Federated Linear Contextual Bandits with User-level Differential Privacy

Ruiquan Huang [1] **Huanyu Zhang** [2] **Luca Melis** [2] **Milan Shen** [2] **Meisam Hejazinia** [3] **Jing Yang** [1]

Abstract

This paper studies federated linear contextual bandits under the notion of user-level differential privacy (DP). We first introduce a unified federated bandits framework that can accommodate various definitions of DP in the sequential decision-making setting. We then formally introduce user-level central DP (CDP) and local DP (LDP) in the federated bandits framework, and investigate the fundamental trade-offs between the learning regrets and the corresponding DP guarantees in a federated linear contextual bandits model. For CDP, we propose a federated algorithm termed as `ROBIN` and show that it is near-optimal in terms of the number of clients M and the privacy budget ε by deriving nearly-matching upper and lower regret bounds when user-level DP is satisfied. For LDP, we obtain several lower bounds, indicating that learning under user-level (ε, δ)-LDP must suffer a regret blow-up factor at least $\min\{1/\varepsilon, M\}$ or $\min\{1/\sqrt{\varepsilon}, \sqrt{M}\}$ under different conditions.

1. Introduction

Federated learning (FL) (McMahan et al., 2017a) has become a trending distributed machine learning paradigm where numerous clients collaboratively train a prediction model under the coordination of a central server while keeping the local training data at each client. FL is motivated by various applications where real-world data are exogenously generated at edge devices, and it is desirable to protect the privacy of local data by only sharing model updates instead of the raw data.

While the majority of FL studies focus on the supervised learning setting, recently, a few researchers begin to extend FL to the multi-armed bandits (MAB) framework (Lai & Robbins, 1985; Auer et al., 2002; Bubeck & Cesa-Bianchi, 2012; Agrawal & Goyal, 2012; 2013), and have proposed several federated bandits models and learning algorithms (Shi & Shen, 2021; Shi et al., 2021; Dubey & Pentland, 2020; Huang et al., 2021; He et al., 2022a). Under the classical MAB model, a player chooses to play one out of a set of arms at each time slot. An arm, if played, will generate a reward that is randomly drawn from a fixed but unknown distribution. With all previous observations, the player needs to decide which arm to pull in each time in order to maximize the cumulative expected reward. MAB thus represents an online learning model that naturally captures the intrinsic exploration-exploitation tradeoff in many sequential decision-making problems. Under the new realm of federated bandits, each client is facing a local bandits model with shared parameters. While classical MAB allows immediate access to the sequentially generated data for decision-making, in federated bandits, local data streams are generated and analyzed at the clients, with periodic information exchange among clients. Such a model is naturally motivated by a corpus of applications, such as recommender systems, clinical trials, and cognitive radio, where the sequential decision-making involves multiple clients and is distributed by nature.

Meanwhile, *user-level* differential privacy (DP) has emerged as a stronger but more realistic notion of privacy, which guarantees the privacy of a user's entire contribution of data instead of individual samples. Roughly speaking, user-level DP requires that the output of an algorithm does not significantly change when a user's entire contribution has been changed. While user-level DP has been studied in applications involving *static* datasets, such as discrete distribution estimation (Acharya et al., 2022; Cummings et al., 2021), learning (Levy et al., 2021; Ghazi et al., 2021), and federated learning (Girgis et al., 2022), to the best of our knowledge, it has not been studied in the *online* setting, where data is generated sequentially as decisions are made and outcomes are observed.

In this work, we take an initial effort to incorporate user-level DP into the framework of federated bandits, where M clients work collaboratively to learn a shared bandits model through a central server. We note that ensuring user-level DP in federated bandits setting is extremely challenging, mainly due to the following reasons. First, due to the interactive sequential decision-making nature of bandits, we do

[1]School of EECS, The Pennsylvania State University, University Park, PA, USA [2]Meta, USA [3]Google, USA. Correspondence to: Jing Yang <yangjing@psu.edu>.

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

not have static datasets at clients. Rather, each local sample is generated online accordingly to the federated learning mechanism adopted by the system. As a result, its distribution depends on not only the historical data at the client, but also the entire history across all clients and the server. Such intricate dependency makes the definition of user-level DP elusive in the federated bandits setting.

Second, due to the bandit feedback, only the reward of the pulled arm is observed, where the arm-pulling decision depends on history and the DP mechanism. Thus, the local samples are non-independent and identically distributed (non-IID), which is in stark contrast to the IID assumption usually adopted in the literature of user-level DP.

Third, in the federated bandits setting, multiple clients jointly learn the shared bandits parameter collaboratively. In general, the number of data samples contributed by a single client grows linearly in time horizon T, which is unbounded. Thus, to achieve user-level DP, vanilla noise-adding DP mechanisms may require the corresponding noise variance scale in the same order. On the other hand, in order to obtain sublinear learning regret, it requires that the estimation error in the estimated model parameters decays fast in time. Thus, it is unclear whether it is possible to still achieve sublinear learning regret while guaranteeing user-level DP.

We tackle those aforementioned challenges explicitly in this work. Specifically, our main contributions can be summarized as follows.

First, we formally introduce a DP oriented federated bandits framework, which provides a principled viewpoint to capture and characterize potential privacy leakage in online decision-making problems. Our general framework can accommodate all previously introduced differential privacy notions in the bandits settings, such as joint DP (Shariff & Sheffet, 2018; Dubey & Pentland, 2020). We then specialize the framework to capture user-level DP, and introduce both central and local user-level DP.

Second, we investigate user-level central differential privacy (CDP), and study the fundamental trade-off between learning regret and DP guarantee. Under standard margin condition and diversity condition studied in conventional linear contextual bandits (Hao et al., 2020; Papini et al., 2021), we propose a near-optimal algorithm termed as `ROBIN` with user-level (ε, δ)-DP guarantee. `ROBIN` enjoys a regret of $\tilde{O}\left(\max\left\{1, \frac{d\log T}{M\varepsilon^2}\right\} C_0 d\log T\right)$, where C_0 is a margin parameter, d is the dimension of the features, M is the number of clients, and T is the time horizon. The near optimality is established by obtaining a minimax lower bound $\Omega\left(\max\left\{1, \frac{1}{M\varepsilon^2}\right\} C_0 d\log T\right)$ under the same CDP constraint and the diversity and margin conditions. Furthermore, we also investigate the lower bound without the margin condition. Compared to the non-private counterpart, our results indicate that when $\varepsilon = O(1/\sqrt{M})$, the regret suffers a blow-up factor of at least $\min\{M, \frac{1}{\varepsilon^2 M}\}$ or $\min\{\sqrt{M}, \frac{1}{\varepsilon\sqrt{M}}\}$, depending on whether the margin condition is imposed or not. When $\varepsilon = \Omega(1/\sqrt{M})$, imposing the user-level CDP constraint does not affect the hardness of the federated linear contextual bandits problem, regardless of whether the margin condition is satisfied.

Third, we study user-level local different privacy (LDP) under several settings and conditions. When $\varepsilon = O(1/M)$, the minimax regret lower bound is either $\Omega(M\log T)$ under the margin condition or $\Omega(M\sqrt{T})$ without the margin condition, suggesting that the best policy is to have the clients independently make arm-pulling decisions based on their own local datasets without information sharing. When $\varepsilon = \Omega(1/M)$, we obtain a minimax lower bound in the order of $\Omega(C_0 d\log T/\varepsilon)$ under the margin condition and $\Omega(\sqrt{dMT/\varepsilon})$ without the margin condition, indicating that any federated linear contextual bandits algorithm satisfying user-level LDP suffers a regret blow-up factor of at least $1/\sqrt{\varepsilon}$ without the margin condition, or $1/\varepsilon$ with the margin condition. Thus, the user-level LDP constraint makes the learning problem strictly harder than the non-private case. Moreover, we also develop a tighter lower bound for pure LDP, i.e. $\delta = 0$, which can be obtained by replacing ε in the CDP lower bound by $\varepsilon/\sqrt{M}$. A summary of our main results is shown in Table 1.

2. The Private Federated Bandits Framework

2.1. Notations

For any $M \in \mathbb{N}$, we denote $[M] := \{1, \ldots, M\}$. For any vector x and symmetric matrix V, we denote $\|x\|$ as the ℓ_2 norm of x, $\|x\|_V := \sqrt{x^\top V x}$, and $\lambda_{\min}(V)$, $\lambda_{\max}(V)$ refer to the minimum and maximum eigenvalues of V, respectively. For any matrix X, we use $\|X\|$ to denote its spectral norm, and use $X^\dagger$ to denote its pseudo-inverse. For any set $\mathcal{A}$, we denote $\mathcal{A}^n$ as its n-fold Cartesian product. For two probability measures $\mathbb{P}, \mathbb{Q}$, we use $d_{TV}(\mathbb{P}, \mathbb{Q})$ and $\mathtt{KL}(\mathbb{P}, \mathbb{Q})$ to denote their total variation distance and Kullback–Leibler distance, respectively. We denote $q_{\le t} := \{q_1, \ldots, q_t\}$. We use $\mathcal{F}(S_1, S_2)$ to denote the set of all possible measurable functions from set S_1 to another set S_2. $\tilde{O}(f)$ hides the logarithm term of f, i.e. $\tilde{O}(f) = O(f|\log f|)$.

2.2. The Federated Bandits Framework

We consider a federated linear contextual bandits setting captured by tuple $([M], \mathcal{A}, \mathcal{C}, \phi, \mathcal{P}, d)$, where $[M]$ is the set of clients, $\mathcal{A}$ is the set of arms, $\mathcal{C}$ is the set of contexts, $\mathcal{P} := \{\rho_i\}_{i\in[M]}$ is a set of distributions over context set $\mathcal{C}$, and $\phi : \mathcal{C} \times \mathcal{A} \to \mathbb{R}^d$ is the feature mapping.

At each time slot t, each client i observes a context $c_{i,t} \in \mathcal{C}$

Table 1: Regret Bounds of Linear Contextual Bandits under Different (ε, δ)-DP Constraints.

ALGORITHM	ASM.3.2	MODEL	CONSTRAINT	REGRET
LOWER BOUND (HE ET AL., 2022B)	✗	SING	ITEM-LEVEL JDP	$\Omega\left(\sqrt{dT} + d/\varepsilon\right)$
FEDUCB (DUBEY & PENTLAND, 2020)	✗	FED	ITEM-LEVEL JDP	$\tilde{O}\left(d^{3/4}M^{3/4}\sqrt{T/\varepsilon}\right)^{\dagger}$
P-FEDLINUCB (ZHOU & CHOWDHURY, 2023)	✗	FED	ITEM-LEVEL CDP	$\tilde{O}\left(d^{3/4}\sqrt{MT/\varepsilon} + d\sqrt{MT}\right)$
ROBIN (THEOREM 3.3)	✓	FED	USER-LEVEL CDP	$\tilde{O}\left(\min\left\{M, \max\left\{1, \frac{d\log T}{M\varepsilon^2}\right\}\right\} C_0 d\log T\right)$
LOWER BOUND (THEOREMS 4.5 AND 5.3)	✓	FED	USER-LEVEL CDP	$\Omega(\min\left\{M, \max\left\{1, \frac{1}{M\varepsilon^2}\right\}\right\} C_0 d\log T)$
	✗	FED	USER-LEVEL CDP	$\Omega\left(\min\left\{M, \max\left\{\sqrt{M}, \frac{1}{\varepsilon}\right\}\right\}\sqrt{dT}\right)$
	✓	FED	USER-LEVEL LDP*	$\Omega\left(\min\left\{M, 1/\varepsilon\right\} C_0 d\log T\right)$
	✗	FED	USER-LEVEL LDP*	$\Omega\left(\min\left\{M, \sqrt{M/\varepsilon}\right\}\sqrt{dT}\right)$

C_0: parameter of Asm 3.2, d: dimension of model parameter, M: number of clients, T: time horizon. SING and Fed stand for single-client and multi-client settings, respectively. JDP stands for jointly differential privacy. The standard non-private lower bounds are $\Omega(\sqrt{dMT})$ and $\Omega(C_0 d\log T)$, without or with Asm 3.2, respectively. $*$: The result for user-level $(\varepsilon, 0)$-LDP is presented in Corollary 5.5. †: We adopt the result in Zhou & Chowdhury (2023).

drawn according to distribution $\rho_i \in \mathcal{P}$. Then, client i pulls arm $a_{i,t} \in \mathcal{A}$ and receives a reward $r_{i,t} = \phi(c_{i,t}, a_{i,t})^\intercal \theta^* + \eta_{i,t}$, where $\theta^* \in \mathbb{R}^d$ is an unknown parameter vector shared among clients, $\eta_{i,t}$ is a random noise, and $\phi(c_{i,t}, a_{i,t}) \in \mathbb{R}^d$ is the feature vector associated with arm $a_{i,t}$ and context $c_{i,t}$. For simplicity, we denote $\phi(c_{i,t}, a)$ as $x_{i,t,a}$. We assume $\|\phi(c,a)\|_2 \le 1$, $\|\theta^*\|_2 \le 1$, and $\eta_{i,t}$ is an IID standard Gaussian random variable, i.e., $\eta_{i,t} \sim N(0,1)$. Such assumptions are standard in the linear contextual bandits literature (Abbasi-Yadkori et al., 2011; Chu et al., 2011).

We assume there exists a central server in the system, and similar to FL, the clients can communicate with the server periodically with zero latency. Specifically, the clients can send "local model updates" to the central server, which then aggregates and broadcasts the updated "global model" to the clients. (We will specify these components later.) We also assume that clients and server are fully synchronized (McMahan et al., 2017a).

For federated linear contextual bandits, the utility of primary interest is the expected cumulative regret among all clients, defined as:

$$\text{Regret}(M,T) = \mathbb{E}\left[\sum_{i=1}^{M}\sum_{t=1}^{T}\left(x_{i,t,a_{i,t}^*}^\intercal \theta^* - x_{i,t,a_{i,t}}^\intercal \theta^*\right)\right],$$

where $a_{i,t}^* \in \mathcal{A}$ is an optimal arm for client i given context $c_{i,t}$: $\forall b \ne a_{i,t}^*$, $x_{i,t,a_{i,t}^*}^\intercal \theta^* - x_{i,t,b}^\intercal \theta^* \ge 0$.

2.3. User-level Differential Privacy

In order to formally introduce user-level DP into the federated bandits framework, we consider a federated algorithm that consists of $2M+1$ sub-algorithms denoted as $\texttt{Alg} := (\texttt{R}_0, \texttt{Alg}_1, \texttt{R}_1, \ldots, \texttt{Alg}_M, \texttt{R}_M)$, where $\texttt{Alg}_i$ is an online decision-making algorithm adopted by client i, $\texttt{R}_i$ is a channel that shares the information from client i to the central server, and $\texttt{R}_0$ is a channel that broadcasts the aggregated information from the server to all clients.

Mathematically, let $H_{i,t}$ be the local historical observations and actions of client i before it makes decision at time t, i.e. $H_{i,t} = \{c_{i,\tau}, a_{i,\tau}, r_{i,\tau}\}_{\tau=1}^{t-1}$, and $q_{i,t}$ and q_t be the outputs of $\texttt{R}_i$ and $\texttt{R}_0$ at time t, respectively.

With a specified $\texttt{Alg}$, at each time step t, the learning procedure proceeds as follows. First, the server aggregates up-to-date local updates from the clients and broadcasts the aggregated information to all clients through channel $\texttt{R}_0$, i.e., $\texttt{R}_0 : \{q_{i,\le t}\}_i \mapsto q_t$. Upon receiving the broadcast information, each client performs local online decision-making by executing $\texttt{Alg}_i : \{c_{i,t}\} \cup H_{i,t} \cup \{q_{\le t}\} \mapsto a_{i,t}$, and obtains $r_{i,t}$. Finally, each client generates a local update and uploads it to the server through channel $\texttt{R}_i : (H_{i,t+1}, q_{\le t}) \mapsto q_{i,t+1}$.

We note that the general federated bandits learning framework can accommodate various bandits and communication models. For example, depending on whether communication happens in a step t or not, we can let $q_{i,t} = 0$ to indicate that client i does not upload any information at time t, and $q_t = 0$ if the server does not broadcast at time t.

Based on the specified federated bandits learning framework, we then introduce the user-level DP notions as follows. Let $H_t = \{H_{i,t}\}_{i\in[M]}$ be the entire history across all clients. Note that H_t is a streaming dataset, i.e. $H_{t'} \subseteq H_t$ for any $t' \le t$. Due to the online setting, we follow the definition of differential privacy *under continual observation (Dwork et al., 2010a)*. For simplicity, we denote $\texttt{R}_i(\{H_{i,\tau}, q_{\tau-1}\}_{\tau\le t}) := (\texttt{R}_i(H_{i,1}, q_0), \ldots, \texttt{R}_i(H_{i,t}, q_{\le t-1})), \forall i \in [M]$. With a little abuse of notation, we denote $\texttt{R}_0(\{H_\tau\}_{\tau\le t}) := (\texttt{R}_0(\{q_{i,1}\}_{i\in[M]}), \ldots, \texttt{R}_0(\{q_{i,\le t}\}_{i\in[M]}))$ to indicate the end-to-end relationship between the entire history $\{H_\tau\}_{\tau\le t}$ and the global information $\{q_1, \ldots, q_t\}$ produced by $\texttt{R}_0$. With-

out loss of generality, we use $\mathcal{Q}$ to denote the range of channel $\mathtt{R}_i$ for any $i \in [M] \cup \{0\}$.

Definition 2.1 (i-neighboring datasets)**.** We say $H_t = \{H_{j,t}\}_{j\in[M]}$ is i-neighboring to $H'_t = \{H'_{j,t}\}_{j\in[M]}$ if $H_{j,t} = H'_{j,t}$ for all $j \neq i$.

Definition 2.2 (User-level central DP)**.** Consider a time horizon T. A federated algorithm $\mathtt{Alg} = (\mathtt{R}_0, \mathtt{Alg}_1, \mathtt{R}_1, \ldots, \mathtt{Alg}_M, \mathtt{R}_M)$ is (ε, δ)-central user-level differentially private if for any i-neighboring streaming datasets $\{H_t\}_{t\leq T}$ and $\{H'_t\}_{t\leq T}$, and any subset $Q_{\leq T} := (Q_1, \ldots, Q_T) \subset \mathcal{Q}^T$, we have

$$\mathbb{P}[\mathtt{R}_0(\{H_t\}_{t\leq T}) \in Q_{\leq T}] \leq e^{\varepsilon}\mathbb{P}[\mathtt{R}_0(\{H'_t\}_{t\leq T}) \in Q_{\leq T}] + \delta.$$

Besides the user-level central DP and local DP (to be introduced in Section 5), we note that the proposed federated bandits framework can accommodate various DP notions, as elaborated in Appendix A.

3. Algorithm Design and Analysis for CDP

In this section, we aim to design a collaborative learning algorithm for the federated linear contextual bandits that achieves sublinear learning regret under user-level CDP.

3.1. Challenges of Adopting Gram Matrix in Algorithms

To gain a better understanding of our algorithm design, we first elaborate the difficulties encountered by prevalent Gram-Matrix (GM)-based approaches in federated linear contextual bandits when user-level CDP is taken into account. It is worth noting that all current privacy-preserving algorithms utilized in federated linear contextual bandits rely on GM-based approaches. Specifically, we focus on the prominent challenges associated with the widely adopted upper confidence bound (UCB) methods.

The fundamental task in the design of bandits algorithms is to balance the exploration-exploitation trade-off. UCB-type algorithms achieve this through constructing an upper confidence bound of the estimated reward of each arm, and then picking the arm with the highest UCB in each step t. In the linear contextual bandits setting, such UCB is usually in the form of $x^\top\hat{\theta} + \alpha\|x\|_{\bar{V}_t^{-1}}$, where x is the feature vector associated with individual arms and the incoming context, $\hat{\theta}$ is the estimated model parameter, α is a constant, and $\bar{V}_t := I_d + \sum_{\tau<t} x_\tau x_\tau^\intercal$ is the matrix defined by the encountered feature vectors that are used to estimate $\hat{\theta}$. Roughly speaking, $\bar{V}_t$ captures the uncertainty in the estimate $\hat{\theta}$, i.e., $\|\hat{\theta} - \theta^*\|_{\bar{V}_t} = \tilde{O}(\sqrt{d})$. By selecting the arm associated with the highest UCB, it ensures that the directions along which $\bar{V}_t$ has small eigenvalues can be sufficiently explored, thus reducing the uncertainty in $\hat{\theta}$ efficiently.

In the federated setting, each client i collects $\{(x_{i,\tau,a_\tau}, r_{i,\tau})\}_\tau$ locally and exchanges information with the server for expedited estimation of θ^*. Under the UCB-based framework, in order to characterize the estimation uncertainty and facilitate efficient exploration in subsequent time steps, it in general requires the server to collect and aggregate the Gram matrices $V_{i,t} := \sum_{\tau\in\mathcal{T}(t)} x_{i,t,a_{i,t}} x_{i,t,a_{i,t}}^\intercal$, where $\mathcal{T}(t)$ is a subset of $[t]$ determined by the specific algorithm design, and then broadcast the privatized aggregated Gram matrix $\tilde{V}_t^{-1}$ to all clients, along with the privatized estimate $\tilde{\theta}$.

