, for some constant $c > 0$.

(ii) $\|\widehat{\beta}_{\mathcal{S}_{i,t}}(\tau) - \beta_i\|_2 \leq \frac{h}{4x_{\max}}$ with probability at least $1 - \frac{5}{t^4} - 2e^{-ctp_*^2}$, for some constant $c > 0$.

Proof. Part (i): Huber-Lasso Estimator. Apply Proposition 3.1 with $\mathcal{A} = \mathcal{S}_{i,t}$, $\mathcal{A}' = \mathcal{S}'_{i,t}$, where $|\mathcal{S}'_{i,t}| \geq \frac{tp_*}{4}$ with probability at least $1 - e^{-tp_*^2/36}$. Set:

$$\lambda = \frac{hp\phi}{192x_{\max}s_0}, \quad \alpha = \frac{1}{t^4}, \quad z = \left(\frac{hp\phi}{768x_{\max}^2 s_0\tau_0}\right)^{\frac{1+\delta}{\delta}} |\mathcal{S}_{i,t}|, \quad \tau = \left(\frac{hp\phi}{768x_{\max}^2 s_0}\right)^{-1/\delta}. \tag{B.15}$$

We can see that

$$4x_{\max}\tau_0\left(\frac{z}{|\mathcal{S}_{i,t}|}\right)^{\delta/(1+\delta)} = \frac{hp\phi}{192x_{\max}s_0}, \quad \frac{p\phi}{60x_{\max}s_0}\tau_0\left(\frac{|\mathcal{S}_{i,t}|}{z}\right)^{1/(1+\delta)} = \frac{64x_{\max}\tau_0^2}{5h} \geq \lambda. \tag{B.16}$$

and hence

$$\|\widehat{\beta}_{\mathcal{S}_{i,t}}(\tau,\lambda) - \beta_i\|_1 \leq 48\frac{s_0\lambda}{p\phi} = \frac{h}{4x_{\max}}, \tag{B.17}$$

with probability:

$$\begin{aligned} &1 - \frac{1}{t^4} - 2de^{-\left(\frac{hp\phi}{768x_{\max}^2 s_0\tau_0}\right)^{\frac{1+\delta}{\delta}}\frac{tp_*}{4}} - e^{-tp_*^2/36} \geq 1 - \frac{5}{t^4} - 2e^{-ctp_*^2}, \\ &c = \min\left(\frac{1}{36}, \frac{1}{4}\left(\frac{hp\phi}{768x_{\max}^2 s_0\tau_0}\right)^{\frac{1+\delta}{\delta}}\right). \end{aligned} \tag{B.18}$$

Part (ii): Huber Estimator. We have

$$\|\widehat{\beta}_{\mathcal{S}_{i,t}}(\tau) - \beta_i\|_2 \leq \frac{2}{c_l|\mathcal{S}_{i,t}|}\left\|\sum_{r\in\mathcal{S}_{i,t}}\ell'_\tau(\varepsilon_r)X_r\right\|_2 \leq \frac{2d^{1/2}\tau}{c_l|\mathcal{S}_{i,t}|}\max_{1\leq j\leq d}\left|\sum_{r\in\mathcal{S}_{i,t}}\ell'_\tau(\varepsilon_r)\frac{X_{rj}}{\tau}\right|, \tag{B.19}$$

with $c_l \geq \frac{p\phi}{8}$. Use the supermartingale:

$$M_t = \exp\left(\sum_{r=1}^{t}\left[\ell'_\tau(\varepsilon_r)\frac{X_{rj}}{\tau} - \frac{\nu_\delta}{\tau^{1+\delta}}\right]\right), \tag{B.20}$$

yielding:

$$\mathbb{P}\left(\max_{1\le j\le d}\left|\sum_{r\in\mathcal{S}_{i,t}}\ell'_\tau(\varepsilon_r)\frac{X_{rj}}{\tau}\right|>\frac{\nu_\delta|\mathcal{S}_{i,t}|}{2\tau_0^{1+\delta}}\right)\le 2de^{-\nu_\delta\frac{tp_*}{18}\tau_0^{-(1+\delta)}}. \tag{B.21}$$