According to Definition 2.2, in order to achieve *user-level* CDP, it is necessary to ensure that if $V_{i,t}$ is replaced by another Gram matrix $V'_{i,t}$, the broadcast information would remain similar. Standard additive noise mechanisms require that the variance of the additive noise in each dimension scales in $\sup \|V_{i,t} - V'_{i,t}\|$, which, without any additional assumption, scales in $\Omega(t)$ in general. If a noise with $\Omega(t)$ variance is added, the corresponding UCB, denoted as $x^\top\tilde{\theta} + \alpha\|x\|_{\tilde{V}_t^{-1}}$, would be largely different from its non-privatized counterpart, resulting in wrong arm-pulling at clients and linear regret.

3.2. Additional Assumptions

The aforementioned challenges in balancing user-level CDP and regret faced by UCB-type algorithms motivate us to strive for a possible approach where local Gram matrices are not required for collaborative parameter estimation in the federated linear contextual bandits setting. Towards that, we introduce the following two standard assumptions in the literature of linear contextual bandits.

Assumption 3.1 (Context diversity (Hao et al., 2020))**.** For any client i, let $a^*_{c_i} = \arg\max_a \phi(c_i, a)^\intercal\theta^*$ be the optimal arm under context c_i. Then,

$$\lambda_{\min}\left(\mathbb{E}_{c_i\sim\rho_i}\left[\phi(c_i, a^*_{c_i})\phi(c_i, a^*_{c_i})^\intercal\right]\right) \geq \lambda_0 > 0.$$

Assumption 3.1 essentially indicates that the minimum eigenvalue of the expected Gram matrix under the *optimal* policy is bounded away from zero. Thus, it ensures that with high probability, every possible direction in the parameter space can be adequately explored under the optimal policy for each client i. Intuitively, if the optimal policy were known beforehand, each client could collect sufficient information in each dimension of the parameter space and obtain an estimate of θ with favorable accuracy guarantee in each dimension. Therefore, such estimates can be directly aggregated at the server with certain accuracy guarantee even without knowing the corresponding Gram matrices. This could potentially avoid adding strong noises to the Gram matrices and degrading the regret performance. Besides, as we discuss in Appendix D, Assumption 3.1 is actually necessary in order to achieve sublinear regret for a broad class of algorithms in the federated linear contextual

bandits setting.

We remark that Assumption 3.1 does not guarantee that the distributions of local datasets collected by the client are identical or even close in terms of total variation distance, since the covariance matrices still depend on how the clients are making decisions. Even if they focus on the optimal arms, the matrix $\mathbb{E}[\phi(c_{i,t}, a^*_{c_{i,t}})\phi(c_{i,t}, a^*_{c_{i,t}})^\intercal]$ can still vary drastically across clients, as different clients can potentially have very different context distributions. This fact is in stark contrast to previous user-level DP works, e.g., Levy et al. (2021), and makes our problem more challenging.

Assumption 3.2 (Margin condition (Rigollet & Zeevi, 2010; Reeve et al., 2018))**.** Let $a^*_{c_i} = \arg\max_a \phi(c_i, a)^\intercal \theta^*$ be the optimal arm under context c_i. Then, there exists a constant C_0 such that, for any $\epsilon > 0$ and any $i \in [M]$,

$$\mathbb{P}_{c_i \sim \rho_i}\left[\forall a \neq a^*_{c_i},\, [\phi(c_i, a^*_{c_i}) - \phi(c_i, a)]^\intercal \theta^* \leq \epsilon\right] \leq C_0 \epsilon.$$

Roughly speaking, Assumption 3.2 ensures that for a randomly generated context, the expected reward under the corresponding optimal arm and that under any sub-optimal arm are statistically well separated. This is a standard assumption for *instance-dependent* analysis of linear contextual bandits. Similar to the minimum reward gap between the optimal arm and any sub-optimal arm in the stochastic MAB setting, C_0 controls the hardness of the problem: the larger C_0 is, the more challenging to distinguish the optimal arm and the second sub-optimal arm under a given context. In the following, we investigate upper bound *with* Assumption 3.2 and lower bounds *with* and *without* Assumption 3.2.

3.3. The ROBIN Algorithm

In this section, we propose a gReedy explOitation Based prIvatized averagiNg (ROBIN) Algorithm under the federated linear contextual bandits setting. Our objective is to leverage Assumption 3.1 and Assumption 3.2 to achieve sub-linear learning regret and guarantee user-level CDP at the same time.

ROBIN works in phases. In total, it has P phases, and each phse $p \in [P]$ contains 2^p time indices. Denote $\mathcal{T}_p$ as the set of time indices in phase p. Then, it proceeds as follows.

LinUCB-based Initialization: In the first U phases, each client performs the classical LinUCB algorithm (Abbasi-Yadkori et al., 2011) locally. Specifically, at the beginning, client i initializes an all-zero matrix $V_{i,1}$ and an all-zero vector $Y_{i,1}$. Then, at each time t in phase $p \in U$, upon observing a context $c_{i,t}$ and the corresponding decision set $\mathcal{D}_{i,t} = \{\phi(c_{i,t}, a) : a \in \mathcal{A}\}$, the client i chooses action $x_{i,t,a_{i,t}}$ according to

$$x_{i,t,a_{i,t}} = \arg\max_{x \in \mathcal{D}_{i,t}} x^\intercal \hat{\theta}_{i,t} + \alpha \|x\|_{(I_d + V_{i,t})^{-1}}, \qquad (1)$$

where $\hat{\theta}_{i,t} = (I_d + V_{i,t})^{-1} Y_{i,t}$, and α is a parameter to be specified later. After receiving a reward $r_{i,t}$, each client updates matrix $V_{i,t}$ and vector $Y_{i,t}$ according to

$$\begin{cases} V_{i,t+1} = V_{i,t} + x_{i,t,a_{i,t}} x_{i,t,a_{i,t}}^\intercal, \\ Y_{i,t+1} = Y_{i,t} + x_{i,t,a_{i,t}} r_{i,t}. \end{cases} \qquad (2)$$

The benefit of such initialization is that the minimum eigenvalue of $V_{i,t}$ can be guaranteed to grow linearly in t.

Local Greedy Exploitation: For any phase $p > U$ after the initialization phase, each client receives a private global estimate $\hat{\theta}^p$ from the central server at the beginning of phase p. We will specify how to construct such a global estimate in the next several paragraphs. Meanwhile, each client resets matrix $V_{i,t}$ and vector $Y_{i,t}$ to be zero. Then, for all $t \in \mathcal{T}_p$, each client greedily takes actions with respect the global estimate, i.e. $a_{i,t} = \arg\max_a \phi(c_{i,t}, a)^\intercal \hat{\theta}^p$, and collects a reward $r_{i,t}$. Again, $V_{i,t}$ and $Y_{i,t}$ are updated according to Equation (2).

Upload Channel $\mathtt{R}_i$: At the end of each phase p such that $p \geq U$, each client constructs a local estimator $\tilde{\theta}_{i,p} = \tilde{V}_{i,p}^\dagger \tilde{Y}_{i,p}$ based on the ordinary least squares method, where $\tilde{V}_{i,p}$ and $\tilde{Y}_{i,p}$ are copies of statistics $V_{i,t}$ and $Y_{i,t}$ at the end of phase p. These local estimators are then sent to the central server for privatized aggregation.

Privatized Aggregation $\mathtt{R}_0$ at the Server: Once the local estimates $\{\tilde{\theta}_{i,p}\}_{i \in [M]}$ are received at the end of phase p, the server needs to aggregate those local estimates and obtain a global estimate $\hat{\theta}^{p+1}$, which will be broadcast to the clients to facilitate their decision-making in phase $p+1$. There are two objectives for this aggregation step: One one hand, due to the *user-level CDP* constraint, the corresponding aggregation needs to ensure that $\hat{\theta}^{p+1}$ does not change significantly if $\tilde{\theta}_{i,p}$ is replaced by another possible local estimate $\tilde{\theta}'_{i,p}$ for any $i \in [M]$. On the other hand, in order to achieve low learning regret, it requires that $\hat{\theta}^{p+1}$ is sufficiently close to the ground truth parameter θ^*. In particular, it is desirable to have $\|\hat{\theta}^{p+1} - \theta^*\|$ scales in $\tilde{O}\left(1/\sqrt{M|\mathcal{T}_p|}\right)$ with high probability as p increase.

However, in general, the sensitivity of $\frac{1}{M}\sum_i \tilde{\theta}_{i,p}$ scales in $O(1/M)$. This is because $\tilde{\theta}_{i,p}$ is a random variable, whose distribution expands the entire parameter space $\{\theta : \|\theta\| \leq 1\}$. Thus, if a vanilla additive noise mechanism is adopted, the variance of the noise should scale in $\Omega(1/M)$. As a result, $\|\hat{\theta}^{p+1} - \theta^*\|$ scales in $\Omega(1/\sqrt{M})$, which cannot provide the desired estimation accuracy in order to achieve sublinear regret.

To overcome this challenge, we aim to leverage the concentration property of $\tilde{\theta}_{i,p}$ to have a more delicate analysis of the sensitivity of $\frac{1}{M}\sum_i \tilde{\theta}_{i,p}$. The intuition is, under assump-

Algorithm 1 The ROBIN Algorithm

Input: P, β, c_1, $\delta_0 = \delta/(2P)$, $\varepsilon_0 = \varepsilon/\sqrt{6P\log(2/\delta)}$.
while not reaching the time horizon T **do**
  if $p \le U$ **then**
    ▷ Initialization phases
    $D_{i,p} = \emptyset$
    for each client i and t in Phase p **do**
      Observe $c_{i,t}$
      Choose $x_{i,t,a_{i,t}}$ according to Eqn. (1), and receive $r_{i,t}$
      Update $V_{i,t}$ and $Y_{i,t}$ according to Eqn. (2)
    end for
  else
    ▷ Greedy exploitation
    for each client i **do**
      Receive $\hat{\theta}^{(p)}$ from the server
      Set $V_{i,t} = 0, Y_{i,t} = 0$
      for $t \in$ Phase p **do**
        Receive decision set $\mathcal{D}_{i,t} = \{x_{i,t,a}\}_{a\in\mathcal{A}}$
        Pull arm $x_{i,t,a_t} = \arg\max_{x\in\mathcal{D}_{i,t}} x^\intercal \hat{\theta}^{(p)}$
        Receive reward $r_{i,t}$
        Update $V_{i,t}$ and $Y_{i,t}$ according to Eqn. (2)
      end for
    end for
  end if
  if $p \ge U$ **then**
    for each client i **do**
      ▷ Upload channel: $\mathtt{R}_i$
      $\tilde{\theta}_{i,p} \leftarrow \tilde{V}_{i,p}^{\dagger}\tilde{Y}_{i,p}$
      Send $\tilde{\theta}_{i,p}$ to the server
    end for
    ▷ Privatized aggregation: $\mathtt{R}_0$
    $\hat{\theta}^{p+1} = \mathtt{WMHD}\left(\{\tilde{\theta}_{i,p}\}_{i\in[M]}, \frac{c_1}{\sqrt{|\mathcal{T}_p|}}, \frac{\beta}{16P}, \varepsilon_0, \delta_0\right)$
  end if
  $p \leftarrow p+1$
end while

tion 3.1, $\tilde{\theta}_{i,p+1}$ will be concentrated in a ball centered at θ^* with radius $\tilde{O}(1/\sqrt{|\mathcal{T}_p|})$ with high probability. Therefore, if we are able to leverage such concentration property and reduce the sensitivity of $\frac{1}{M}\sum_i \tilde{\theta}_{i,p}$, we will be able to adaptively reduce the variance of the additive noise in the DP mechanism and achieve sublinear regret.

Motivated by this intuition, we adopt the WinsorizedMean-HighD (WMHD) Algorithm from Levy et al. (2021) for the private aggregation at the server. Roughly speaking, WMHD projects local estimates $\{\tilde{\theta}_{i,p}\}_{i\in[M]}$ into a privatized range in the parameter space and then adds noise accordingly. By properly choosing the size of the range and the noise level, it outputs a privatized average of $\{\tilde{\theta}_{i,p}\}_{i\in[M]}$ with sufficient estimation accuracy and CDP guarantee. The details of WMHD can be found in Algorithm 4 in Appendix B.

The ROBIN algorithm is summarized in Algorithm 1.

3.4. Regret Analysis

Theorem 3.3. *Fix* $\beta \in (0,1)$ *and let* $c_1 = \frac{4\sqrt{2d\log(16d(M+1)P/\beta)}}{\lambda_0}$. *Then, under Assumptions 3.1 and 3.2, when* $\varepsilon \ge \Omega(\frac{\sqrt{d}\log^{1.5}T}{M})$, *Algorithm 1 (i) satisfies user-level* (ε,δ)*-CDP, and (ii) with probability at least* $1-\beta$, *achieves a regret upper bounded by*

$$\tilde{O}\left(\max\left(1, \frac{d\log T\log(\frac{1}{\delta})\log^3(\frac{1}{\beta})}{M\varepsilon^2}\right)\frac{C_0 d\log(\frac{1}{\beta})\log T}{\lambda_0^2}\right).$$

Proof of the user-level CDP guarantee: Since $\hat{\theta}^p$ is a (ε_0,δ_0)-differentially private estimation (Theorem 2 in Levy et al. (2021)), and there are total P phases, by the advanced composition rule (Lemma G.1), we conclude that the entire algorithm achieves user-level $(\varepsilon_0\sqrt{6P\log(1/\delta')}, \delta'+\delta_0 P)$-CDP for any $\delta' > 0$. The proof is finished by noting that $\delta_0 = \delta/(2P)$, $\varepsilon_0 = \varepsilon/\sqrt{6P\log(2/\delta)}$ and choosing $\delta' = \delta/2$.

Proof sketch of the regret upper bound. To upper bound the regret, we need to characterize the estimation error of the global estimator $\hat{\theta}^p$ for $p > U$. We inductively show that the following two claims hold with high probability.

- Claim 1: $\lambda_{\min}(\tilde{V}_{i,p}) \ge \Omega(|\mathcal{T}_p|)$.
- Claim 2: $\|\hat{\theta}^{p+1} - \theta^*\| \le \tilde{O}\left(\frac{1}{\varepsilon M\sqrt{|\mathcal{T}_p|}} + \frac{1}{\sqrt{M|\mathcal{T}_p|}}\right)$.

If Claim 1 holds, it is straightforward to conclude that each local estimator is sufficiently accurate, i.e. $\|\tilde{\theta}_{i,p} - \theta^*\| \le \tilde{O}\left(1/\sqrt{|\mathcal{T}_p|}\right)$. Then, due to Theorem 2 in Levy et al. (2021), WMHD guarantees that $\hat{\theta}^{p+1}$ is "close" to the average of local estimators $\{\tilde{\theta}_{i,p}\}_{i\in[M]}$. Specifically, we can show that with high probability,

$$\left\|\hat{\theta}^{p+1} - \tfrac{1}{M}\textstyle\sum_{i\in[M]}\tilde{\theta}_{i,p}\right\| \le \tilde{O}\left(\frac{1}{\varepsilon M\sqrt{|\mathcal{T}_p|}}\right).$$

Thus, Claim 2 follows by applying Claim 1 on the average estimator, i.e. $\|\frac{1}{M}\sum_i \tilde{\theta}_{i,p} - \theta^*\| \le \tilde{O}\left(1/\sqrt{M|\mathcal{T}_p|}\right)$.

If Claim 2 holds, combining with Assumption 3.2, with high probability, the covariance matrix $\sum_{t\in\mathcal{T}_p} x_{i,t,a_{i,t}} x_{i,t,a_{i,t}}^\intercal$ is close to $|\mathcal{T}_p|\mathbb{E}_{c_i\sim\rho_i}\left[\phi(c_i,a_{c_i}^*)\phi(c_i,a_{c_i}^*)^\intercal\right]$. Hence, Claim 1 follows directly from Assumption 3.1.

Finally, the LinUCB algorithm in the first U phases guarantees that Claim 1 holds with high probability, and by induction, both claims hold for all phases $p \ge U$. Therefore, the regret can be upper bounded by the product of estimation error and the probability of playing sub-optimal arms.

Due to Assumption 3.2, the probability is again controlled by the estimation error, and thus the regret upper bound is $\tilde{O}\left((1+1/(M\varepsilon^2))\log T\right)$. □

The full proof can be found in Appendix C.

Remark 3.4. In general, the difficulty of deriving regret upper bounds without Assumption 3.2 is due to the weak diversity Assumption 3.1, where we only require the covariance matrix associated with *the optimal arm* to be sufficiently diverse. Therefore, without Assumption 3.2, if the sub-optimal gap is too small, it is highly likely that the feature vector associated with the decision of the agent is very different from the optimal feature vector, although the corresponding reward (inner product of the feature vector and θ) is close to the optimal reward. If this is the case, then the agent cannot leverage Assumption 3.1 to construct an accurate estimation of θ, as Assumption 3.1 only applies to the feature vectors associated with the optimal arm.

4. Lower Bounds under CDP Constraint

In this section, we present regret lower bounds for any *user-level central differentially private* federated algorithms. All proofs in this section can be found in Appendix E..

To be more precise, we separate the federated algorithms into two categories: almost-memoryless algorithms and with-memory algorithms. Without additional assumptions, the general framework defined in Section 2.3 is a with-memory algorithm, where the decision-making algorithm $\texttt{Alg}_i$ depends on both the local history and the global information $q_{\le t}$. If the available information is restricted, we define almost-memoryless algorithm as follows.

Definition 4.1 (Almost-memoryless algorithm)**.** A federated algorithm $(\texttt{R}_0, \texttt{Alg}_1, \texttt{R}_1, \ldots, \texttt{R}_M)$ is almost **memoryless** if there exists a constant $u = o(T)$ such that for any time step $t \ge u$, the decision-making does not depend on local history data, i.e. $\texttt{Alg}_i(H_{i,t}, q_{\le t}) = \texttt{Alg}_i(q_{\le t})$, for any $i \in [M]$ and $H_{i,t}$.

We note that a typical memoryless algorithm is phased elimination type of algorithms (Shi & Shen, 2021; Shi et al., 2021; Huang et al., 2021), where the policy in a single phase does not change and only depends on the information broadcast from the last communication round. Moreover, Algorithm 1 proposed in this work is also almost-memoryless.

Theorem 4.2. *If $\varepsilon < \log 2$, $\delta = \tilde{O}(\frac{1}{M\sqrt{T}})$, then, there exists a federated linear contextual bandits instance satisfying Assumptions 3.1 and 3.2, such that any* **almost-memoryless** *federated algorithm satisfying user-level (ε, δ)-CDP must incur a regret lower bounded by*

$$\Omega\left(\max\left\{1, \tfrac{1}{M\varepsilon^2}\right\} C_0 d \log T + e^{-M\varepsilon} C_0 MT\right).$$

Remark 4.3. First, we note that in the regime $\varepsilon = \Omega\left(\frac{\sqrt{d}\log^{1.5} T}{M}\right)$ considered in Theorem 3.3, the first term of the lower bound dominates the other. Thus, the regret upper bound under ROBIN nearly matches with the lower bound, indicating that ROBIN is near-optimal in terms of M and ε in this regime. On the other hand, when $\varepsilon = \tilde{O}\left(\frac{\log(MT)}{M}\right)$, the second term $e^{-M\varepsilon}C_0 MT$ dominates the first term and grows linearly in T. Thus, it is impossible to achieve sublinear regret for any almost-memoryless algorithm in this regime. Since ROBIN is an almost-memoryless algorithm, it arguably achieves the best we could hope for among all almost-memoryless algorithms.

The lower bound also indicates the number of samples contributed by each client (i.e., T) cannot be arbitrarily large (i.e., cannot scale faster than $\tilde{O}(e^{M\varepsilon}/M)$) in order to achieve sublinear regret for any almost-memoryless algorithm. A similar phenomenon in offline supervised learning is also observed in Levy et al. (2021), which states that learning with user-level DP cannot reach zero error when the number of clients is fixed.

The proof of Theorem 4.4 is built upon a generic lower bound developed by He et al. (2022b), as informally stated in Theorem 4.4, and the fingerprinting lemma (Kamath et al., 2019; Bun et al., 2017).

Theorem 4.4 (Informal)**.** *Let θ_1^* be uniformly sampled from a two-dimensional sphere $\Theta = \{x \in \mathbb{R}^2 : \|x\| = r\}$. Then, there exists a federated linear contextual bandits model such that the total regret is lower bounded by*

$$\Omega\left(\textstyle\sum_{i\in[M],t\in[T]} \inf_{\theta_{i,t}\in\mathcal{F}(\mathcal{I}_i,\Theta)} \frac{1}{r}\mathbb{E}_v\left[\|\theta_1^* - \theta_{i,t}\|^2\right]\right),$$

where $\mathcal{I}_i$ is a set of available information provided for client i (e.g., $(H_{i,t}, q_{\le t})$).