Set $\tau=\tau_0^{2/(1+\delta)}(2|\mathcal{S}_{i,t}|)^{1/(1+\delta)}$, so:

$$\left\|\sum_{r\in\mathcal{S}_{i,t}}\ell'_\tau(\varepsilon_r)X_r\right\|_2\le\frac{\nu_\delta d^{1/2}\tau_0^{(1-\delta)/(1+\delta)}}{2^{1-1/(1+\delta)}}|\mathcal{S}_{i,t}|^{1+1/(1+\delta)}. \tag{B.22}$$

Thus:

$$\|\widehat{\beta}_{\mathcal{S}_{i,t}}(\tau)-\beta_i\|_2\le\frac{16\nu_\delta^{(1-\delta)/(1+\delta)^2}d^{1/2}}{p\underline{\phi}2^{1-1/(1+\delta)}}\left(\frac{tp_*}{9}\right)^{1/(1+\delta)}\le\frac{h}{4x_{\max}}, \tag{B.23}$$

with probability:

$$1-e^{-8tp_*^2/243}-2de^{-\nu_\delta\frac{tp_*}{18}\tau_0^{-(1+\delta)}}\ge 1-\frac{5}{t^4}-2e^{-ctp_*^2},\quad c=\min\left(\frac{8}{243},\frac{\nu_\delta}{18}\tau_0^{-(1+\delta)}\right). \tag{B.24}$$

$\square$

Lemma B.10. For $t>4(Kq)^2$, we have

1. $\mathbb{P}(|\mathcal{S}'_{i,t}|\ge\frac{tp}{4})\ge 1-e^{tp^2/36}$
2. $\mathbb{P}(|\mathcal{S}''_{i,t}|\ge\frac{tp_*}{9})\ge 1-e^{-\frac{8tp_*^2}{243}}$.

Proof. We will only present the proof of part 2 since the proofs for two parts are similar.

By definition of $\mathcal{S}''_{i,t}$, we have

$$\mathbb{1}_{r\in\mathcal{S}''_{i,t}}=1_{A_{r-1}}\cdot\mathbb{1}_{X_r\in U_i}\cdot\mathbb{1}_{r\notin\cup_{j\in[K]}\mathcal{T}_{j,t}}. \tag{B.25}$$

Observe that by construction of the forced-sampling sets, for each round of forced-sampling $n\in\{0,1,2,\dots\}$, the time periods $r\in[2^nKq+1,(2^{n+1}-1)Kq]$ are played contiguously without any forced-sampling (or updates to the forced-sampling estimators). Then, for any time $r\notin\{\mathcal{T}_{j,t'}\}_{j\in[K],t'\in[T]}$ where we do not perform forced-sampling. Let N_t be the largest integer satisfying $t>2^{N_t+1}Kq$. We define

$$V_{1,t}:=[2^{N_t}Kq+1,(2^{N_t+1}-1)Kq],\quad V_{2,t}=[2^{N_t+1}Kq+1,\min\{t,(2^{N_t+2}-1)Kq\}] \tag{B.26}$$

$$M_{i,t}:=\sum_{r\in V_{1,t}}\mathbb{1}_{r\in\mathcal{S}''_{i,t}}+\sum_{r\in V_{2,t}}\mathbb{1}_{r\in\mathcal{S}''_{i,t}}<\sum_{r=1}^{t}\mathbb{1}_{r\in\mathcal{S}''_{i,t}}=|\mathcal{S}''_{i,t}|. \tag{B.27}$$

By construction, both $V_{1,t}$ and $V_{2,t}$ are not containing the forced-sampling time steps and hence the forced-sample estimator are never updated within those intervals. As a result, we have

$$M_{i,t} = \sum_{r\in V_{1,t}} \mathbb{1}_{r\in A_{2^{N_t}Kq}} \cdot \mathbb{1}_{X_r\in U_i} \cdot \mathbb{1}_{r\notin \cup_{j\in[K]}\mathcal{T}_{j,t}} + \sum_{r\in V_{2,t}} \mathbb{1}_{r\in A_{2^{N_t+1}Kq}} \cdot \mathbb{1}_{X_r\in U_i} \cdot \mathbb{1}_{r\notin \cup_{j\in[K]}\mathcal{T}_{j,t}} \tag{B.28}$$

$$\geq \mathbb{1}_{r\in A_{2^{N_t}Kq}} \cdot \mathbb{1}_{r\in A_{2^{N_t+1}Kq}} \cdot \sum_{r\in V_{1,t}\cup V_{2,t}} \mathbb{1}_{X_r\in U_i}. \tag{B.29}$$