Proof sketch of Theorem 4.2. Theorem 4.4 states that the regret is lower bounded by the performance of estimating the "direction" of the true parameter θ^*. Note that in general, the error of estimating direction cannot be directly lower bounded by the standard estimation error, especially when the norm of parameters are $\Omega(1)$. However, under the margin condition in Assumption 3.2, the norm of θ^* is in general $O(1)$. Hence, to lower bound the estimation error of the direction, we follow an alternative approach that re-parameterizes θ_s^* by its angle γ_s^* and provide an upper bound of the expected inner product $\mathbb{E}_v[\theta_{i,t,s}^\intercal \theta_s^*]$. By leveraging the fingerprinting lemma, we show that if $\theta_{i,t,s}$ is a private estimator, then, $\mathbb{E}_v[\theta_{i,t,s}^\intercal \theta_s^*] \le O(\varepsilon M\sqrt{(t-1)\mathbb{E}_v[\|\theta_s^* - \theta_{i,t,s}\|^2]})$. Combining with the fact that $\mathbb{E}_v[\|\theta_s^* - \theta_{i,t,s}\|^2] = 2 - 2\mathbb{E}_v[\theta_{i,t,s}^\intercal \theta_s^*]$, we conclude that $\mathbb{E}_v[\|\theta_s^* - \theta_{i,t,s}\|^2] \ge \Omega(1/(\varepsilon^2 M^2(t-1)+1))$. Taking the summation over t and i, and by Theorem 4.4, we

obtain a lower bound $\Omega(\log T/(M\varepsilon^2))$. The second term $e^{-M\varepsilon}C_0MT$ can be derived by calculating the estimation error under the case when all clients collect dummy information (e.g. $x_{i,t,a_t} = 0$). The final result then follows by taking the non-private regret lower bound $\Omega(\log T)$ into consideration. □

Our results can be generalized to obtain minimax lower bounds for the with-memory algorithms by optimizing the norm r of the true parameter or separately analyze the information contained in local datasets. The results are summarized as follows.

Theorem 4.5. *Fix any $\varepsilon \in (0, \log 2)$, $\delta = \tilde{O}\left(\frac{1}{M\sqrt{T}}\right)$, $T \geq d^2$. Then, there exists a federated linear contextual bandits instance satisfying Assumptions 3.1 and 3.2 such that any* **with-memory** *federated algorithm satisfying user-level (ε, δ)-CDP must incur a regret lower bounded by*

$$\Omega\left(\min\left\{M, \max\left\{1, \frac{1}{M\varepsilon^2}\right\}\right\} C_0 d \log T\right).$$

If Assumption 3.2 is not satisfied, then the minimax regret lower bound becomes

$$\Omega\left(\min\left\{M, \max\left\{\sqrt{M}, \tfrac{1}{\varepsilon}\right\}\right\}\sqrt{dT}\right).$$

Remark 4.6. Since with-memory algorithms encompass almost-memoryless algorithms as special cases, the first part of Theorem 4.5 verifies that ROBIN is nearly optimal compared to *any* algorithms when $\varepsilon = \Omega\left(\frac{\sqrt{d}\log^{1.5}T}{M}\right)$. In fact, ROBIN can be easily adapted to a with-memory algorithm to remove the constraint on the parameters. Specifically, if at the beginning of the entire algorithm, ROBIN is allowed to adaptively decides that clients follow Algorithm 1 when $\varepsilon = \Omega(\frac{\sqrt{d}\log^{1.5}T}{M})$, or independently adopt LinUCB without sharing any information when $\varepsilon = O(\frac{\sqrt{d}\log^{1.5}T}{M})$, then, the algorithm achieves an upper bound that nearly matches with the lower bound in terms of M and ε for any $\varepsilon \in (0, \log 2)$. In addition, this adaptation preserves the user-level CDP guarantee, since performing LinUCB locally without information exchange guarantees zero information leakage.

Theorem 4.5 also indicates that, when $\varepsilon = \Omega(\frac{1}{\sqrt{M}})$, the lower bound reduces to $\Omega(C_0 d \log T)$ with Assumption 3.2 and $\Omega(\sqrt{dMT})$ without Assumption 3.2. This suggests that imposing user-level (ε, δ)-CDP constraint does not increase the hardness of learning compared with its non-private counterpart for the general with-memory algorithms in this regime. However, when $\varepsilon = O(\frac{1}{\sqrt{M}})$, imposing the CDP constraint incurs a blow-up factor at least $\min\{M, \frac{1}{\varepsilon^2 M}\}$ with Assumption 3.1 or $\min\{\sqrt{M}, \frac{1}{\varepsilon\sqrt{M}}\}$ without Assumption 3.1, indicating the hardness of learning strictly increases.

5. Lower Bounds under LDP Constraint

In this section, we present regret lower bounds under the non-interactive local differential privacy constraint, which provides an initial view of this more challenging problem. The full proofs in this section can be found in Appendix F.

Definition 5.1 (Non-interactive upload channel)**.** The upload channels $\mathtt{R}_i$ of a federated algorithm $\mathtt{Alg} = (\mathtt{R}_0, \mathtt{Alg}_1, \mathtt{R}_1, \ldots, \mathtt{Alg}_M, \mathtt{R}_M)$ is non-interactive if $\mathtt{R}_i$ does not depend on global information $q_{\leq t-1}$ conditioned on the local history $H_{i,t}$, i.e. $\mathtt{R}_i(H_{i,t}, q_{\leq t-1}) = \mathtt{R}_i(H_{i,t})$.

We note that existing phased elimination-based federated bandits algorithms are non-interactive (Shi & Shen, 2021; Shi et al., 2021; Huang et al., 2021). In contrary, UCB-type algorithms are interactive and with-memory (Dubey & Pentland, 2020; Li & Wang, 2022b; Li et al., 2020).

Definition 5.2 (Non-interactive user-level local DP)**.** Consider a time horizon T. A federated algorithm $\mathtt{Alg} = (\mathtt{R}_0, \mathtt{Alg}_1, \mathtt{R}_1, \ldots, \mathtt{Alg}_M, \mathtt{R}_M)$ is user-level (ε, δ)-locally differentially private if for any i-neighboring streaming datasets $\{H_t\}_{t\leq T}$ and $\{H'_t\}_{t\leq T}$ (see Definition 2.1), and any subset $Q_{i,\leq T} = (Q_{i,1}, \ldots, Q_{i,T}) \subset \mathcal{Q}^T$, we have

$$\begin{aligned}&\mathbb{P}[\mathtt{R}_i(\{H_{i,t}\}_{t\leq T}) \in Q_{i,\leq T}] \\ &\quad\leq e^{\varepsilon}\mathbb{P}[\mathtt{R}_i(\{H'_{i,t}\}_{t\leq T}) \in Q_{i,\leq T}] + \delta.\end{aligned}$$

We focus on the with-memory setting, since any lower bound of with-memory algorithms must be a lower bound of memoryless algorithms.

Theorem 5.3. *If $\varepsilon \in (0, \log 2)$, $\delta = \tilde{O}(1/M\sqrt{T})$, there exists a federated linear contextual bandits instance satisfying Assumptions 3.1 and 3.2 such that any* **with-memory** *federated algorithm satisfying user-level (ε, δ)-LDP must incur a regret lower bounded by*

$$\Omega\left(\min\left\{1/\varepsilon, M\right\} C_0 d \log T\right).$$

If Assumption 3.2 is not satisfied, then the minimax regret lower bound becomes

$$\Omega\left(\min\left\{\sqrt{M/\varepsilon}, M\right\}\sqrt{dT}\right).$$

Remark 5.4. We note that when $\varepsilon = O(1/M)$, the regret lower bound is either $\Omega(M \log T)$ under Assumption 3.2 or $\Omega(M\sqrt{T})$ without Assumption 3.2, suggesting that the best policy is to have clients independently make arm-pulling decisions without information exchange. When $\varepsilon = \Omega(1/M)$, the lower bound becomes $\Omega(C_0 d \log T/\varepsilon)$ under Assumption 3.2, or $\Omega(\sqrt{dMT/\varepsilon})$ without Assumption 3.2. This indicates that under the (ε, δ)-LDP constraint, as long as $\varepsilon \in (0, \log 2)$, the regret of any federated algorithm must suffer a blow-up factor at least $\min\{1/\sqrt{\varepsilon}, \sqrt{M}\}$ without Assumption 3.2, or $\min\{1/\varepsilon, M\}$ with Assumption 3.2, compared with the optimal regrets in the non-private setting.

For completeness, we also investigate the regret lower bound under user-level pure LDP constraint, i.e. $\delta = 0$.

Corollary 5.5. *For any $\varepsilon \in (0, \log 2)$, there exists a federated linear contextual bandits instance satisfying Assumptions 3.1 and 3.2 such that any* **with-memory** *federated algorithm satisfying ε-LDP must incur a regret lower bounded by*

$$\Omega\left(\min\left\{M, 1/\varepsilon^2\right\} C_0 d \log T\right).$$

If Assumption 3.2 is not satisfied, then the minimax regret lower bound becomes

$$\Omega\left(\min\left\{M, \sqrt{M}/\varepsilon\right\} \sqrt{dT}\right).$$

Compared with the results in Theorem 4.5, we note that the regret lower bound under pure LDP constraint is generally higher than that under CDP constraint. Specifically, the results in Corollary 5.5 can be obtained by replacing ε in Theorem 4.5 by $\varepsilon/\sqrt{M}$[1].

6. Related Work

Differential Privacy in Bandits. There is a line of research focusing on differentially private multi-armed bandits (DP-MAB). Mishra & Thakurta (2015) first introduce the problem of DP-MAB with algorithms that achieve sublinear regret. Later, Tossou & Dimitrakakis (2016); Sajed & Sheffet (2019); Azize & Basu (2022) improve the analysis and propose several different algorithms that enjoy the optimal regret. In addition, Hu & Hegde (2022) achieve similar near-optimal regret based on a Thompson-sampling based approach, Tao et al. (2022) consider heavy-tailed rewards case, and Chowdhury & Zhou (2022a) provide an optimal regret in distributed DP-MAB. Local DP constraint is also studied by Ren et al. (2020) in MAB and by Zheng et al. (2020) in both MAB and linear contextual bandits. Shariff & Sheffet (2018) propose LinUCB with changing perturbation to satisfy jointly differential privacy. Later, Wang et al. (2020a) consider pure DP in both global and local setting. Wang et al. (2022b) propose an algorithm with dynamic global sensitivity. Other models including linear and generalized linear bandits under DP constraints are studied by Hanna et al. (2022) and Han et al. (2021). The shuffle model has also been addressed by Chen et al. (2020); Chowdhury & Zhou (2022b); Garcelon et al. (2022); Tenenbaum et al. (2021).

Federated Bandits. There is a growing body of research studying item-level DP based local data privacy protection in federated bandits. Li et al. (2020); Zhu et al. (2021) study federated bandits with DP guarantee. Dubey & Pentland (2022) consider private and byzantine-proof cooperative decision making in multi-armed bandits. Dubey & Pentland (2020); Zhou & Chowdhury (2023) consider the linear contextual bandit model with joint DP guarantee. Li et al. (2022a) study private distributed bandits with partial feedback.

Federated bandits without explicit DP constraints have also been studied by Wang et al. (2020b); Li & Wang (2022a); Shi & Shen (2021); Shi et al. (2021); Huang et al. (2021); Wang et al. (2022a); He et al. (2022a); Li et al. (2022b).

User-level DP. First introduced by Dwork et al. (2010b), user-level DP has attracted increased attention recently (McMahan et al., 2017b; Wang et al., 2019). Liu et al. (2020); Acharya et al. (2022) study discrete distribution estimation under user-level DP. Ghazi et al. (2021) investigate the number of users required for learning under user-level DP constraint. Amin et al. (2019); Epasto et al. (2020) study approaches to bound individual users' contributions in order to achieve good trade-off between utility and user-level privacy guarantee. Levy et al. (2021) consider various learning tasks, such as mean estimation, under user-level DP constraint. To the best of our knowledge, user-level DP has not been studied in the online learning setting before.

7. Conclusions

In this paper, we investigated federated linear contextual bandits under user-level differential privacy constraints. We first introduced a general federated sequential decision-making framework that can accommodate various notations of DP in the bandits setting. We then proposed an algorithm termed as `ROBIN` and showed that it is near-optimal under the user-level CDP constraint. We further provided various lower bounds for federated algorithms under user-level LDP constraint, which imply that learning under LDP constraint is strictly harder than the non-private case. Designing federated algorithms to approach such lower bounds under LDP constraint would be an interesting direction to pursue in the future.

Acknowledgments

The authors would like to thank Ilya Mironov and Abhimanyu Dubey for the helpful and inspiring discussions.

The work of R. Huang and J. Yang was supported in part by the U.S. National Science Foundation under grants 2030026, 2114542, 2133170 and a gift from Meta.

The work of Meisam Hejazinia was performed when he was at Meta.

[1] Another lens to see this phenomenon is the privacy amplification by shuffling (Feldman et al., 2022). Loosely speaking, an $(\varepsilon/\sqrt{M}, \delta)$-CDP lower bound leads to an (ε, δ)-LDP lower bound, when ε is sufficiently small (Acharya et al., 2022).

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Improved algorithms for linear stochastic bandits. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, pp. 2312–2320, 2011.

Acharya, J., Sun, Z., and Zhang, H. Differentially private Assouad, Fano, and Le Cam. In *Algorithmic Learning Theory*, pp. 48–78. PMLR, 2021.

Acharya, J., Liu, Y., and Sun, Z. Discrete distribution estimation under user-level local differential privacy. *arXiv preprint arXiv:2211.03757*, 2022.

Agrawal, S. and Goyal, N. Analysis of Thompson sampling for the multi-armed bandit problem. In *Conference on Learning Theory*, pp. 39–1, 2012.

Agrawal, S. and Goyal, N. Further optimal regret bounds for Thompson sampling. In *Artificial Intelligence and Statistics*, pp. 99–107, 2013.

Amin, K., Kulesza, A., Munoz, A., and Vassilvtiskii, S. Bounding user contributions: A bias-variance trade-off in differential privacy. In *International Conference on Machine Learning*, pp. 263–271. PMLR, 2019.

Asoodeh, S., Aliakbarpour, M., and Calmon, F. P. Local differential privacy is equivalent to contraction of e_γ-divergence. *arXiv preprint arXiv:2102.01258*, 2021.

Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

Azize, A. and Basu, D. When privacy meets partial information: A refined analysis of differentially private bandits. *arXiv preprint arXiv:2209.02570*, 2022.

Bubeck, S. and Cesa-Bianchi, N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

Bun, M., Steinke, T., and Ullman, J. Make up your mind: The price of online queries in differential privacy. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1306–1325. SIAM, 2017.

Carpentier, A., Vernade, C., and Abbasi-Yadkori, Y. The elliptical potential lemma revisited. *arXiv preprint arXiv:2010.10182*, 2020.

Chen, X., Zheng, K., Zhou, Z., Yang, Y., Chen, W., and Wang, L. (Locally) differentially private combinatorial semi-bandits. In *International Conference on Machine Learning*, pp. 1757–1767. PMLR, 2020.

Chowdhury, S. R. and Zhou, X. Distributed differential privacy in multi-armed bandits. *arXiv preprint arXiv:2206.05772*, 2022a.

Chowdhury, S. R. and Zhou, X. Shuffle private linear contextual bandits. *arXiv preprint arXiv:2202.05567*, 2022b.

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

Cummings, R., Feldman, V., McMillan, A., and Talwar, K. Mean estimation with user-level privacy under data heterogeneity. In *Advances in Neural Information Processing Systems*, 2021.

Den Hollander, F. Probability theory: The coupling method. *Lecture notes available online (http://websites. math. leidenuniv. nl/probability/lecturenotes/CouplingLectures. pdf)*, 2012.

Dubey, A. and Pentland, A. Differentially-private federated linear bandits. *Advances in Neural Information Processing Systems*, 33:6003–6014, 2020.

Dubey, A. and Pentland, A. Private and Byzantine-proof cooperative decision-making. *arXiv preprint arXiv:2205.14174*, 2022.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. Local privacy, data processing inequalities, and minimax rates. *arXiv preprint arXiv:1302.3203*, 2013.

Dwork, C., Naor, M., Pitassi, T., and Rothblum, G. N. Differential privacy under continual observation. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pp. 715–724, 2010a.

Dwork, C., Naor, M., Pitassi, T., Rothblum, G. N., and Yekhanin, S. Pan-private streaming algorithms. In *ics*, pp. 66–80, 2010b.

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

Epasto, A., Mahdian, M., Mao, J., Mirrokni, V., and Ren, L. Smoothly bounding user contributions in differential privacy. *Advances in Neural Information Processing Systems*, 33:13999–14010, 2020.

Feldman, V. and Steinke, T. Generalization for adaptively-chosen estimators via stable median. In *Conference on Learning Theory*, pp. 728–757. PMLR, 2017.

Feldman, V., McMillan, A., and Talwar, K. Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 954–964. IEEE, 2022.

Garcelon, E., Chaudhuri, K., Perchet, V., and Pirotta, M. Privacy amplification via shuffling for linear contextual bandits. In *International Conference on Algorithmic Learning Theory*, pp. 381–407. PMLR, 2022.

Ghazi, B., Kumar, R., and Manurangsi, P. User-level differentially private learning via correlated sampling. *Advances in Neural Information Processing Systems*, 34: 20172–20184, 2021.

Girgis, A. M., Data, D., and Diggavi, S. Distributed user-level private mean estimation. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pp. 2196–2201. IEEE, 2022.

Han, Y., Liang, Z., Wang, Y., and Zhang, J. Generalized linear bandits with local differential privacy. *Advances in Neural Information Processing Systems*, 34:26511–26522, 2021.

Hanna, O. A., Girgis, A. M., Fragouli, C., and Diggavi, S. Differentially private stochastic linear bandits:(almost) for free. *arXiv preprint arXiv:2207.03445*, 2022.

Hao, B., Lattimore, T., and Szepesvari, C. Adaptive exploration in linear contextual bandit. In *International Conference on Artificial Intelligence and Statistics*, pp. 3536–3545. PMLR, 2020.

He, J., Wang, T., Min, Y., and Gu, Q. A simple and provably efficient algorithm for asynchronous federated contextual linear bandits. *arXiv preprint arXiv:2207.03106*, 2022a.

He, J., Zhang, J., and Zhang, R. A reduction from linear contextual bandit lower bounds to estimation lower bounds. In *International Conference on Machine Learning*, pp. 8660–8677. PMLR, 2022b.

Hu, B. and Hegde, N. Near-optimal Thompson sampling-based algorithms for differentially private stochastic bandits. In *Uncertainty in Artificial Intelligence*, pp. 844–852. PMLR, 2022.

Huang, R., Wu, W., Yang, J., and Shen, C. Federated linear contextual bandits. *Advances in Neural Information Processing Systems*, 34:27057–27068, 2021.

Kamath, G., Li, J., Singhal, V., and Ullman, J. Privately learning high-dimensional distributions. In *Conference on Learning Theory*, pp. 1853–1902. PMLR, 2019.

Lai, T. L. and Robbins, H. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1): 4–22, 1985.

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

Levy, D., Sun, Z., Amin, K., Kale, S., Kulesza, A., Mohri, M., and Suresh, A. T. Learning with user-level privacy. *Advances in Neural Information Processing Systems*, 34: 12466–12479, 2021.

Li, C. and Wang, H. Asynchronous upper confidence bound algorithms for federated linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 6529–6553. PMLR, 2022a.

Li, C. and Wang, H. Communication efficient federated learning for generalized linear bandits. *arXiv preprint arXiv:2202.01087*, 2022b.

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

Li, T., Song, L., and Fragouli, C. Federated recommendation system via differential privacy. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 2592–2597. IEEE, 2020.

Li, W., Song, Q., Honorio, J., and Lin, G. Federated x-armed bandit. *arXiv preprint arXiv:2205.15268*, 2022b.

Liu, Y., Suresh, A. T., Yu, F. X. X., Kumar, S., and Riley, M. Learning discrete distributions: user vs item-level privacy. *Advances in Neural Information Processing Systems*, 33: 20965–20976, 2020.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 1273–1282, Fort Lauderdale, FL, USA, Apr. 2017a.

McMahan, H. B., Ramage, D., Talwar, K., and Zhang, L. Learning differentially private recurrent language models. *arXiv preprint arXiv:1710.06963*, 2017b.

Mishra, N. and Thakurta, A. (Nearly) optimal differentially private stochastic multi-arm bandits. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pp. 592–601, 2015.

Papini, M., Tirinzoni, A., Restelli, M., Lazaric, A., and Pirotta, M. Leveraging good representations in linear contextual bandits. In *International Conference on Machine Learning*, pp. 8371–8380. PMLR, 2021.

Reeve, H., Mellor, J., and Brown, G. The k-nearest neighbour UCB algorithm for multi-armed bandits with covariates. In *Algorithmic Learning Theory*, pp. 725–752. PMLR, 2018.

Ren, W., Zhou, X., Liu, J., and Shroff, N. B. Multi-armed bandits with local differential privacy. *arXiv preprint arXiv:2007.03121*, 2020.

Rigollet, P. and Zeevi, A. Nonparametric bandits with covariates. *arXiv preprint arXiv:1003.1630*, 2010.

Sajed, T. and Sheffet, O. An optimal private stochastic-mab algorithm based on optimal private stopping rule. In *International Conference on Machine Learning*, pp. 5579–5588. PMLR, 2019.

Shariff, R. and Sheffet, O. Differentially private contextual linear bandits. *Advances in Neural Information Processing Systems*, 31, 2018.

Shi, C. and Shen, C. Federated multi-armed bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 9603–9611, 2021.

Shi, C., Shen, C., and Yang, J. Federated multi-armed bandits with personalization. In *Proceedings of the 24rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, April 2021.

Tao, Y., Wu, Y., Zhao, P., and Wang, D. Optimal rates of (locally) differentially private heavy-tailed multi-armed bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 1546–1574. PMLR, 2022.

Tenenbaum, J., Kaplan, H., Mansour, Y., and Stemmer, U. Differentially private multi-armed bandits in the shuffle model. *Advances in Neural Information Processing Systems*, 34:24956–24967, 2021.

Tossou, A. C. and Dimitrakakis, C. Algorithms for differentially private multi-armed bandits. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

Tropp, J. Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16:262–270, 2011.

Wang, C.-H., Li, W., Cheng, G., and Lin, G. Federated online sparse decision making. *arXiv preprint arXiv:2202.13448*, 2022a.

Wang, H., Zhao, Q., Wu, Q., Chopra, S., Khaitan, A., and Wang, H. Global and local differential privacy for collaborative bandits. In *Fourteenth ACM Conference on Recommender Systems*, pp. 150–159, 2020a.

Wang, H., Zhao, D., and Wang, H. Dynamic global sensitivity for differentially private contextual bandits. In *Proceedings of the 16th ACM Conference on Recommender Systems*, pp. 179–187, 2022b.

Wang, Y., Hu, J., Chen, X., and Wang, L. Distributed bandit learning: Near-optimal regret with efficient communication. In *International Conference on Learning Representations*, 2020b.

Wang, Z., Song, M., Zhang, Z., Song, Y., Wang, Q., and Qi, H. Beyond inferring class representatives: User-level privacy leakage from federated learning. In *IEEE INFOCOM 2019-IEEE conference on computer communications*, pp. 2512–2520. IEEE, 2019.

Zheng, K., Cai, T., Huang, W., Li, Z., and Wang, L. Locally differentially private (contextual) bandits learning. *Advances in Neural Information Processing Systems*, 33: 12300–12310, 2020.

Zhou, X. and Chowdhury, S. R. On differentially private federated linear contextual bandits. *arXiv preprint arXiv:2302.13945*, 2023.

Zhu, Z., Zhu, J., Liu, J., and Liu, Y. Federated bandit: A gossiping approach. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 5(1):1–29, Feb 2021. ISSN 2476-1249. doi: 10.1145/3447380. URL http://dx.doi.org/10.1145/3447380.

A. Relationship with other DP Notions in Bandits

We use Figure 1 to demonstrate all existing DP channels in the bandits literature.

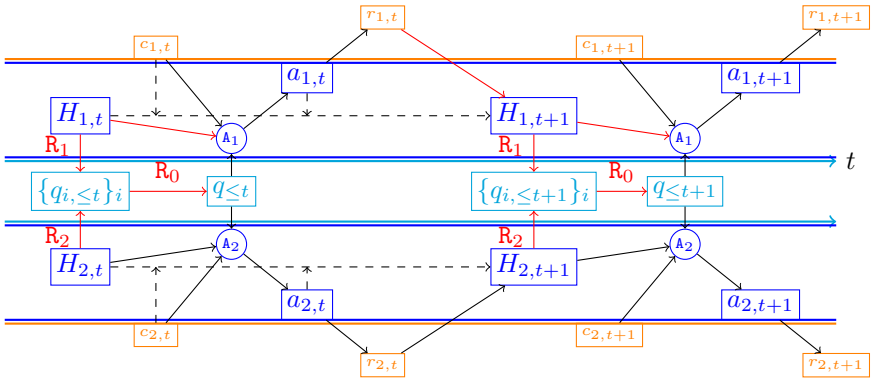

Figure 1: Graphical structure of all possible private channels in a bandits model. We assume that $\mathtt{Alg}_i$ produces some intermediate random variable $\mathtt{A}_i$ which determines the action.

Recall that $H_{i,t} = \{c_{i,\tau}, a_{i,\tau}, r_{i,\tau}\}_{\tau=1}^{t-1}$. We separately discuss different DP mechanisms in single-client and multi-client settings.

Single-client. In this case, the federated bandits framework defined in Section 2.3 reduces to the upper half of Figure 1.

- A single-client bandits algorithm is called **locally differentially private** (Ren et al., 2020; Zheng et al., 2020), if the red arrow from $r_{1,t}$ to the dataset $H_{1,t+1}$ is a DP channel. Mathematically, for any two rewards $r_{1,t}$ and $r'_{1,t}$, and any measurable set S containing $H_{1,t+1}$, we have

$$\mathbb{P}(H_{1,t+1} \in S | r_{1,t}) \le e^{\varepsilon}\mathbb{P}(H_{1,t+1} \in S | r'_{1,t}) + \delta.$$

- A single-client stochastic multi-armed bandits algorithm is called **globally differentially private** (Tossou & Dimitrakakis, 2016; Azize & Basu, 2022), if the red arrow from $H_{1,t}$ to $\mathtt{A}_1$ is a DP channel with respect to rewards. More precisely, for any t'-neighboring dataset $H_{1,t}$ and $H'_{1,t}$ satisfying $r_{1,t_0} = r'_{1,t_0}$ for all $t_0 \le t$ except at $t_0 = t'$, and any measurable set S containing $\mathtt{A}_1$, we have

$$\mathbb{P}(\mathtt{A}_1 \in S | H_{1,t}) \le e^{\varepsilon}\mathbb{P}(\mathtt{A}_1 \in S | H'_{1,t}) + \delta.$$

- A single-client linear contextual bandits algorithm is called **jointly differentially private** (Shariff & Sheffet, 2018), if the red arrow from $H_{1,t}$ to $\mathtt{A}_1$ is a DP channel with respect to one datum. More precisely, for any t'-neighboring dataset $H_{1,t}$ and $H'_{1,t}$ satisfying $(c_{i,t_0}, a_{i,t_0}, r_{1,t_0}) = (c'_{i,t_0}, a'_{i,t_0}, r'_{1,t_0})$ for all $t_0 \le t$ except at $t_0 = t'$, and any measurable set S containing $\mathtt{A}_1$, we have

$$\mathbb{P}(\mathtt{A}_1 \in S | H_{1,t}, c_{i,t}) \le e^{\varepsilon}\mathbb{P}(\mathtt{A}_1 \in S | H'_{1,t}, c_{i,t}) + \delta.$$

Multi-clients. In this case, we focus on the red arrows between clients and the server, i.e. $\mathtt{R}_i$ and $\mathtt{R}_0$.

- A multi-client bandits algorithm is called **item-level locally (jointly) differentially private** (Dubey & Pentland, 2020), if the red arrow $\mathtt{R}_i$ from $H_{i,t}$ to $q_{i,\le t}$ is a DP channel, and each client can fully rely on its own local dataset (with-memory setting). More precisely, if for any client i and t', $H_{i,t}$ and $H'_{i,t}$ only differ in one datum at time t', i.e. $(c_{i,t_0}, a_{i,t_0}, r_{i,t_0}) = (c'_{i,t_0}, a'_{i,t_0}, r'_{i,t_0})$ for all $t_0 \le t$ except at $t_0 = t'$, and for any measurable set containing $q_{i,t}$, we have

$$\mathbb{P}(q_{i,t} \in S | H_{i,t}) \le e^{\varepsilon}\mathbb{P}(q_{i,t} \in S | H'_{i,t}) + \delta.$$

We close this section by noting that there are many other types of DP that can be deduced from Figure 1. We list several key ingredients that form a DP type, such as whether it is item-level, whether it is locally differentially private, whether it is jointly differentially private, and whether it is memoryless.

B. DP Algorithms

In this section, we list all the differentially private mechanisms that are used in Algorithm 1 and provide their guarantees. First, we introduce the formal definition of (r, β)-concentration.

Definition B.1 ((r, β)-concentration)**.** $\{x_i\}_i \subset \mathbb{R}^d$ is (r, β) concentrated if there exists $\mu \in \mathbb{R}^d$ such that $\mathbb{P}(\forall i, \ \|x_i - \mu\| \geq r) \leq \beta$.

The PrivateRange algorithm outputs a private interval with length $4r$ if the input dataset is (r, β)-concentration such that the dataset falls into this interval with high probability.

Algorithm 2 PrivateRange($\{x_i\}_{i\in[M]}, \varepsilon, r, B$) (Feldman & Steinke, 2017)

1: **Input:** $\{x_i\}_{i\in[M]} \in [-B, B]^M$, r: concentration radius, ε: privacy parameter.
2: Divide interval $[-B, B]$ into $\ell = B/r$ disjoint bins, each with length $2r$. Let S be the set of middle points of those bins.
3: For $i \in [M]$, let $x_i' = \arg\min_{x\in S} |x - x_i|$ be the point in S closest to x_i.
4: For any $x \in S$, define cost function

$$c(x) = \max\left\{\left|i \in [M] : x_i' < x\right|, \left|i \in [M] : x_i' > x\right|\right\}.$$

5: Sample $\bar{x} \in S$ from the distribution:

$$\mathbb{P}(\bar{x} = x) = \frac{\exp(-\varepsilon c(x)/2)}{\sum_{x'\in S} \exp(-\varepsilon c(x')/2)}.$$

6: Return $[\bar{x} - 2r, \bar{x} + 2r]$.

Lemma B.2. *PrivateRange($\{x_i\}_{i\in[M]}, \varepsilon, r, B$) is $(\varepsilon, 0)$ differentially private.*

Algorithm 3 WinsorizedMean1D($\{x_i\}_{i\in[M]}, r, \beta, \varepsilon, B$) (Levy et al., 2021)

1: **Input:** $\{x_i\}_{i\in[M]} \in [-B, B]^M$, r: concentration radius, γ: concentration error probability, ε, δ: privacy parameter.
2: $[a, b]$ = PrivateRange($\{x_i\}_i, \varepsilon/2, r, B$), where $b - a = 4r$.
3: Sample $\xi \sim Lap(0, 8r/(M\varepsilon))$ and return

$$\bar{x} = \xi + \frac{1}{M}\sum_{i=1}^{M} \max\{a, \min\{b, x_i\}\}.$$

Lemma B.3 (Theorem 1 in Levy et al. (2021))**.** *WinsorizedMean1D is $(\varepsilon, 0)$-differentially private.*

We slightly modify a constant in the following WinsorizedMeanHighD (`WMHD`) algorithm. Compared to the original version, this modification is due to a tighter parameter choice in the advanced composition rule G.2.

Algorithm 4 WinsorizedMeanHighD($\{x_i\}_{i\in[M]}, r, \beta, \varepsilon, \delta$) (Levy et al., 2021)

1: **Input:** $\{x_i\}_{i\in[M]} \subset \mathbb{S}_1 \subset \mathbb{R}^d$, r: concentration radius, γ: concentration error probability, ε, δ: privacy parameter.
2: Let $D = \text{diag}(w_1, \ldots, w_d)$, where w_s are sampled independently and uniformly from $\{1, -1\}$ for all $s \in [d]$.
3: Set $\mathbf{U} = \frac{1}{\sqrt{d}}\mathbf{H}D$, where $\mathbf{H}$ is a d-dimensional Hadamard matrix.
4: For all $i \in [M]$, $s \in [d]$, compute

$$y_{i,s} = e_s^\top \mathbf{U} X_i.$$

5: Let $\varepsilon' = \frac{\varepsilon}{\sqrt{6d\log(1/\delta)}}$, $r' = 10r\sqrt{\frac{\log(dM/\beta)}{d}}$.
6: For all $s \in [d]$, compute

$$Y_s = \text{WinsorizedMean1D}(\{y_{i,s}\}_{i\in[M]}, r', \beta, \varepsilon', B\sqrt{d}).$$

7: Let $\bar{Y} = (Y_1, \ldots, Y_d)^\intercal$ and return

$$\bar{X} = \mathbf{U}^{-1}\bar{Y}.$$

Theorem B.4 (Theorem 2 in Levy et al. (2021)). *If $\hat{\theta} \in \mathbb{R}^d$ is the output of WinsorizedMeanHighD$(\{\theta_i\}_i, r, \beta, \varepsilon, \delta, B)$, then $\hat{\theta}$ is (ε, δ)-differentially private. Moreover, let each coordinate of $\boldsymbol{\xi} \in \mathbb{R}^d$ be independently sampled from Lap$(8r'/(M\varepsilon'))$ and $\bar{\theta} = \sum_i \theta_i/M$ be the sample mean, where $\varepsilon' = \frac{\varepsilon}{\sqrt{6d\log(1/\delta)}}$ and $r' = 10r\sqrt{(\log(dM/\beta))/d}$. Then, we have*

$$d_{TV}\left(\mathbb{P}\left(\|\hat{\theta}-\bar{\theta}\|\big|\{\theta_i\}_i\right), \mathbb{P}\left(\|\boldsymbol{\xi}\|\big|\{\theta_i\}_i\right)\right) \leq \beta + \frac{d^2 B}{10r\sqrt{(\log(dM/\beta))}}\exp\left(-\frac{M\varepsilon}{8\sqrt{6d\log(1/\delta)}}\right).$$

Corollary B.5. *Under the same setting as in Theorem B.4, we have*

$$\mathbb{P}\left(\|\hat{\theta}-\bar{\theta}\| \geq \frac{80r\log(d/\beta)\sqrt{6d\log(dM/\beta)\log(1/\delta)}}{M\varepsilon}\right) \leq 3\beta + \frac{d^2 B}{10r\sqrt{(\log(dM/\beta))}}\exp\left(-\frac{M\varepsilon}{8\sqrt{6d\log(1/\delta)}}\right).$$

Proof. By the definition of (r,β)-concentration, we have

$$\mathbb{P}\big(\{\theta_i\}_i \text{ is not } (r,0)\text{-concentrated}\big) \leq \beta.$$

Therefore,

$$\begin{aligned}
&\mathbb{P}\left(\|\hat{\theta}-\bar{\theta}\| \geq \frac{80r\log(d/\beta)\sqrt{6d\log(dM/\beta)\log(1/\delta)}}{M\varepsilon}\right)\\
&\leq \mathbb{P}\left(\|\hat{\theta}-\bar{\theta}\| \geq \frac{80r\log(d/\beta)\sqrt{6d\log(dM/\beta)\log(1/\delta)}}{M\varepsilon}\Big|\{\theta_i\}_i \text{ is } (r,0)\text{-concentrated}\right)\mathbb{P}\left(\{\theta_i\}_i \text{ is } (r,0)\text{-concentrated}\right)\\
&\quad + \mathbb{P}\left(\{\theta_i\}_i \text{ is not } (r,0)\text{-concentrated}\right)\\
&\overset{(a)}{\leq} 2\beta + \frac{d^2 B}{10r\sqrt{(\log(dM/\beta))}}\exp\left(-\frac{M\varepsilon}{8\sqrt{6d\log(1/\delta)}}\right) + \mathbb{P}\left(\|\boldsymbol{\xi}\| \geq \frac{80r\log(d/\beta)\sqrt{6d\log(dM/\beta)\log(1/\delta)}}{M\varepsilon}\right)\\
&= 2\beta + \frac{d^2 B}{10r\sqrt{(\log(dM/\beta))}}\exp\left(-\frac{M\varepsilon}{8\sqrt{6d\log(1/\delta)}}\right) + \mathbb{P}\left(\|\boldsymbol{\xi}\| \geq \frac{8r'\sqrt{d}}{M\varepsilon'}\log(d/\beta)\right)\\
&\overset{(b)}{\leq} 3\beta + \frac{d^2 B}{10r\sqrt{(\log(dM/\beta))}}\exp\left(-\frac{M\varepsilon}{8\sqrt{6d\log(1/\delta)}}\right),
\end{aligned}$$

where (a) is due to Theorem B.4, and (b) follows from the tail bound for Laplace random vectors developed in Lemma G.3.

□

C. Proof of the Regret Upper Bound of ROBIN

In this section, we provide the analysis for the upper bound of the regret of Algorithm 1. The key idea is to show that the global estimator is sufficiently accurate.

Outline of this section: In **Step 1**, we first develop Lemma C.2 aided by Lemma C.1, and show the minimum eigenvalue of the Gram matrix $V_{i,U}$ at the end of the initialization phase U scales linearly in $|\mathcal{T}_U|$. Then, in **Step 2**, Proposition C.3 shows that linearly growing minimum eigenvalues implies accurate global estimators and vice versa, which inductively verifies that the estimation error of $\hat{\theta}^{p+1}$ decays in the order of $O(1/\sqrt{M|\mathcal{T}_p|})$. Finally, with the accurate global estimator, Theorem C.6 in **Step 3** presents the regret upper bound.

Recall that the Gram matrix of client i at the end of phase p is

$$\tilde{V}_{i,p} = \sum_{\tau\in\mathcal{T}_p} x_{i,\tau,a_{i,\tau}} x_{i,\tau,a_{i,\tau}}^{\intercal}.$$

Step 1: Bound the Minimum Eigenvalue in the Initialization Phase.

This step is akin to Lemma 12 and Lemma 13 in Papini et al. (2021) and consists of two parts. The first part states that when the sub-optimal gap is larger that Δ, the number of times that LinUCB does not choose the optimal arm is upper bounded by $\tilde{O}(1/\Delta)$.

Lemma C.1. *Given any $\beta \in (0,1)$, with probability at least $1-\beta$, and for any $\Delta > 0$, the following inequality holds for any client i.*

$$\mathbb{E}\left[\sum_{\tau\in\mathcal{T}_U} \mathbb{1}\left\{a_{i,\tau} \neq a_{c^*_{i,\tau}}, \forall b \neq a^*_{c_{i,\tau}}, (\phi(c_{i,\tau}, a^*_{c_{i,\tau}}) - \phi(c_{i,\tau}, b))^\intercal \theta^* \geq \Delta\right\}\right] \leq \frac{2\alpha\sqrt{d|\mathcal{T}_U|\log|\mathcal{T}_U|}}{\Delta}, \tag{3}$$

where $\alpha = 1 + \sqrt{2\log(M/\beta) + d\log|\mathcal{T}_U|}$.

Proof. By Theorem 20.5 in Lattimore & Szepesvári (2020), if $\hat{\theta}_{i,t}$ is the local estimator in the initialization phase ($t \in \mathcal{T}_U$), we have for each client i, and any $\beta > 0$,

$$\mathbb{P}\left(\forall t \in \mathcal{T}_U, \|\hat{\theta}_{i,t} - \theta^*\|_{I+V_{i,t}} \leq 1 + \sqrt{2\log(1/\beta) + d\log|\mathcal{T}_U|}\right) \geq 1-\beta.$$

By rescaling β to β/M and taking the union bound, we have

$$\mathbb{P}\left(\forall i \in [M], t \in \mathcal{T}_U, \|\hat{\theta}_{i,t} - \theta^*\|_{I+V_{i,t}} \leq 1 + \sqrt{2\log(M/\beta) + d\log|\mathcal{T}_U|}\right) \geq 1-\beta,$$

where β is any positive number.

Let $\alpha = 1 + \sqrt{2\log(M/\beta) + d\log|\mathcal{T}_U|}$. Hence, the instantaneous regret can be upper bounded by

$$\begin{aligned}
&(\phi(c_{i,\tau}, a^*_{c_{i,\tau}}) - \phi(c_{i,\tau}, a_{i,\tau}))^\intercal \theta^* \\
&= \phi(c_{i,\tau}, a^*_{c_{i,\tau}})^\intercal(\theta^* - \hat{\theta}_{i,\tau}) - \phi(c_{i,\tau}, a_{i,\tau})^\intercal(\theta^* - \hat{\theta}_{i,\tau}) + \phi(c_{i,\tau}, a^*_{c_{i,\tau}})^\intercal\hat{\theta}_{i,\tau} - \phi(c_{i,\tau}, a_{i,\tau})^\intercal\hat{\theta}_{i,\tau} \\
&\leq \|\phi(c_{i,\tau}, a^*_{c_{i,\tau}})\|_{(I+V_{i,\tau})^{-1}}\|\theta^* - \hat{\theta}_{i,\tau}\|_{I+V_{i,\tau}} + \|\phi(c_{i,\tau}, a_{i,\tau})\|_{(I+V_{i,\tau})^{-1}}\|\theta^* - \hat{\theta}_{i,\tau}\|_{I+V_{i,\tau}} \\
&\quad + \phi(c_{i,\tau}, a^*_{c_{i,\tau}})^\intercal\hat{\theta}_{i,\tau} - \phi(c_{i,\tau}, a_{i,\tau})^\intercal\hat{\theta}_{i,\tau} \\
&\leq \alpha\|\phi(c_{i,\tau}, a^*_{c_{i,\tau}})\|_{(I+V_{i,\tau})^{-1}} + \alpha\|\phi(c_{i,\tau}, a_{i,\tau})\|_{(I+V_{i,\tau})^{-1}} + \phi(c_{i,\tau}, a^*_{c_{i,\tau}})^\intercal\hat{\theta}_{i,\tau} - \phi(c_{i,\tau}, a_{i,\tau})^\intercal\hat{\theta}_{i,\tau} \\
&\leq 2\alpha\|\phi(c_{i,\tau}, a_{i,\tau})\|_{(I+V_{i,\tau})^{-1}}.
\end{aligned}$$

Due the boundness assumption, we can conclude that

$$(\phi(c_{i,\tau}, a^*_{c_{i,\tau}}) - \phi(c_{i,\tau}, a_{i,\tau}))^\intercal \theta^* \leq 2\min\left\{\alpha\|\phi(c_{i,\tau}, a_{i,\tau})\|_{(I+V_{i,\tau})^{-1}}, 1\right\}.$$

Therefore, we can bound Equation (3) as follows:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{\tau\in\mathcal{T}_U} \mathbb{1}\left\{a_{i,\tau} \neq a_{c^*_{i,\tau}}, \forall b \neq a^*_{c_{i,\tau}}, (\phi(c_{i,\tau}, a^*_{c_{i,\tau}}) - \phi(c_{i,\tau}, b))^\intercal \theta^* \geq \Delta\right\}\right] \\
&\leq \mathbb{E}\left[\sum_{\tau\in\mathcal{T}_U} \mathbb{1}\left\{a_{i,\tau} \neq a_{c^*_{i,\tau}}, \forall b \neq a^*_{c_{i,\tau}}, (\phi(c_{i,\tau}, a^*_{c_{i,\tau}}) - \phi(c_{i,\tau}, b))^\intercal \theta^* \geq \Delta\right\} \frac{(\phi(c_{i,\tau}, a^*_{c_{i,\tau}}) - \phi(c_{i,\tau}, a_{i,\tau}))^\intercal\theta^*}{\Delta}\right] \\
&\leq \mathbb{E}\left[\sum_{\tau\in\mathcal{T}_U} \mathbb{1}\left\{a_{i,\tau} \neq a_{c^*_{i,\tau}}, \forall b \neq a^*_{c_{i,\tau}}, (\phi(c_{i,\tau}, a^*_{c_{i,\tau}}) - \phi(c_{i,\tau}, b))^\intercal \theta^* \geq \Delta\right\} \frac{2\min\{\alpha\|x_{i,\tau,a_{i,\tau}}\|_{(I+V_{i,\tau})^{-1}}, 1\}}{\Delta}\right] \\
&\leq \mathbb{E}\left[\sum_{\tau\in\mathcal{T}_U} \frac{2\min\{\alpha\|x_{i,\tau,a_{i,\tau}}\|_{(I+V_{i,\tau})^{-1}}, 1\}}{\Delta}\right] \\
&\overset{(a)}{\leq} \frac{2\alpha\sqrt{d|\mathcal{T}_U|\log|\mathcal{T}_U|}}{\Delta},
\end{aligned}$$

where (a) follows from the elliptical potential lemma in Lemma G.5. □

Based on Lemma C.1, we prove the second part that the minimum eigenvalue of the Gram matrix $\tilde{V}_{i,U}$ scales linearly in $|\mathcal{T}_U|$ with high probability.