Now, we deal with a sum over independent random variables drawn from $\mathcal{P}_X$, each happening with probability p_* by Assumption 2.3. Since N_t is the largest integer such that $t > 2^{N_t+1}Kq$, we have

$$\begin{aligned}
|V_{1,t}\cup V_{2,t}| &= \Big(\min\{t, 2^{N_t+2}\}\Big)Kq - 2^{N_t+1}Kq + \Big(2^{N_t+1}Kq - Kq - 2^{N_t}Kq\Big) && \text{(B.30)}\\
&= \min\Big\{t - 2^{N_t}Kq - 2Kq, 3\cdot 2^{N_t}Kq - 2Kq\Big\} && \text{(B.31)}\\
&\geq \min\Big\{t - \frac{t}{2} - 2Kq, \frac{3}{4}\cdot 2^{N_t+2}Kq - 2Kq\Big\} && \text{(B.32)}\\
&\geq \min\Big\{\frac{t}{2} - 2Kq, \frac{3}{4}t - 2Kq\Big\} && \text{(B.33)}\\
&\geq \min\Big\{\frac{t}{2} - \frac{t}{4}, \frac{3}{4}t - \frac{t}{4}\Big\} = \frac{t}{4}. && \text{(B.34)}
\end{aligned}$$

The first inequality comes from $2^{N_t+2}Kq \geq t$ and the last inequaltiy comes from $t \geq 8Kq$. Moreover, we have

$$\begin{aligned}
V_{1,t}\cup V_{2,t} &\leq t - 2^{N_t}Kq - Kq && \text{(B.35)}\\
&\leq t - \frac{t}{4} - Kq && \text{(B.36)}\\
&\leq \frac{3t}{4}. && \text{(B.37)}
\end{aligned}$$

Observe that we have

$$2^{N_t+1}Kq > 2^{N_t}Kq > \frac{t}{4} > (Kq)^2 \tag{B.38}$$

since $t > 4(Kq)^2$ and $t < 2^{N_t+2}Kq$ by the definition of N_t. Hence, we can apply Lemma **??** combine with a union bound, we obtain

$$\begin{aligned}
\mathbb{P}(B_{2^{N_t}Kq} \text{ and } B_{2^{N_t+1}Kq}) &\geq 1 - \frac{5K}{(t/4)^2} - \frac{5K}{(t/2)^2} && \text{(B.39)}\\
&\geq 1 - \frac{100}{K^3q^4} > \frac{8}{9}. && \text{(B.40)}
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{E}(M_{i,t}) &\geq \mathbb{P}(B_{2^{N_t}Kq} \text{ and } B_{2^{N_t+1}Kq}) \cdot p_*|V_{1,t}\cup V_{2,t}| && \text{(B.41)}\\
&\geq \frac{2tp_*}{9}. && \text{(B.42)}
\end{aligned}$$

Using the Hoeffding inequality and (B.35), we have

$$\mathbb{P}(M_{i,t} - \mathbb{E}(M_{i,t}) \le -\eta) \le e^{-\frac{2\eta^2}{|V_{1,t} \cup V_{2,t}|}} \tag{B.43}$$

$$\le e^{-\frac{8\eta^2}{3t}}. \tag{B.44}$$

By choosing $\eta = \frac{tp_*}{9}$ and from (B.41), we obtain

$$\mathbb{P}(M_{i,t} \le \frac{2tp_*}{9} - \frac{tp_*}{9}) \le e^{-\frac{8t^2p_*^2}{9^2 3t}} = e^{-\frac{8tp_*^2}{243}}. \tag{B.45}$$

From the fact that $M_{i,t} \le |\mathcal{S}''_{i,t}|$, we have

$$\mathbb{P}(|\mathcal{S}''_{i,t}| < \frac{tp_*}{9}) \le e^{-\frac{8tp_*^2}{243}}. \tag{B.46}$$

□

References

Bastani, H. and Bayati, M. (2020). Online decision making with high-dimensional covariates. *Operations Research* **68** 276–294.

Fan, J., Wang, Z., Yang, Z. and Ye, C. (2023). Provably efficient high-dimensional bandit learning with batched feedbacks. *arXiv:2311.13180*.

Huber, P. J. (1973). Robust estimation: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics,* **12**, 389–434.

Kang, M. and Kim, G.-S. (2023). Heavy-tailed linear bandit with Huber regression. In *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, PMLR **216** 1027–1036.