Lemma C.2 (Guarantee in the first U phases)**.** *If U satisfies*

$$\frac{12(dC_0+\lambda_0)\sqrt{|\mathcal{T}_U|\log(2dM/\beta)}\log|\mathcal{T}_U|}{\lambda_0} \le \frac{\lambda_0}{4}|\mathcal{T}_U|,$$

then, for any $\beta \in (0,1)$, with probability at least $1-\beta$, the Gram matrix of any client i at the end of phase U has full rank, and

$$\forall i \in [M],\ \ \lambda_{\min}\left(\tilde{V}_{i,U}\right) \ge \frac{\lambda_0}{4}|\mathcal{T}_U|.$$

Proof. We analyze the expectation of the Gram matrix $\tilde{V}_{i,U}$ as follows.

$$\begin{aligned}
\mathbb{E}[\tilde{V}_{i,U}] &= \mathbb{E}\left[\sum_{\tau\in\mathcal{T}_U} x_{i,\tau,a_{i,\tau}} x_{i,\tau,a_{i,\tau}}^\intercal\right] \\
&\ge \mathbb{E}\left[\sum_{\tau\in\mathcal{T}_U} \phi(c_{i,\tau},a_{i,\tau})\phi(c_{i,\tau},a_{i,\tau})^\intercal \mathbb{1}\{a_{i,\tau}=a_{c_{i,\tau}^*},\forall b\neq a_{c_{i,\tau}^*}^*,(\phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)-\phi(c_{i,\tau},b))^\intercal\theta^*\ge\Delta\}\right] \\
&= \mathbb{E}\left[\sum_{\tau\in\mathcal{T}_U} \phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)\phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)^\intercal \mathbb{1}\{a_{i,\tau}=a_{c_{i,\tau}^*},\forall b\neq a_{c_{i,\tau}^*}^*,(\phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)-\phi(c_{i,\tau},b))^\intercal\theta^*\ge\Delta\}\right] \\
&= \mathbb{E}\left[\sum_{\tau\in\mathcal{T}_U} \phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)\phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)^\intercal\right] \\
&\quad - \mathbb{E}\left[\sum_{\tau\in\mathcal{T}_U} \phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)\phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)^\intercal \mathbb{1}\{\exists b\neq a_{c_{i,\tau}^*}^*,(\phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)-\phi(c_{i,\tau},b))^\intercal\theta^*<\Delta\}\right] \\
&\quad - \mathbb{E}\left[\sum_{\tau\in\mathcal{T}_U} \phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)\phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)^\intercal \mathbb{1}\{a_{i,\tau}\neq a_{c_{i,\tau}^*},\forall b\neq a_{c_{i,\tau}^*}^*,(\phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)-\phi(c_{i,\tau},b))^\intercal\theta^*\ge\Delta\}\right] \\
&\overset{(a)}{\ge} (\lambda_0 - C_0\Delta)|\mathcal{T}_U| I_d - I_d\mathbb{E}\left[\sum_{\tau\in\mathcal{T}_U} \mathbb{1}\{a_{i,\tau}\neq a_{c_{i,\tau}^*},\forall b\neq a_{c_{i,\tau}^*}^*,(\phi(c_{i,\tau},a_{c_{i,\tau}^*}^*)-\phi(c_{i,\tau},b))^\intercal\theta^*\ge\Delta\}\right] \\
&\overset{(b)}{\ge} (\lambda_0 - C_0\Delta)|\mathcal{T}_U| I_d - \frac{2\alpha\sqrt{d|\mathcal{T}_U|\log|\mathcal{T}_U|}}{\Delta} I_d, \text{ with probability at least } 1-\beta_0,
\end{aligned}$$

where $\alpha = 1+\sqrt{2\log(M/\beta_0)+d\log|\mathcal{T}_U|}$, (a) is due to the diversity condition in Assumption 3.1 and the margin condition in Assumption 3.2, and (b) follows from Lemma C.1.

Now choose $\Delta = \frac{\lambda_0}{2C_0}$. We have, with probability at least $1-\beta_0$,

$$\mathbb{E}[\tilde{V}_{i,U}] \ge \left(\frac{\lambda_0|\mathcal{T}_U|}{2} - \frac{4\alpha C_0\sqrt{d|\mathcal{T}_U|\log|\mathcal{T}_U|}}{\lambda_0}\right) I_d.$$

Consider the difference $\tilde{V}_{i,U} - \mathbb{E}[\tilde{V}_{i,U}]$, which is a summation of zero-mean matrices. Thus, we can apply matrix concentration inequality in Lemma G.4 to obtain that, for any $\beta_1 > 0$, with probability at least $1-\beta_1-\beta_0$, we have

$$\begin{aligned}
\lambda_{\min}(\tilde{V}_{i,U}) &= \lambda_{\min}\left(\tilde{V}_{i,U} - \mathbb{E}[\tilde{V}_{i,U}] + \mathbb{E}[\tilde{V}_{i,U}]\right) \\
&\ge \lambda_{\min}\left(\mathbb{E}[\tilde{V}_{i,U}]) - \lambda_{\max}(\tilde{V}_{i,U} - \mathbb{E}[\tilde{V}_{i,U}]\right) \\
&\ge \frac{\lambda_0|\mathcal{T}_U|}{2} - \frac{4\alpha C_0\sqrt{d|\mathcal{T}_U|\log|\mathcal{T}_U|}}{\lambda_0} - \sqrt{2|\mathcal{T}_U|\log(d/\beta_1)} - 2/3.
\end{aligned}$$

Choosing $\beta_1 = 1/(M\beta_0)$ and taking the union bound, we have, with probability at least $1 - 2\beta_0$,

$$\lambda_{\min}(\tilde{V}_{i,U}) \geq \frac{\lambda_0|\mathcal{T}_U|}{2} - \frac{12(dC_0 + \lambda_0)\sqrt{|\mathcal{T}_U|\log(dM/\beta_0)}\log|\mathcal{T}_U|}{\lambda_0}.$$

We finish the proof by choosing $\beta_0 = \beta/2$ and noting that

$$\frac{12(dC_0 + \lambda_0)\sqrt{|\mathcal{T}_U|\log(2dM/\beta)}\log|\mathcal{T}_U|}{\lambda_0} \leq \frac{\lambda_0}{4}|\mathcal{T}_U|.$$

□

Step 2: Inductively Bound the Minimum Eigenvalue and Global Estimation Error after Initialization.

Proposition C.3. *Fix $\beta > 0$, ε_0, δ_0, and $c_1 = \frac{4\sqrt{2d\log(16d(M+1)P/\beta)}}{\lambda_0}$ as defined in Algorithm 1. If there are total P phases, consider the following events:*

$$\mathcal{E}_V = \left\{\forall i \in [M], p \in [U:P],\ \ \tilde{V}_{i,p} \geq \frac{\lambda_0|\mathcal{T}_p|}{4} I_d\right\},$$
$$\mathcal{E}_\theta = \left\{\forall p \in [U+1:P],\ \ \|\hat{\theta}^p - \theta^*\| \leq \frac{c_2}{\sqrt{M|\mathcal{T}_p|}}\right\},$$

where the parameters are set as

$$|\mathcal{T}_p| = 2^p, \forall p \in [P],$$
$$U = \left\lceil \max\left\{\log_2\left(\frac{64\log(2dMP/\beta)}{\lambda_0^2}\right), \log_2\left(\frac{144dU^2(C_0+\lambda_0)^2\log(2dMP/\beta)\log 2}{\lambda_0^4}\right)\right\}\right\rceil$$
$$c_2 = \frac{4\sqrt{2d\log(16d(M+1)P/\beta)}}{\lambda_0}\left(\frac{80\log(16dP/\beta)\sqrt{6d\log(16dMP/\beta)\log(1/\delta_0)}}{\sqrt{M}\varepsilon_0} + 1\right),$$
$$M \geq \max\left\{\frac{8\sqrt{6d\log(1/\delta_0)}}{\varepsilon_0}\log\left(\frac{4d^2P\sqrt{|\mathcal{T}_P|}}{10\beta c_1\sqrt{\log(16dMP/\beta)}}\right), \frac{32C_0^2c_2^2}{\lambda_0^2|\mathcal{T}_U|}\right\}.$$

Then, Algorithm 1 guarantees that

$$\mathbb{P}\left(\mathcal{E}_V\mathcal{E}_\theta\right) \geq 1 - \beta.$$

Proof. Let us divide $\mathcal{E}_V$ and $\mathcal{E}_\theta$ into disjoint sub-events:

$$\mathcal{E}_{V,p} = \left\{\forall i \in [M],\ \ \tilde{V}_{i,p} \geq \frac{\lambda_0|\mathcal{T}_p|}{4} I_d\right\},$$
$$\mathcal{E}_{\theta,p} = \left\{\forall i \in [M], \|\hat{\theta}^{p+1} - \theta^*\| \leq \frac{c_2}{\sqrt{M|\mathcal{T}_p|}}\right\},$$

such that $\mathcal{E}_V = \cap_{p\geq U}\mathcal{E}_{V,p}$ and $\mathcal{E}_\theta = \cap_{p\geq U}\mathcal{E}_{\theta,p}$. We aim to verify the following two claims.

Claim C.4. *Under event $\mathcal{E}_{V,p}$, for any $\beta > 0$, we have $\mathbb{P}(\mathcal{E}_{\theta,p}) \geq 1 - \beta/(2P)$.*

Claim C.5. *Under event $\mathcal{E}_{\theta,p}$, for any $\beta > 0$, we have $\mathbb{P}(\mathcal{E}_{V,p+1}) \geq 1 - \beta/(2P)$.*

Proof of Claim C.4:

We recall the definition of local estimators $\tilde{\theta}_{i,p}$:

$$\tilde{\theta}_{i,p} = \tilde{V}_{i,p}^{\dagger}\left(\sum_{\tau\in\mathcal{T}_p} x_{i,\tau,a_{i,\tau}} r_{i,\tau}\right) = \tilde{V}_{i,p}^{\dagger}\left(\sum_{\tau\in\mathcal{T}_p} x_{i,\tau,a_{i,\tau}}\left(\eta_{i,\tau} + x_{i,\tau,a_{i,\tau}}^{\intercal}\theta^*\right)\right),$$

where $\eta_{i,\tau} = r_{i,\tau} - x_{i,\tau,a_{i,\tau}}^{\intercal}\theta^*$ is the IID Gaussian noise. Since $\mathcal{E}_{V,p}$ asserts that the covariance matrix $\tilde{V}_{i,p}$ is full rank, we have

$$\tilde{\theta}_{i,p} - \theta^* = \tilde{V}_{i,p}^{-1}\left(\sum_{\tau\in\mathcal{T}_p} x_{i,\tau,a_{i,\tau}}\eta_{i,\tau}\right),$$

which is a summation of independent $\sigma_{i,\tau}^2$-sub-Gaussian random variables conditioned on $\tilde{V}_{i,p}$, and

$$\sigma_{i,\tau}^2 \le \|\tilde{V}_{i,p}^{-1}x_{i,\tau,a_{i,\tau}}\|^2 \le \frac{1}{\lambda_{\min}(\tilde{V}_{i,p})^2} \le \frac{16}{\lambda_0^2|\mathcal{T}_p|^2}.$$

Due to the independence of $\{\eta_{i,\tau}\}_\tau$ conditioned on $\{x_{i,\tau}\}$, we have that $\tilde{\theta}_{i,p} - \theta^*$ is a σ^2-sub-Gaussian random vector with $\sigma^2 = \frac{16}{\lambda_0^2|\mathcal{T}_p|}$. Thus, for any $\beta_1 > 0$, we have with probability at least $1-\beta_1$,

$$\|\tilde{\theta}_{i,p} - \theta^*\| \le \frac{4\sqrt{2d\log(2d/\beta_1)}}{\lambda_0\sqrt{|\mathcal{T}_p|}}.$$

In addition, note that

$$\frac{1}{M}\sum_i\left(\tilde{\theta}_{i,p} - \theta^*\right) = \frac{1}{M}\sum_{i,\tau}\tilde{V}_{i,p}^{-1}\left(x_{i,\tau,a_{i,\tau}}\eta_{i,\tau}\right),$$

which is a sub-Gaussian random vector conditioned on covariance matrices $\{\tilde{V}_{i,p}\}_i$. Following the same argument, we have

$$\mathbb{P}\left(\left\|\frac{1}{M}\sum_i\left(\tilde{\theta}_{i,p} - \theta^*\right)\right\| \le \frac{4\sqrt{2d\log(2d/\beta_1)}}{\lambda_0\sqrt{M|\mathcal{T}_p|}}\right) \ge 1-\beta_1, \forall\beta_1 > 0.$$

So far, by rescaling β_1 to $\beta_1/(M+1)$, we have verified that for any $\beta_1 > 0$, under event $\mathcal{E}_{V,p} = \{\forall i,\ \ \lambda_{\min}(\tilde{V}_{i,p}) \ge \lambda_0|\mathcal{T}_p|/4\}$,

$$\mathbb{P}\left(\forall i,\ \ \|\tilde{\theta}_{i,p} - \theta^*\| \le \frac{4\sqrt{2d\log(2d(M+1)/\beta_1)}}{\lambda_0\sqrt{|\mathcal{T}_p|}},\ \text{and}\ \left\|\frac{1}{M}\sum_i\left(\tilde{\theta}_{i,p} - \theta^*\right)\right\| \le \frac{4\sqrt{2d\log(2d(M+1)/\beta_1)}}{\lambda_0\sqrt{M|\mathcal{T}_p|}}\right) \ge 1-\beta_1. \tag{4}$$

Let $c_1 = 4\sqrt{2d\log(2d(M+1)/\beta_1)}/\lambda_0$. Then, $\{\tilde{\theta}_{i,p}\}_{i\in[M]}$ are $(c_1/\sqrt{|\mathcal{T}_p|}, \beta_1)$-concentrated. By the design of Algorithm 1 and Corollary B.5, we have

$$\begin{aligned}
&\mathbb{P}\left(\|\hat{\theta}^{p+1} - \frac{1}{M}\sum_i\tilde{\theta}_{i,p}\| \ge \frac{80c_1\log(d/\beta_1)\sqrt{6d\log(dM/\beta_1)\log(1/\delta_0)}}{M\varepsilon_0\sqrt{|\mathcal{T}_p|}}\right)\\
&\le 3\beta_1 + \frac{d^2\sqrt{|\mathcal{T}_p|}}{10c_1\sqrt{(\log(dM/\beta_1))}}\exp\left(-\frac{M\varepsilon_0}{8\sqrt{6d\log(1/\delta_0)}}\right).
\end{aligned}$$

Combining with Equation (4), we conclude that, under event $\mathcal{E}_{V,p}$,

$$\mathbb{P}\left(\|(\hat{\theta}^{p+1} - \theta^*\| \ge \frac{80c_1\log(d/\beta_1)\sqrt{6d\log(dM/(\beta_1))\log(1/\delta_0)}}{\sqrt{|\mathcal{T}_p|}M\varepsilon_0} + \frac{c_1}{\sqrt{M|\mathcal{T}_p|}}\right)$$

$$\leq 4\beta_1 + \frac{d^2\sqrt{|\mathcal{T}_p|}}{10c_1\sqrt{(\log(dM/\beta_1))}}\exp\left(-\frac{M\varepsilon_0}{8\sqrt{6d\log(1/\delta_0)}}\right),$$

where $c_1 = \frac{4\sqrt{2d\log(d(M+1)/\beta_1)}}{\lambda_0}$, and β_1 is an arbitrary positive number.

Let

$$\begin{aligned}
&\beta_1 = \beta/(16P),\\
&c_1 = \frac{4\sqrt{2d\log(16d(M+1)P/\beta)}}{\lambda_0},\\
&c_2 = c_1\left(\frac{80\log(d/\beta_1)\sqrt{6d\log(16dMP/\beta)\log(1/\delta_0)}}{\sqrt{M}\varepsilon_0} + 1\right).
\end{aligned}$$

Note that

$$M\varepsilon_0 \geq 8\sqrt{6d\log(1/\delta_0)}\log\left(\frac{4d^2P\sqrt{|\mathcal{T}_p|}}{10\beta c_1\sqrt{\log(16dMP/\beta)}}\right).$$

We simplify the result as follows. Under event $\mathcal{E}_{V,p}$,

$$\mathbb{P}\left(\|(\hat{\theta}^{p+1} - \theta^* \| \geq \frac{c_2}{\sqrt{M|\mathcal{T}_p|}}\right) \leq \frac{\beta}{2P},$$

which completes the proof of Claim C.4.

Next, we prove Claim C.5.

Proof of Claim C.5:

Recall that at each time slot $t \in \mathcal{T}_{p+1}$, client i greedily chooses an arm $a_{i,t}$ with respect to the estimator $\hat{\theta}^{p+1}$ under context $c_{i,t}$ drawn from ρ_i. We consider the matrix $\mathbb{E}_{c_{i,t}\sim\rho_i}\left[\phi(c_{i,t},a_{i,t})\phi(c_{i,t},a_{i,t})^\intercal\right]$ under the assumption that $\mathcal{E}_{\theta,p}$ holds. We have

$$\begin{aligned}
&\mathbb{E}_{c_{i,t}\sim\rho}\left[\phi(c_{i,t},a_{i,t})\phi(c_{i,t},a_{i,t})^\intercal\right]\\
&\quad= \sum_{a\in\mathcal{A}}\mathbb{E}_{c\sim\rho_i}\left[\phi(c,a)\phi(c,a)^\intercal\mathbb{1}\left\{\forall b\neq a,\ (\phi(c,a)-\phi(c,b))^\intercal\hat{\theta}^{p+1}\geq 0\right\}\right]\\
&\quad= \sum_{a\in\mathcal{A}}\mathbb{E}_{c\sim\rho}\left[\phi(c,a)\phi(c,a)^\intercal\mathbb{1}\left\{\forall b\neq a,\ (\phi(c,a)-\phi(c,b))^\intercal\theta^*\geq(\phi(c,a)-\phi(c,b))^\intercal(\theta^*-\hat{\theta}^{p+1})\right\}\right]\\
&\quad\overset{(a)}{\geq} \sum_{a\in\mathcal{A}}\mathbb{E}\left[\phi(c,a)\phi(c,a)^\intercal\mathbb{1}\left\{\forall b\neq a,\ (\phi(c,a)-\phi(c,b))^\intercal\theta^*\geq\frac{2c_2}{\sqrt{M|\mathcal{T}_p|}}\right\}\right]\\
&\quad= \sum_{a\in\mathcal{A}}\mathbb{E}\left[\phi(c,a)\phi(c,a)^\intercal\mathbb{1}\left\{\forall b\neq a,\ (\phi(c,a)-\phi(c,b))^\intercal\theta^*\geq\frac{2c_2}{\sqrt{M|\mathcal{T}_p|}}\right\}\right]\\
&\quad= \mathbb{E}\left[\phi(c,a_c^*)\phi(c,a_c^*)^\intercal\mathbb{1}\left\{\forall b\neq a_c^*,\ (\phi(c,a_c^*)-\phi(c,b))^\intercal\theta^*\geq\frac{2c_2}{\sqrt{M|\mathcal{T}_p|}}\right\}\right]\\
&\quad= \mathbb{E}\left[\phi(c,a_c^*)\phi(c,a_c^*)^\intercal\right]\\
&\qquad-\mathbb{E}\left[\phi(c,a_c^*)\phi(c,a_c^*)^\intercal\mathbb{1}\left\{\exists b\neq a_c^*,\ (\phi(c,a_c^*)-\phi(c,b))^\intercal\theta^*<\frac{2c_2}{\sqrt{M|\mathcal{T}_p|}}\right\}\right]\\
&\quad\geq \lambda_0 I - I\mathbb{P}\left(\exists b\neq a_c^*,\ (\phi(c,a_c^*)-\phi(c,b))^\intercal\theta^*<\frac{2c_2}{\sqrt{M|\mathcal{T}_p|}}\right)
\end{aligned}$$

$$\overset{(b)}{\geq} \left(\lambda_0 - \frac{2C_0c_2}{\sqrt{M|\mathcal{T}_p|}}\right) I_d,$$

where (a) follows from $\mathcal{E}_{\theta,p}$, and (b) is due to the margin condition in Assumption 3.2.

Therefore, for any $p \geq U$, we have

$$\mathbb{E}\left[\tilde{V}_{i,p+1}\right] \geq \lambda_0|\mathcal{T}_{p+1}| - |\mathcal{T}_{p+1}|\frac{2C_0c_2}{\sqrt{M|\mathcal{T}_p|}} \overset{(a)}{=} \lambda_0|\mathcal{T}_{p+1}| - 2C_0c_2\sqrt{\frac{2|\mathcal{T}_{p+1}|}{M}},$$

where (a) is due to $|\mathcal{T}_{p+1}| = 2^{p+1} = 2|\mathcal{T}_p|$.

Thus, we can apply the matrix concentration inequality Lemma G.4 on the martingale difference $\phi(c_{i,t}, a_{i,t})\phi(c_{i,t}, a_{i,t})^\intercal - \mathbb{E}[\phi(c_{i,t}, a_{i,t})\phi(c_{i,t}, a_{i,t})^\intercal]$ for $t \in \mathcal{T}_p/\mathcal{T}_U$ to make the following conclusions: With probability at least $1 - \beta_0$, where $\beta_0 > 0$ is any positive number, we have

$$\begin{aligned}
\lambda_{\min}(\tilde{V}_{i,p+1}) &= \lambda_{\min}\left(\tilde{V}_{i,p+1} - \mathbb{E}[\tilde{V}_{i,p+1}] + \mathbb{E}[\tilde{V}_{i,p+1}]\right) \\
&\geq \lambda_{\min}\left(\mathbb{E}[\tilde{V}_{i,p+1}]\right) - \lambda_{\max}(\tilde{V}_{i,p+1} - \mathbb{E}[\tilde{V}_{i,p+1}]) \\
&\geq \lambda_0|\mathcal{T}_{p+1}| - 2C_0c_2\sqrt{\frac{2|\mathcal{T}_{p+1}|}{M}} - \sqrt{2|\mathcal{T}_{p+1}|\log(d/\beta_0)} - 2/3.
\end{aligned}$$

Now, set $\beta_0 = \beta/(2MP)$, and take the union bound over all clients. Then, with probability at least $1 - \beta/(2P)$, for all $i \in [M]$ the following holds.

$$\lambda_{\min}(\tilde{V}_{i,p+1}) \geq \lambda_0|\mathcal{T}_{p+1}| - 2C_0c_2\sqrt{\frac{2|\mathcal{T}_{p+1}|}{M}} - \sqrt{2|\mathcal{T}_{p+1}|\log(2dMP/\beta)} - 2/3,$$

where $c_2 = \frac{8\sqrt{2d\log(16d(M+1)P/\beta)}}{\lambda_0}\left(\frac{80\log(16dP/\beta)\sqrt{6d\log(16dMP/\beta)\log(1/\delta_0)}}{\sqrt{M}\varepsilon_0} + 1\right)$.

Note that $p \geq U$, and $|\mathcal{T}_U|$ satisfies the following inequality:

$$\sqrt{2|\mathcal{T}_U|\log(2dMP/\beta)} + 2/3 \leq \frac{\lambda_0|\mathcal{T}_U|}{4}.$$

In addition, due to $p \geq U$, we have that

$$2\sqrt{2}C_0c_2 \leq \frac{\lambda_0\sqrt{M|\mathcal{T}_{p+1}|}}{2}$$

holds for any p.

Thus, we conclude that, under event $\mathcal{E}_{\theta,p}$,

$$\mathbb{P}\left(\forall i \in [M],\ \ \lambda_{\min}(\tilde{V}_{i,p+1}) \geq \frac{\lambda_0|\mathcal{T}_{p+1}|}{4}\right) \geq 1 - \frac{\beta}{2P},$$

which finishes the proof of Claim C.5.

Then, based on Claim C.5 and Claim C.4, combining with Lemma C.2, we conclude that

$$\mathbb{P}\left(\mathcal{E}_V\mathcal{E}_\theta\right) = \mathbb{P}\left(\left(\cap_{p\geq U}\,\mathcal{E}_{V,p}\right) \cap \left(\cap_{p\geq U}\,\mathcal{E}_{\theta,p}\right)\right) \geq 1 - \beta.$$

□

Step 3: Upper Bound the Regret.

Now, we are ready to provide the final result on the upper bound of the regret of Algorithm 1.

Theorem C.6 (Regret). *Under the parameter setting in Proposition C.3, with probability at least $1-\beta$, the total regret of Algorithm 1 is upper bounded by*

$$\textit{Regret}(M,T) \leq \tilde{O}\left(\max\left(1, \frac{d\log T\log(1/\delta)\log^3(1/\beta)}{\varepsilon^2 M}\right)\frac{C_0 d\log(1/\beta)\log T}{\lambda_0^2}\right).$$

Proof. Consider a phase $p > U$. Under the events $\mathcal{E}_V$ and $\mathcal{E}_\theta$ defined in Proposition C.3, we have

$$\begin{aligned}
\mathbb{E}[r^*_{i,t} - r_{i,t}] &= \mathbb{E}\left[\left(\phi(c_{i,t}, a^*_{c_{i,t}}) - \phi(c_{i,t}, a_{i,t})\right)^\intercal \theta^*\right] \\
&= \mathbb{E}\left[\mathbb{1}\left\{a_{i,t} \neq a^*_{c_{i,t}}\right\}\left(\phi(c_{i,t}, a^*_{c_{i,t}}) - \phi(c_{i,t}, a_{i,t})\right)^\intercal \theta^*\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\left\{a_{i,t} \neq a^*_{c_{i,t}}\right\}\left(\phi(c_{i,t}, a^*_{c_{i,t}}) - \phi(c_{i,t}, a_{i,t})\right)^\intercal (\theta^* - \hat{\theta}^{(p)})\right] \\
&\overset{(a)}{\leq} \mathbb{E}\left[\mathbb{1}\left\{a_{i,t} \neq a^*_{c_{i,t}}\right\}\frac{2c_2}{\sqrt{M|\mathcal{T}_{p-1}|}}\right] \\
&= \frac{2c_2}{\sqrt{M|\mathcal{T}_{p-1}|}}\mathbb{P}(a_{i,t} \text{ is not optimal }) \\
&= \frac{2c_2}{\sqrt{M|\mathcal{T}_{p-1}|}}\mathbb{P}\left(\left(\phi(c_{i,t}, a_{i,t}) - \phi(c_{i,t}, a^*_{c_{i,t}})\right)^\intercal \hat{\theta}^{(p)} \geq 0\right) \\
&= \frac{2c_2}{\sqrt{M|\mathcal{T}_{p-1}|}}\mathbb{P}\left(0 > \left(\phi(c_{i,t}, a_{i,t}) - \phi(c_{i,t}, a^*_{c_{i,t}})\right)^\intercal \theta^* \geq \left(\phi(c_{i,t}, a_{i,t}) - \phi(c_{i,t}, a^*_{c_{i,t}})\right)^\intercal (\theta^* - \hat{\theta}^{(p)})\right) \\
&\overset{(b)}{\leq} \frac{2c_2}{\sqrt{M|\mathcal{T}_{p-1}|}}\mathbb{P}\left(0 < \left(\phi(c_{i,t}, a^*_{c_{i,t}}) - \phi(c_{i,t}, a_{i,t})\right)^\intercal \theta^* \leq \frac{2c_2}{\sqrt{M|\mathcal{T}_{p-1}|}}\right) \\
&\overset{(c)}{\leq} \frac{4C_0 c_2^2}{M|\mathcal{T}_{p-1}|},
\end{aligned}$$

where (a) and (b) follow from event $\mathcal{E}_\theta$, and (c) is due to the margin condition in Assumption 3.2.

Therefore, with probability at least $1-\beta$,

$$\begin{aligned}
\text{Regret}(M,T) &= \sum_{p\in[P]}\sum_{i\in[M], t\in\mathcal{T}_p}\mathbb{E}[r^*_{i,t} - r_{i,t}] \\
&\leq \sum_{p>U} M|\mathcal{T}_p| \cdot \frac{4C_0c_2^2}{M|\mathcal{T}_{p-1}|} + M|\mathcal{T}_U| \\
&= \sum_{p>U} 8C_0c_2^2 + M|\mathcal{T}_U| \\
&= 8C_0c_2^2 P + M|\mathcal{T}_U|.
\end{aligned}$$

We complete the proof by noting that

$$\begin{aligned}
&P = O(\log T), \\
&\varepsilon_0 = \varepsilon/\sqrt{P}, \\
&U = O(\log\log\log T), \\
&c_2 = \frac{4\sqrt{2d\log(16d(M+1)P/\beta)}}{\lambda_0}\left(\frac{80\log(16dP/\beta)\sqrt{6d\log(16dMP/\beta)\log(1/\delta_0)}}{\sqrt{M}\varepsilon_0} + 1\right) \\
&\quad = \tilde{O}\left(\frac{\sqrt{d\log(1/\beta)}}{\lambda_0}\max\left(1, \frac{\log(1/\beta)\sqrt{d\log T\log(1/\beta)\log(1/\delta)}}{\varepsilon\sqrt{M}}\right)\right).
\end{aligned}$$

□

D. Necessity of Diversity Assumption for Memoryless Algorithms under User-level CDP Constraint

In this section, we prove that the diversity condition in Assumption 3.1 is necessary for achieving sublinear regret when adopting almost-memoryless algorithms under the user-level CDP constraint.

Proposition D.1. *If $\varepsilon < 0.1$ and $\delta < 0.001$, then, there exists a federated linear contextual bandits instance not satisfying Assumption 3.1 such that any almost-memoryless algorithm must incur a regret lower bounded by $\Omega(T)$.*

Proof. We consider a federated linear contextual bandits with two arms $\{1, 2\}$ and M clients, where the context is fixed for each client. For each client $i \in [M]$, let features $\phi_i(c, 1), \phi_i(c, 2) \in \mathbb{R}^2$ be defined as follows. For the first client ($i = 1$), $\phi_1(c, 1) = (0.5, 0)^\intercal$ and $\phi_1(c, 2) = (-0.5, 0)^\intercal$. For any other client ($i > 1$), $\phi_i(c, 1) = (0, 0.5)^\intercal$ and $\phi_i(c, 2) = (0, -0.5)^\intercal$. Let the model parameter $\theta^* \in \Theta = \{\pm 1\} \times \{\pm 1\}$, and the reward $r_{i,t} \sim N\left(\phi_i(c, a_{i,t})^T \theta^*, 1\right)$.

We note that the feature distribution does not satisfy the **diversity assumption** (Assumption 3.1), since the smallest eigenvalue is 0 for each $\phi_i(c, 1)$ and $\phi_i(c, 2)$, while it satisfies the margin condition (Assumption 3.2) with $C_0 = 1$.

To reduce the dependency between each client, we again consider the extended history similarly in Appendix E.1. Let $H_{i,t} = \{c, a_{i,\tau}, r_{i,\tau}\}_{\tau<t}$ be the history of client i, and $\bar{H}_{i,t} = \{c, a = \{1, 2\}, r_{i,\tau,a}\}_{\tau<t} \supset H_{i,t}$ be the extended history of client i at time t. $\bar{H}_{i,t}$ is independent of $\bar{H}_{j,t}$ conditioned on model parameter θ^*.

Note that the total regret is lower bounded by the regret of the first user, and the regret of play sub-optimal arm is 1. Then, we have

$$\begin{aligned}\text{Regret}(M, T) &\geq \sum_{t\in[T]} \mathbb{P}(a_{1,t} \neq e_1^\intercal \theta^*) \\ &\geq \sum_{t\in[T]} \inf_{\substack{\hat{\theta}_{1,t} \in \mathcal{F}(\bar{\mathcal{I}}, \{1,-1\}) \\ \{\bar{R}_i\}_{i\in[0,M]}}} \mathbb{P}\left(e_1^\intercal \theta^* \neq \hat{\theta}_{1,t}(\bar{q}_{\leq t})\right).\end{aligned}$$

Now we aim to show that due to the user-level DP constraint, the estimation error cannot be too small for client 1.

Consider θ^* and θ', such that $e_1^\intercal \theta^* = -e_1^\intercal \theta' = 1$, and $e_2^\intercal \theta^* = e_2^\intercal \theta'$. We construct a coupling between $\bar{q}_{\leq t}|\theta^*$ and $\bar{q}_{\leq t}|\theta^{'}$. Specifically, if we flip the distribution of $r_{1,t,1}$ from $N(0.5, 1)$ to $N(-0.5, 1)$, and $r_{1,t,2}$'s distribution from $N(-0.5, 1)$ to $N(0.5, 1)$, we have changed the distribution from $\bar{H}_{1,t}|\theta^*$ to $\bar{H}_{1,t}|\theta^{'}$ while keeping the other $\bar{H}_{j,t} (j \neq 1)$ unchanged. Furthermore, the expected Hamming distance of this coupling is 1, since only client 1's data has been changed.

By leveraging the private Le Cam's method (Theorem 1 in (Acharya et al., 2021)),

$$\inf_{\substack{\hat{\theta}_{1,t} \in \mathcal{F}(\bar{\mathcal{I}}, \Theta) \\ \{\bar{R}_i\}_{i\in[0,M]}}} \mathbb{P}\left(e_1^\intercal \theta^* \neq \hat{\theta}_{1,t}(\bar{q}_{\leq t})\right) \geq 0.9e^{-\epsilon} - 10\delta.$$

Therefore, we have

$$\text{Regret}(M, T) \geq T\left(0.9e^{-\epsilon} - 10\delta\right).$$

If $\epsilon < 1$ and $\delta < 0.001$, we conclude that $\text{Regret}(M, T) = \Omega(T)$.

□

E. Proof of the Regret Lower Bounds Under User-level CDP Constraint

In this section, we provide the full analysis for CDP lower bounds. The first subsection describes a hard instance of linear contextual bandits model that will be used in the proofs. The remaining subsections provide the proofs of Theorem 4.2 and Theorem 4.5.

E.1. General Setting of Hard Instance

First, we introduce the notation of truncated Gaussian distributions. If a Gaussian random variable $X \sim N(0, I_d)$ is truncated to $\{x : \|x\|_2 \leq r\}$, then we denote the truncated Gaussian distribution of X as $N(0, I_d|r)$.

In the lower bound analysis, we follow the setting in He et al. (2022b), as specified below.

Arms and Dimension: There are 2 arms: $\{1, 2\}$, and the dimension d is an even number.

Feature Vectors and the Context Distribution: The feature vector of the second arm is always 0. For the feature vector of the first arm, let the distribution of the context c_i for any client i satisfy $\phi(c_i, 1) = (0, \ldots, z_{i,s}^\intercal, \ldots, 0)^\intercal$ with s uniformly distributed over $[d/2]$, and $\{z_{i,s}\}_{i,s} \subset \mathbb{R}^2$ independently sampled from a truncated normal $N(0, I_2|1)$.

Model Parameter and Its Distribution: The model parameter $\theta^* = (\theta_1^{*\intercal}, \ldots, \theta_{d/2}^{*\intercal})^\intercal \in \mathbb{R}^d$ with each $\theta_s^* \in \mathbb{R}^2$ sampled independently and uniformly from a sphere $\mathbb{S}_r = \{x \in \mathbb{R}^2 : \|x\| = r\}$, where $r \in [0, 1/\sqrt{d}]$. The constraint on r is due to the boundness assumption that $\|\theta^*\| \leq 1$. Moreover, the parameter C_0 defined in Assumption 3.2 satisfies $C_0 = \Omega(1/r)$.

Notations of Available Information Used to Make Decisions: Recall that $H_{i,t} = \{c_{i,\tau}, a_{i,\tau}, r_{i,\tau}\}_{\tau<t}$ is the history of client i. Note that $r_{i,t}$ is sampled from a Gaussian distribution with mean $\phi(c_{i,t}, a_{i,t})^\intercal \theta^*$ and variance 1, if the true model is θ^*. We further denote that $\phi(c_{i,t}, a) = x_{i,t,a}$.

To reduce the dependency between histories of different clients, we define $\bar{H}_{i,t} = \{c_{i,\tau}, a = \{1, 2\}, r_{i,\tau,a}\}_{\tau<t} \supset H_{i,t}$ to be the extended history of client i at time t, where $r_{i,t,a}$ is the (virtual) reward sampled from (un-played) pulled arm, such that $\bar{H}_{i,t}$ provides full information. It is worth pointing out that $\bar{H}_{i,t}$ is independent with $\bar{H}_{j,t}$ conditioned on model parameter θ^*. With these notations, we introduce $\bar{q}_{i,\leq t}$ and $\bar{q}_{\leq t}$, which are outputs from the "extended" DP channels $\bar{R}_i$ and $\bar{R}_0$ with inputs $\bar{H}_{i,t}$ and $\{\bar{q}_{i,t}\}_{i\in[M]}$, respectively. Here, "extended channels" implies that $\bar{R}_i(H_{i,t}) = R_i(H_{i,t})$ and $\bar{R}_0(\{q_{i,\leq t}\}_i) = R_0(\{q_{i,\leq t}\}_i)$.

With the general setting described above, we present the generic regret lower bound modified from Proposition 3.5 in He et al. (2022b). Note that while the original result holds for the single-client setting, it is straightforward to extend the result into the federated setting.

Theorem E.1. *For the hard instance described in Appendix E.1, the total regret of M clients can be lower bounded by*

$$\Omega\left(\sum_{i\in[M],t\in[T],s\in[d/2]} \inf_{\substack{\theta_{i,t,s}\in\mathcal{F}(\bar{\mathcal{I}}_i,\mathbb{S}_r)\\ \{\bar{R}_i\}_{i\in[0,M]}}} \frac{1}{rd}\mathbb{E}_v\left[\|\theta_s^* - \theta_{i,t,s}\|^2\right]\right),$$

where $\bar{\mathcal{I}}_i$ is a set of all possible information $(\bar{H}_{i,t}, \bar{q}_{\leq t})$ provided to client i.

Note that θ_s^* are independently and identically distributed, and $\{\theta_{i,t,s}\}_s$ are from the same set of measurable functions. Without loss of generality, it suffices to lower bound the estimation error for the first parameter θ_1^*, which leads to the following corollary.

Corollary E.2. *Under the same setting in Theorem E.1, the total regret of all clients can be lower bounded by*

$$\Omega\left(\sum_{i\in[M],t\in[T]} \inf_{\substack{\theta_{i,t}\in\mathcal{F}(\bar{\mathcal{I}}_i,\mathbb{S}_r)\\ \{\bar{R}_i\}_{i\in[0,M]}}} \frac{1}{2r}\mathbb{E}_v\left[\|\theta_1^* - \theta_{i,t}\|^2\right]\right),$$

where $r \in [0, 1/\sqrt{d}]$.

Corollary E.2 suggests that it suffices to lower bound the estimation error for each single time step t and client i, where the estimator $\theta_{i,t}$ is constructed from both the local data $\bar{H}_{i,t}$ and global information $\bar{q}_{\leq t}$. Since $q_{\leq t}$ is a private output from an (ε, δ)-CDP mechanism, by the post-processing property, $\theta_{i,t}$ is also an (ε, δ)-CDP estimator. However, we have to be cautious that it is differentilly private only with respect to client j's local data, where $j \neq i$. Mathematically, for any subset $S \subset \mathbb{S}_r$ and j-neighboring dataset H_t, H_t', where $j \neq i$, we have

$$\mathbb{P}\left(\theta_{i,t} \in S|H_t\right) \leq e^\varepsilon \mathbb{P}\left(\theta_{i,t} \in S|H_t'\right) + \delta.$$

E.2. Proof of Theorem 4.5

Equipped with Corollary E.2, we are able to lower bound the regret by the estimation error.

Note that both true parameter and the estimator are two dimensional vectors with constant norm. We point out that our setting is different from other works on lower bound of estimation error where each coordinate of the true parameter is sampled from an interval independently (Levy et al., 2021; Kamath et al., 2019). Hence, we re-parameterize θ_1^* by its angle, i.e. $\theta_1^* = r(\cos\gamma^*, \sin\gamma^*)^\intercal$, and γ^* is sampled uniformly from the interval $[0, 2\pi)$. We further denote $e_1 = (1, 0)$ and $e_2 = (0, 1)$ that form the canonical basis of $\mathbb{R}^2$.

Proof Outline: **Step 1** is to decompose the estimation error to the expectations of M random variables $\{Z_i\}_i$ which capture the covariance of the global estimator and local data of each client i. **Step 2** upper bounds $\mathbb{E}[Z_i]$ for all $i \in [M]$ under the CDP constraint, indicating that the estimation error is bounded from below. **Step 3** combines the previous steps to prove the final regret lower bound.

To simplify notations, in Step 1 and Step 2, we fix a time step t and a client i_0, and aim to bound the estimation error $\mathbb{E}\left[\left\|\theta_1^* - \hat{\theta}\right\|^2\right]$, where $\hat{\theta}$ is the optimal solution of $\inf_{\substack{\hat{\theta}\in\mathcal{F}(\bar{\mathcal{I}}_{i_0}, \mathbb{S}_r) \\ \{\bar{R}_i\}_{i\in[0,M]}}} \mathbb{E}\left[\left\|\theta_1^* - \hat{\theta}\right\|^2\right]$.

Step 1: Decompose the Estimation Error.

We note that similar result is obtained when each coordinate of θ^* is independently sampled. (See Lemma 6.8 in (Kamath et al., 2019) and Lemma 3.6 in (Bun et al., 2017).)

Lemma E.3 (Fingerprinting Lemma[2])**.** *Define random variables Z_i for each i as follows.*

$$Z_i = (\hat{\theta} - \theta_1^*)^\intercal(-e_1\sin\gamma^* + e_2\cos\gamma^*)(-e_1\sin\gamma^* + e_2\cos\gamma^*)^\intercal \bar{V}_i(\bar{\theta}_i - \theta^*),$$

where $\bar{V}_i = \sum_{\tau<t} x_{i,\tau,1}x_{i,\tau,1}^\intercal$, and $\bar{\theta}_i = \bar{V}_i^\dagger\left(\sum_{\tau<t} x_{i,\tau,1}r_{i,\tau,1}\right)$, and recall that $r_{i,\tau,1}$ is sampled from $N(x_{i,\tau,1}^\intercal\theta_1^*, 1)$.

Then, we have

$$\mathbb{E}\left[\left\|\theta_1^* - \hat{\theta}\right\|^2\right] = 2r^2 - 2r^2\sum_i \mathbb{E}[Z_i].$$

Proof. Due to $\|\theta_1^*\| = \|\hat{\theta}\| = r$, it suffices to analyze the term $\mathbb{E}\left[\hat{\theta}^\intercal\theta_1^*\right]$. Note that $\theta_1^* = r(\cos\gamma^*, \sin\gamma^*)^\intercal$. Then, we have

$$\begin{aligned}
\mathbb{E}\left[\hat{\theta}^\intercal\theta_1^*\right] &= \frac{r}{2\pi}\int_0^{2\pi} e_1^\intercal\mathbb{E}[\hat{\theta}|\gamma^*]\cos\gamma^* + e_2^\intercal\mathbb{E}[\hat{\theta}|\gamma^*]\sin\gamma^* d\gamma^* \\
&= \frac{r}{2\pi}\left(e_1^\intercal\mathbb{E}[\hat{\theta}|\gamma^*]\sin\gamma^* - e_2^\intercal\mathbb{E}[\hat{\theta}|\gamma^*]\cos\gamma^*\right)\Big|_{\gamma^*=0}^{\gamma^*=2\pi} \\
&\quad - \frac{r}{2\pi}\int_0^{2\pi} e_1^\intercal\frac{\partial}{\partial\gamma^*}\mathbb{E}[\hat{\theta}|\gamma^*]\sin\gamma^* + e_2^\intercal\frac{\partial}{\partial\gamma^*}\mathbb{E}[\hat{\theta}|\gamma^*]\cos\gamma^* d\gamma^* \\
&= r\mathbb{E}_{\gamma^*}\left[(-e_1\sin\gamma^* + e_2\cos\gamma^*)^\intercal\frac{\partial}{\partial\gamma^*}\mathbb{E}[\hat{\theta}|\gamma^*]\right].
\end{aligned}$$

For the derivative, it is worth noting that $\mathbb{E}[\hat{\theta}|\gamma^*] = \mathbb{E}\left[\mathbb{E}\left[\hat{\theta}\middle|\{\bar{H}_{i,t}\}_{i\in[M]}\right]\middle|\gamma^*\right]$. We have

$$\begin{aligned}
\frac{\partial}{\partial\gamma^*}\mathbb{E}[\hat{\theta}|\gamma^*] &= \int_{\{\bar{H}_{i,t}\}_i} \mathbb{E}\left[\hat{\theta}\middle|\{\bar{H}_{i,t}\}_i\right]\frac{1}{(2\pi)^{M(t-1)/2}}\frac{\partial}{\partial\gamma^*}\exp\left(-\frac{1}{2}\sum_{i\in[M],\tau<t}\left(r_{i,\tau,1} - x_{i,\tau,1}^\intercal\theta^*\right)^2\right) \\
&= r\mathbb{E}\left[\mathbb{E}\left[\hat{\theta}\middle|\{\bar{H}_{i,t}\}_i\right](-e_1\sin\gamma^* + e_2\cos\gamma^*)^\intercal\sum_{i,\tau<t} x_{i,\tau,1}(r_{i,\tau,1} - x_{i,\tau,1}^\intercal\theta^*)\middle|\theta^*\right] \\
&= r\mathbb{E}\left[\hat{\theta}(-e_1\sin\gamma^* + e_2\cos\gamma^*)^\intercal\sum_i \bar{V}_i(\bar{\theta}_i - \theta^*)\middle|\theta^*\right].
\end{aligned}$$

[2]We adopt the fingerprinting lemma rather than DP Assouad's method (Acharya et al., 2021), because in general, the lower bound obtained by DP Assouad's method has an additional blow-up factor $\sqrt{d}$ compared to that obtained from the fingerprinting lemma.

Combining with the fact that $\mathbb{E}[\bar{V}_i(\bar{\theta}_i - \theta^*)|\theta^*, \bar{V}_i] = 0$, we have

$$\begin{aligned}\mathbb{E}\left[\hat{\theta}^\intercal \theta_1^*\right] &= r^2 \mathbb{E}\left[(-e_1 \sin\gamma^* + e_2\cos\gamma^*)^\intercal \left(\hat{\theta} - \theta^*\right)(-e_1\sin\gamma^* + e_2\cos\gamma^*)^\intercal \sum_i \bar{V}_i(\bar{\theta}_i - \theta^*)\right] \\ &= r^2 \sum_i \mathbb{E}[Z_i].\end{aligned}$$

Therefore,

$$\mathbb{E}\left[\left\|\theta_1^* - \hat{\theta}\right\|^2\right] = 2r^2 - 2\mathbb{E}\left[\hat{\theta}^\intercal\theta_1^*\right] = 2r^2 - 2r^2\sum_i \mathbb{E}[Z_i],$$

which completes the proof.

□

Step 2: Upper Bound Each $\mathbb{E}[Z_i]$ under the CDP Constraint.

Lemma E.4. *Under the same setting as in Lemma E.3, if the federated algorithm satisfies user-level (ε, δ)-CDP, we have*

$$\mathbb{E}[Z_j] \le (e^\varepsilon - 1)\sqrt{\frac{2(t-1)}{d}\mathbb{E}\left[\|\hat{\theta} - \theta^*\|^2\right]} + 6r\delta\sqrt{2(t-1)}\log(1/\delta), \quad \forall j \neq i_0, \tag{5}$$

$$\mathbb{E}[Z_{i_0}] \le \sqrt{\frac{2(t-1)}{d}\mathbb{E}\left[\|\hat{\theta} - \theta^*\|^2\right]}. \tag{6}$$

Proof. Recall that

$$Z_i = (\hat{\theta} - \theta_1^*)^\intercal(-e_1\sin\gamma^* + e_2\cos\gamma^*)(-e_1\sin\gamma^* + e_2\cos\gamma^*)^\intercal \sum_{\tau<t} x_{i,\tau,1}(r_{i,\tau,1} - x_{i,\tau,1}^\intercal\theta^*),$$

where $r_{i,\tau,1}$ is sampled from $N(x_{i,\tau,1}^\intercal\theta_1^*, 1)$, and $x_{i,\tau,1} = (0, \ldots, z_{i,\tau,s}, \ldots, 0)$ with probability $2/d$, and $z_{i,\tau,s}$ is sampled independently from a truncated normal $N(0, I_2|1)$.

We have

$$\begin{aligned}\mathbb{E}[Z_{i_0}]^2 &\le \mathbb{E}\left[\|\hat{\theta} - \theta_1^*\|^2\right]\mathbb{E}\left[\left(\left(-e_1\sin\gamma^* + e_2\cos\gamma^*\right)^\intercal \sum_{\tau<t} x_{i,\tau,1}(r_{i,\tau,1} - x_{i,\tau,1}^\intercal\theta^*)\right)^2\right] \\ &= \mathbb{E}\left[\|\hat{\theta} - \theta_1^*\|^2\right]\mathbb{E}\left[\sum_{\tau<t}\left(\left(-e_1\sin\gamma^* + e_2\cos\gamma^*\right)^\intercal x_{i,\tau,1}\right)^2\right] \\ &= \frac{2(t-1)}{d}\mathbb{E}\left[\|\hat{\theta} - \theta_1^*\|^2\right],\end{aligned}$$

which verifies the second Equation (6) of Lemma E.4.

To prove the first part of Lemma E.4, the approach is akin to Kamath et al. (2019). We introduce a statistically indistinguishable random variable $\tilde{Z}_j$ for each Z_j, $j \neq i$. Let $\bar{H}'_{j,t}$ be sampled independently and identically with $\bar{H}_{j,t}$, and $\hat{\theta}^{-j}$ be the estimator constructed from $\{\bar{H}_{i,t}\}_{i\neq j} \cup \bar{H}'_{j,t}$. By the definition of CDP, $\hat{\theta}^{-j}$ is statistically indistinguishable compared with $\hat{\theta}$. Then, define

$$\tilde{Z}_j = (\hat{\theta}^{-j} - \theta_1^*)^\intercal(-e_1\sin\gamma^* + e_2\cos\gamma^*)(-e_1\sin\gamma^* + e_2\cos\gamma^*)^\intercal\bar{V}_j(\bar{\theta}_j - \theta^*). \tag{7}$$

We have several properties about $\tilde{Z}_j$. First, due to the independence between $H_{j,t}$ and $H'_{j,t}$, the expectation of $\tilde{Z}_j$ is 0, i.e. $\mathbb{E}[\tilde{Z}_j] = 0$.

Second, the variance of $\tilde{Z}_j$ is upper bounded by the estimation error, because

$$\begin{aligned}
\mathbb{E}[\tilde{Z}_j^2] &\overset{(a)}{\leq} \mathbb{E}\left[\|\hat{\theta}^{-j} - \theta_1^*\|^2 \left((-e_1 \sin\gamma^* + e_2 \cos\gamma^*)^\intercal \bar{V}_j(\bar{\theta}_j - \theta^*)\right)^2\right] \\
&\overset{(b)}{=} \mathbb{E}\left[\mathbb{E}\left[\|\hat{\theta}^{-j} - \theta_1^*\|^2\right] \mathbb{E}\left[\left((-e_1 \sin\gamma^* + e_2 \cos\gamma^*)^\intercal \bar{V}_j(\bar{\theta}_j - \theta^*)\right)^2\right] \middle| \theta_1^*\right] \\
&\overset{(c)}{\leq} \frac{2(t-1)}{d} \mathbb{E}\left[\|\hat{\theta} - \theta_1^*\|^2\right],
\end{aligned}$$

where (a) follows from the Cauchy's inequality, and $(b), (c)$ are due to the fact that $H'_{j,t}$ and $H_{j,t}$ are IID sampled. Then, by choosing a threshold $Z > 0$ which will be specified later, we can bound $\mathbb{E}[Z_j]$ as follows.

$$\begin{aligned}
&\mathbb{E}\left[Z_j\right] = \mathbb{E}\left[Z_j\right] - \mathbb{E}[\tilde{Z}_j] \\
&\overset{(a)}{=} \mathbb{E}\left[\mathbb{E}\left[\int_0^{+\infty} \left[\mathbb{P}\left(Z_j > z | \{\bar{H}_{i,t}\}_i\right) - \mathbb{P}\left(\tilde{Z}_j > z | \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right)\right] dz \middle| \theta_1^*, \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right]\right] \\
&\quad - \mathbb{E}\left[\mathbb{E}\left[\int_{-\infty}^{0} \left[\mathbb{P}\left(Z_j < z | \{\bar{H}_{i,t}\}_i\right) - \mathbb{P}\left(\tilde{Z}_j < z | \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right)\right] dz \middle| \theta_1^*, \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right]\right] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[\int_0^{Z} \left[\mathbb{P}\left(Z_j > z | \{\bar{H}_{i,t}\}_i\right) - \mathbb{P}\left(\tilde{Z}_j > z | \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right)\right] dz \middle| \theta_1^*, \{\bar{H}_{i,t}\}_i, H'_{j,t}\right]\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[\int_Z^{+\infty} \mathbb{P}\left(Z_j > z | \{\bar{H}_{i,t}\}_i\right) dz \middle| \theta_1^*, \{\bar{H}_{i,t}\}_i\right]\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[\int_{-Z}^{0} \left[\mathbb{P}\left(\tilde{Z}_j < z | \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right) - \mathbb{P}\left(Z_j < z | \{\bar{H}_{i,t}\}_i\right)\right] dz \middle| \theta_1^*, \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right]\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[\int_{-\infty}^{-Z} \mathbb{P}\left(\tilde{Z}_j < z | \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right) dz \middle| \theta_1^*, \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right]\right] \\
&\overset{(b)}{\leq} Z\delta + (e^\varepsilon - 1)\mathbb{E}\left[\mathbb{E}\left[\int_0^{Z} \mathbb{P}\left(\tilde{Z}_j > z | \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right) dz \middle| \theta_1^*, \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right]\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[\int_Z^{+\infty} \mathbb{P}\left(Z_j > z | \{\bar{H}_{i,t}\}_i\right) dz \middle| \theta_1^*, \{\bar{H}_{i,t}\}_i\right]\right] \\
&\quad + Z\delta + (1 - e^{-\varepsilon})\mathbb{E}\left[\mathbb{E}\left[\int_{-Z}^{0} \mathbb{P}\left(\tilde{Z}_j < z | \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right) dz \middle| \theta_1^*, \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right]\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[\int_{-\infty}^{-Z} \mathbb{P}\left(\tilde{Z}_j < z | \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right) dz \middle| \theta_1^*, \{\bar{H}_{i,t}\}_i, \bar{H}'_{j,t}\right]\right] \\
&\overset{(c)}{\leq} 2Z\delta + (e^\varepsilon - 1)\mathbb{E}\left[|\tilde{Z}_j|\right] \\
&\quad + \mathbb{E}\left[\mathbb{E}\left[\int_Z^{+\infty} \mathbb{P}\left(Z_j > z\right) dz \middle| \theta_1^*\right]\right] + \mathbb{E}\left[\mathbb{E}\left[\int_{-\infty}^{-Z} \mathbb{P}\left(\tilde{Z}_i < z\right) dz \middle| \theta_1^*\right]\right] \\
&\overset{(d)}{\leq} (e^\varepsilon - 1)\sqrt{\frac{2(t-1)}{d} \mathbb{E}\left[\|\hat{\theta} - \theta_1^*\|^2\right]} + 2Z\delta \\
&\quad \mathbb{E}\left[\mathbb{E}\left[\int_Z^{+\infty} \mathbb{P}\left(Z_j > z\right) dz \middle| \theta_1^*\right]\right] + \mathbb{E}\left[\mathbb{E}\left[\int_{-\infty}^{-Z} \mathbb{P}\left(\tilde{Z}_i < z\right) dz \middle| \theta_1^*\right]\right],
\end{aligned}$$

where (a) follows from $\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}(X > x)dx - \int_{-\infty}^0 \mathbb{P}(X < x)dx$, (b) follows from (ε, δ)-CDP, (c) is due to $1 - e^{-\varepsilon} \leq e^\varepsilon - 1$, and (d) is due to the Cauchy's inequality.

For the term $\int_Z^{+\infty} \mathbb{P}\left(Z_j > z|\theta_1^*\right) dz$, we can bound it as

$$\int_Z^{+\infty} \mathbb{P}\left(Z_j > z|\theta_1^*\right) dz \leq \int_Z^{+\infty} \mathbb{P}\left(2r\left|(-e_1\sin\gamma^* + e_2\cos\gamma^*)^\intercal \bar{V}_j(\bar{\theta}_j - \theta^*)\right| > z\middle|\theta^*\right) dz.$$

Note that $\bar{V}_j\bar{\theta}_j \sim N(\bar{V}_i\theta_1^*, \bar{V}_i)$ conditioned on $(\theta_1^*, \bar{V}_j)$. If we denote $W_j = (-e_1\sin\gamma^* + e_2\cos\gamma^*)^\intercal \bar{V}_j(\bar{\theta}_j - \theta^*)$, then $W_j \sim N(0, \|(-e_1\sin\gamma^* + e_2\cos\gamma^*)\|_{\bar{V}_j}^2)$ and is conditionally independent with other $\bar{\theta}_i$, where $i \neq j$. Hence,

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left[\int_Z^{+\infty} \mathbb{P}\left(Z_j > z|\{\bar{V}_j\}\right) dz\middle|\theta^*\right]\right] \\
&\leq \mathbb{E}\left[\int_Z^{+\infty} \mathbb{P}\left(2rW_j > z|\theta^*, \bar{V}_i\right) dz\right] \\
&= \int_Z^{+\infty} \int_{\frac{z}{2r\|(-e_1\sin\gamma^* + e_2\cos\gamma^*)\|_{\bar{V}_j}}}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right) dxdz \\
&\leq \mathbb{E}\left[\int_{\frac{Z}{2r\|(-e_1\sin\gamma^* + e_2\cos\gamma^*)\|_{\bar{V}_j}}} \frac{2r\|(-e_1\sin\gamma^* + e_2\cos\gamma^*)\|_{\bar{V}_j} \cdot x}{\sqrt{2\pi}} \exp(-x^2/2)dx\right] \\
&= \mathbb{E}\left[\frac{2r\|(-e_1\sin\gamma^* + e_2\cos\gamma^*)\|_{\bar{V}_j} \exp(-\frac{Z^2}{8r^2\|(-e_1\sin\gamma^* + e_2\cos\gamma^*)\|_{\bar{V}_j}^2})}{\sqrt{2\pi}}\right] \\
&\leq 2r\sqrt{(t-1)/\pi}\exp\left(-\frac{Z^2}{8r^2(t-1)}\right),
\end{aligned}$$

where the last inequality again follows from $\|-e_1\sin\gamma^* + e_2\cos\gamma^*\|_{\bar{V}_j} \leq \sqrt{t-1}$.

Following the same reason, we also have

$$\int_{-\infty}^{-Z} \mathbb{P}\left(\tilde{Z}_i < z\right) dz \leq 2r\sqrt{(t-1)/\pi}\exp\left(-\frac{Z^2}{8r^2(t-1)}\right).$$

Thus, we conclude that

$$\mathbb{E}[Z_j] \leq (e^\varepsilon - 1)\sqrt{\frac{2(t-1)}{d}\mathbb{E}\left[\|\hat{\theta} - \theta_1^*\|^2\right]} + 2Z\delta + 4r\sqrt{(t-1)/\pi}\exp\left(-\frac{Z^2}{8r^2(t-1)}\right).$$

We finish the proof by choosing $Z = 2\sqrt{2(t-1)}r\log(1/\delta)$.

□

Step 3: Lower Bound the Total Regret.

Theorem E.5 (Restatement of Theorem 4.5)**.** *Fix any $\varepsilon \in (0, \log 2)$, $\delta = \tilde{O}\left(\frac{1}{M\sqrt{T}}\right)$, $T \geq d^2$. Then, there exists a federated linear contextual bandits instance satisfying Assumptions 3.1 and 3.2 such that any* **with-memory** *federated algorithm satisfying user-level (ε, δ)-CDP must incur a regret lower bounded by*

$$\Omega\left(\min\left\{M, \max\left\{1, \frac{1}{M\varepsilon^2}\right\}\right\} C_0 d\log T\right).$$

If Assumption 3.2 is not satisfied, then the minimax regret lower bound becomes

$$\Omega\left(\min\left\{M, \max\left\{\sqrt{M}, \tfrac{1}{\varepsilon}\right\}\right\}\sqrt{dT}\right).$$

Proof. Combine Step 1 (Lemma E.3) and Step 2 (Lemma E.4), we have,

$$\begin{aligned}2r^2 &= \mathbb{E}\left[\|\hat{\theta}-\theta^*\|^2\right] + 2r^2\sum_i \mathbb{E}\left[Z_i\right] \\ &\le \mathbb{E}\left[\|\hat{\theta}-\theta^*\|^2\right] + 2\left((e^{\varepsilon}-1)(M-1)+1\right) r^2\sqrt{\frac{2(t-1)}{d}\mathbb{E}\left[\|\hat{\theta}-\theta^*\|^2\right]} + 12r^3\delta M\sqrt{2(t-1)}\log(1/\delta).\end{aligned}$$

Consider the case when $\varepsilon \le \log 2$, $\delta \le \frac{\sqrt{\pi}}{12rM\sqrt{2T\log(1/\delta)}}$, we further have

$$r^2 \le \mathbb{E}\left[\|\hat{\theta}-\theta^*\|^2\right] + 2r^2(\varepsilon(M-1)+1)\sqrt{\frac{2(t-1)}{d}\mathbb{E}\left[\|\hat{\theta}-\theta^*\|^2\right]}.$$

Therefore,

$$\mathbb{E}\left[\|\hat{\theta}-\theta^*\|^2\right] \ge \frac{r^2}{8r^2(\varepsilon(M-1)+1)^2(t-1)/d+4}.$$

Substituting the above result into the generic lower bound Corollary E.2, we conclude that

$$\begin{aligned}\text{Regret}(M,T) &\ge \Omega\left(\sum_{i\in[M],t\in[T]}\frac{1}{r}\frac{r^2}{32r^2\max\{\varepsilon^2M^2,1\}(t-1)/d+4}\right) \\ &\ge \Omega\left(\frac{dM}{r\max\{\varepsilon^2M^2,1\}}\log\left(1+r^2\max\{\varepsilon^2M^2,1\}T/d\right)\right).\end{aligned}$$

The rest of the proof consists of two parts, in which two suitable r's are chosen such that we can obtain the desired regret bounds.

First, under the margin condition, note that our setting indicates $C_0 = \Omega(1/r)$. Thus, under the margin condition in Assumption 3.2, we have

$$\text{Regret}(M,T) \ge \Omega\left(\min\left\{M, \frac{1}{\varepsilon^2 M}\right\} C_0 d\log T\right).$$

Thus, the first lower bound is obtained by combining the regret lower bound $\Omega(C_0 d\log T)$ of non-private case (Proposition G.6).

Then, we select $r = \frac{\sqrt{d}}{\max\{\varepsilon M,1\}\sqrt{T}}$, where we require that $T > d^2$. We obtain a worst-case lower bound as follows

$$\text{Regret}(M,T) \ge \Omega\left(\min\left\{M, \frac{1}{\varepsilon}\right\}\sqrt{dT}\right). \tag{8}$$

We finally finish the proof by combining with the non-private regret lower bound $\Omega(\sqrt{dMT})$.

□

E.3. Proof of Theorem 4.2

In this section, we leverage the previous analysis to prove the result for regret lower bound under the CDP constraint, almost-memoryless setting, and the margin condition in Assumption 3.2.

The hard instance is defined the same as that in Appendix E.1. We point out that the only difference of memoryless case is that the estimator $\theta_{i,t}$ is a private estimator with respect to all clients, including itself.

Theorem E.6 (Restatement of Theorem 4.2)**.** *If $\varepsilon < \log 2$, $\delta = \tilde{O}(\frac{1}{M\sqrt{T}})$, then, there exists a federated linear contextual bandits instance satisfying Assumptions 3.1 and 3.2, such that any* **almost-memoryless** *federated algorithm satisfying user-level (ε,δ)-CDP must incur a regret lower bounded by*

$$\Omega\left(\max\left\{1, \tfrac{1}{M\varepsilon^2}\right\} C_0 d\log T + e^{-M\varepsilon} C_0 MT\right).$$

Proof. According to Corollary E.2, and Definition 4.1 of almost-memoryless algorithms, it suffices to analyze the estimation error

$$\inf_{\substack{\theta_{i_0,t}\in\mathcal{F}(\bar{\mathcal{I}},\mathbb{S}_r)\\ \{\bar{R}_i\}_{i\in[0,M]}}} \mathbb{E}_v\left[\|\theta_1^*-\theta_{i_0,t}\|^2\right]$$

for a "memoryless" time t and client i_0, where $\bar{\mathcal{I}}$ is the set of all possible $q_{\le t}$. By the post-processing property, $\theta_{i,t}$ is a (ε,δ)-CDP estimator. Mathematically, for any subset $S\subset\mathbb{S}_r$ and j-neighboring dataset H_t, H_t', where $j\in[M]$, we have

$$\mathbb{P}\left(\theta_{i_0,t}\in S|H_t\right)\le e^{\varepsilon}\mathbb{P}\left(\theta_{i_0,t}\in S|H_t'\right)+\delta.$$

The proof follows the same argument as in Appendix E.2, where we have three steps. The first step remains the same, where we construct M random variables $\{Z_i\}$ (defined in Lemma E.3), and show that

$$\mathbb{E}[\|\hat{\theta}-\theta_1^*\|^2]=2r^2-2r^2\sum_i\mathbb{E}[Z_i],$$

where $\hat{\theta}=\arg\inf_{\substack{\theta_{i_0,t}\in\mathcal{F}(\bar{\mathcal{I}},\mathbb{S}_r)\\ \{\bar{R}_i\}_{i\in[0,M]}}}\mathbb{E}_v\left[\|\theta_1^*-\theta_{i_0,t}\|^2\right]$.

The second step is almost the same as in Lemma E.4, except that we can upper bound $\mathbb{E}[Z_{i,0}]$ using the same inequality in Equation (5).

Therefore, we obtain,

$$\begin{aligned}2r^2&=\mathbb{E}\left[\|\hat{\theta}-\theta^*\|^2\right]+2r^2\sum_i\mathbb{E}\left[Z_i\right]\\&\le\mathbb{E}\left[\|\hat{\theta}-\theta^*\|^2\right]+2(e^{\varepsilon}-1)Mr^2\sqrt{\frac{2(t-1)}{d}\mathbb{E}\left[\|\hat{\theta}-\theta^*\|^2\right]}+12r^3\delta M\sqrt{2(t-1)}\log(1/\delta).\end{aligned}$$

By choosing $\varepsilon<\log 2$, and $\delta<\frac{1}{12rM\sqrt{2T}\log(1/\delta)}$, we have

$$\mathbb{E}\left[\|\hat{\theta}-\theta^*\|^2\right]\ge\frac{r^2}{8r^2\varepsilon^2M^2(t-1)/d+4}.$$

Substituting the above inequality into the generic lower bound in Corollary E.2, and noting that $r=O(1/C_0)$, we conclude that

$$\text{Regret}(M,T)\ge\Omega\left(\frac{C_0d\log T}{\varepsilon^2M}\right).$$

To derive the second term $e^{-\varepsilon M}MT$ in the regret lower bound, we directly analyze the estimation error as follows

$$\begin{aligned}\mathbb{E}[\|\hat{\theta}-\theta_1^*\|^2]&=\mathbb{E}\left[\int_0^{4r^2}\mathbb{P}\left(\|\hat{\theta}-\theta^*\|^2>z|\{\bar{H}_{i,t}\}_i\right)\right]\\&\overset{(a)}{\ge}e^{-M\varepsilon}\mathbb{E}\left[\int_0^{4r^2}\mathbb{P}\left(\|\hat{\theta}-\theta^*\|^2>z\big|\{\bar{H}_{i,t}\}_i=\{x_{i,\tau,a}=0,r_{i,\tau,a}=0\}_{i,\tau}\right)\right]-4r^2M\delta\\&=e^{-M\varepsilon}\mathbb{E}\left[\|\hat{\theta}^0-\theta^*\|^2\right]-4r^2M\delta,\\&\overset{(b)}{\ge}e^{-M\varepsilon}\left(r^2-2\mathbb{E}[\hat{\theta}^{0\intercal}\theta_1^*]\right)-4r^2\delta\\&=e^{-M\varepsilon}r^2-4r^2M\delta,\end{aligned}$$

where (a) is due to the CDP constraint, $\hat{\theta}^0$ is the output from a fixed "zero" dataset, which is independent with θ^*, and (b) follows from the independency, and $\mathbb{E}[\theta^*] = 0$.

Note that $r = O(1/C_0)$. By choosing $\delta < O(\frac{1}{dM^2T})$, we have

$$\text{Regret}(M, T) \geq \Omega\left(e^{-M\varepsilon} C_0 MT - 1\right).$$

We finish the proof by noting that the non-private regret lower bound is $\Omega(C_0 d \log T)$ under the margin condition, according to Proposition G.6.

□

F. Proof of the Regret Lower Bounds Under User-level LDP Constraint

In this section, we provide the proof for regret lower bounds under the user-level LDP constraint. It consists of two subsections. The first subsection lists several general lemmas, which are used to bound the total variation distance between multivariate distributions. The second subsection provides the full proof of the lower bounds.

F.1. Useful Lemmas for the Proof of Theorem 5.3

We first introduce a lemma that bounds the divergence of the output distributions from a DP channel with different input distributions.

Lemma F.1 (Adapted from Lemma 2 in Asoodeh et al. (2021)). *If $q_{i,\leq t}$ is the output of an (ε, δ)-LDP channel $\mathtt{R}_i$ with input $H_{i,t}$, and $H_{i,t}$ follows a prior distribution parameterized by θ, then, for any two different θ, θ', let $\mathbb{P}(q_{i,\leq t}|\theta)$ be the marginal distribution of $q_{i,t}$, and we have*

$$\begin{aligned}
&KL(\mathbb{P}(q_{i,\leq t}|\theta), \mathbb{P}(q_{i,\leq t}|\theta')) \leq \left(1 - e^{-\varepsilon}(1-\delta)\right) KL\left(\mathbb{P}(H_{i,t}|\theta), \mathbb{P}(H_{i,t}|\theta')\right), \\
&d_{TV}(\mathbb{P}(q_{i,\leq t}|\theta), \mathbb{P}(q_{i,\leq t}|\theta')) \leq \left(1 - e^{-\varepsilon}(1-\delta)\right) d_{TV}\left(\mathbb{P}(H_{i,t}|\theta), \mathbb{P}(H_{i,t}|\theta')\right).
\end{aligned}$$

In the following, we give a tighter bound on the total variation distance of two multivariate distributions. It is crucial since the policy depends on both local data $H_{i,t}$ and global information $q_{\leq t}$. While $q_{\leq t}$ is from a DP channel, $H_{i,t}$ is a non-private information and should be analyzed separately.

First, we introduce the notion of coupling and relate the total variation distance with an error probability.

Definition F.2 (Coupling (Den Hollander, 2012)). A coupling of two random variables X, X' is any pair of random variables $(\hat{X}, \hat{X}')$ such that their marginals have the same distribution as X and X', i.e. $\hat{X} \overset{D}{=} X$, and $\hat{X}' \overset{D}{=} X'$. The law $\hat{\mathbb{P}}$ of $(\hat{X}, \hat{X}')$ is a coupling of the laws $\mathbb{P}$ and $\mathbb{P}'$ of X and X'.

Lemma F.3 (Theorem 2.4 & Theorem 2.12 in Den Hollander (2012)). *For any two probability measures $\mathbb{P}$ and $\mathbb{P}'$ on the same measurable space, any coupling $\hat{\mathbb{P}}$ satisfies*

$$d_{TV}(\mathbb{P}, \mathbb{P}') \leq \hat{\mathbb{P}}(\hat{X} \neq \hat{X}').$$

Moreover, there exists a coupling $\hat{\mathbb{P}}_0$ such that

$$d_{TV}(\mathbb{P}, \mathbb{P}') = \hat{\mathbb{P}}_0(\hat{X} \neq \hat{X}').$$

Equipped with the coupling method, we are able to upper bound the total variation distance of two multivariate distributions, as shown in the following lemma.

Lemma F.4 (Total variation distance of two multivariate distributions). *Let $\mathbb{P}$ and $\mathbb{Q}$ be two multivariate distributions defined on $\mathcal{X} \times \mathcal{Y}$, and suppose $\mathbb{P}(X, Y) = \mathbb{P}_1(X|Y)\mathbb{P}_2(Y)$, $\mathbb{Q}(X, Y) = \mathbb{Q}_1(X|Y)\mathbb{Q}_2(Y)$. Then, we have*

$$d_{TV}(\mathbb{P} - \mathbb{Q}) \leq 1 - \left(1 - \max_y d_{TV}\left(\mathbb{P}_1(\cdot|y), \mathbb{Q}_1(\cdot|y)\right)\right)\left(1 - d_{TV}\left(\mathbb{P}_2(Y), \mathbb{Q}_2(Y)\right)\right).$$

Proof. By Lemma F.3, we can find couplings $\hat{\mathbb{P}}_1(\hat{X}, \hat{X}'|\hat{Y}, \hat{Y}')$ and $\hat{\mathbb{P}}_2(\hat{Y}, \hat{Y}')$ such that

- $\hat{\mathbb{P}}_1(\hat{X}, \hat{X}'|\hat{Y}, \hat{Y}')$ is a coupling of $\mathbb{P}_1(X|Y)$ and $\mathbb{Q}_1(X'|Y')$. Moreover,

$$\hat{\mathbb{P}}_1(\hat{X} \neq \hat{X}'|\hat{Y}, \hat{Y}') = d_{TV}\left(\mathbb{P}_1(\cdot|\hat{Y}), \mathbb{Q}(\cdot|\hat{Y}')\right).$$

- $\hat{\mathbb{P}}_2(\hat{Y}, \hat{Y}')$ is a coupling of $\mathbb{P}_2(Y)$ and $\mathbb{Q}_2(Y)$. Moreover,

$$\hat{\mathbb{P}}_2(\hat{Y} \neq \hat{Y}') = d_{TV}\left(\mathbb{P}_2, \mathbb{Q}_2\right).$$

Then, if we define $\hat{\mathbb{P}}(\hat{X}, \hat{X}', \hat{Y}, \hat{Y}') = \hat{\mathbb{P}}_1(\hat{X}, \hat{X}'|\hat{Y}, \hat{Y}')\hat{\mathbb{P}}_2(\hat{Y}, \hat{Y}')$, it can be verified that $\hat{\mathbb{P}}(\hat{X}, \hat{X}', \hat{Y}, \hat{Y}')$ is a coupling of $\mathbb{P}(X, Y)$ and $\mathbb{Q}(X, Y)$, since

$$\begin{aligned}
\int_{\hat{X}', \hat{Y}'} \hat{\mathbb{P}}(\hat{X}, \hat{X}', \hat{Y}, \hat{Y}') &= \int_{\hat{Y}'} \int_{\hat{X}'} \hat{\mathbb{P}}_1(\hat{X}, \hat{X}'|\hat{Y}, \hat{Y}')\hat{\mathbb{P}}_2(Y, \hat{Y}') \\
&= \int_{\hat{Y}'} \mathbb{P}_1(\hat{X}|\hat{Y})\hat{\mathbb{P}}_2(\hat{Y}, \hat{Y}') \\
&= \mathbb{P}_1(\hat{X}|\hat{Y})\mathbb{P}_2(\hat{Y}) \\
&= \mathbb{P}(\hat{X}, \hat{Y}).
\end{aligned}$$

Then, by Lemma F.3, we have

$$\begin{aligned}
d_{TV}\left(\mathbb{P}, \mathbb{Q}\right) &\leq \hat{\mathbb{P}}((\hat{X}, \hat{Y}) \neq (\hat{X}', \hat{Y}')) \\
&= 1 - \hat{\mathbb{P}}((\hat{X}, \hat{Y}) = (\hat{X}', \hat{Y}')) \\
&= 1 - \hat{\mathbb{P}}(\hat{X} = \hat{X}', \hat{Y} = \hat{Y}') \\
&= 1 - \hat{\mathbb{P}}_1(\hat{X} = \hat{X}'|\hat{Y} = \hat{Y}')\hat{\mathbb{P}}_2(\hat{Y} = \hat{Y}') \\
&\leq 1 - \left(1 - \max_y d_{TV}\left(\mathbb{P}_1(\cdot|y), \mathbb{Q}_1(\cdot|y)\right)\right)\left(1 - d_{TV}\left(\mathbb{P}_2, \mathbb{Q}_2\right)\right),
\end{aligned}$$

which completes the proof.

□

F.2. Proof of Theorem 5.3

We follow the same setting defined in Appendix E.1. Hence, it suffices to lower bound the estimation error

$$\inf_{\substack{\theta_{i,t} \in \mathcal{F}(\bar{\mathcal{I}}_i, \mathbb{S}_r) \\ \{\bar{R}_i\}_{i \in [0,M]}}} \mathbb{E}_v\left[\|\theta_1^* - \theta_{i,t}\|^2\right],$$

for each single time step t and each client i. We emphasize that $\theta_{i,t}$ is an LDP estimator with respect to all clients $j \neq i$. Therefore, without loss of generality, we assume $\bar{q}_{\leq t} = \{\bar{q}_{i,\leq t}\}$, i.e. $\mathtt{R}_0$ is an identical map that does not perform any operation on the aggregated information.

Proof of Theorem 5.3. Recall that the DP mechanism R_i is non-interactive, i.e. $q_{i,\leq t}$ is independent with other client's data conditioned on the client i's own data.

Moreover, the full information data set $\bar{H}_{i,t}$ ($\bar{H}_{i,t}$ contains rewards sampled from all un-played arms) is independent with the global information $q_{\leq t}$ conditioned on θ_1^*. Thus, for any θ and θ',

$$\begin{aligned}
&\mathtt{KL}\left[\mathbb{P}(\bar{H}_{i,t}|\theta, q_{\leq t})\|\mathbb{P}(\bar{H}_{i,t}|\theta', q_{\leq t})\right] \\
&= \mathbb{E}\left[\log \frac{\mathbb{P}\left(\bar{H}_{i,t}|\theta\right)}{\mathbb{P}\left(\bar{H}_{i,t}|\theta'\right)}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\tau\in[t-1],a\in\{1,2\}}^{t-1} \mathbb{E}\left[\log\frac{\mathbb{P}\left(c_{i,\tau},a,r_{i,\tau,a}|\theta\right)}{\mathbb{P}\left(c_{i,\tau},a,r_{i,\tau,a}|\theta'\right)}\right]\\
&\overset{(a)}{=} \sum_{\tau\in[t-1]} \mathbb{E}\left[\log\frac{\mathbb{P}\left(r_{i,\tau,1}|\theta,x_{i,\tau,1}\right)}{\mathbb{P}\left(r_{i,\tau,a}|\theta',x_{i,\tau,1}\right)}\right]\\
&\overset{(b)}{\leq} \|\theta-\theta'\|^2(t-1)/d,
\end{aligned}
$$

where (b) follows from that only the reward of the first arm depends on the model parameter, and (a) is due to the fact that the KL-divergence of two Gaussian random variables is upper bounded by the squared difference of their expectations.

Similarly, due to $\theta^* - \{H_{i,t}\}_{i\in[M]} - \{q_{\leq t}\}$, we have

$$
\texttt{KL}\left[\mathbb{P}(H_{1,t},...,H_{M,t}|\theta)\|\mathbb{P}(H_{1,t},...,H_{M,t}|\theta')\right] \leq \|\theta-\theta'\|^2 M(t-1)/d.
$$

Then, we apply Lemma F.1 and the chain rule of KL-divergence on the total variation distance between $\mathbb{P}(\bar{q}_{\leq t}|\theta)$ and $\mathbb{P}(\bar{q}_{\leq t}|\theta')$.

$$
\begin{aligned}
&d_{TV}\left(\mathbb{P}(\bar{q}_{\leq t}|\theta),\mathbb{P}(\bar{q}_{\leq t}|\theta')\right)\\
&\quad\overset{(a)}{\leq} \sqrt{1-\exp\left(-\texttt{KL}\left[\mathbb{P}(\bar{q}_{\leq t}|\theta)\|\mathbb{P}(\bar{q}_{\leq t}|\theta')\right]\right)}\\
&\quad= \sqrt{1-\exp\left(-\sum_{i\in[M]}\texttt{KL}\left[\mathbb{P}(\bar{q}_{i,\leq t}|\theta)\|\mathbb{P}(\bar{q}_{i,\leq t}|\theta')\right]\right)}\\
&\quad\leq \sqrt{1-\exp\left(-(1-e^{-\varepsilon}(1-\delta))\sum_{i\in[M]}\texttt{KL}\left[\mathbb{P}(\bar{H}_{i,\leq t}|\theta)\|\mathbb{P}(\bar{H}_{i,\leq t}|\theta')\right]\right)}\\
&\quad\leq \sqrt{1-\exp\left(-(1-e^{-\varepsilon}(1-\delta))M\|\theta-\theta'\|^2(t-1)/d\right)},
\end{aligned}
$$

where (a) is due to the Bretagnolle–Huber inequality.

Let $\varepsilon' = (1-e^{-\varepsilon}(1-\delta))$. Based on Lemma F.4, we characterize the total variation distance of the joint distribution of local data $H_{i,t}$ and global information $\bar{q}_{\leq t}$ as follows.

$$
\begin{aligned}
&d_{TV}\left(\mathbb{P}(\bar{H}_{i,t},\bar{q}_{\leq t}|\theta),\mathbb{P}(\bar{H}_{i,t},\bar{q}_{\leq t}|\theta')\right)\\
&\quad\leq 1-\left(1-d_{TV}\left(\mathbb{P}(\bar{H}_{i,t}|\theta),Pr(\bar{H}_{i,t}|\theta')\right)\right)\left(1-d_{TV}\left(\mathbb{P}(\bar{q}_{\leq t}\big|\theta),\mathbb{P}(\bar{q}_{\leq t}\big|\theta')\right)\right)\\
&\quad\leq 1-\left(1-\sqrt{1-e^{-2\|\theta-\theta'\|^2(t-1)/d}}\right)\left(1-\sqrt{1-e^{-2\varepsilon_0\|\theta-\theta'\|^2M(t-1)/d}}\right).
\end{aligned}
$$

Now, applying the technique in Proposition 4.1 in He et al. (2022b), we have for any $\theta,\theta'\in\Theta$,

$$
\begin{aligned}
&\mathbb{P}\left(\|\theta^*-\theta_{i,t}\|_2^2 \geq \frac{1}{4}\|\theta-\theta^{'}\|_2^2|\theta^*=\theta\right)\\
&\quad\geq \frac{1}{2}\left(\left(1-d_{TV}\left(\mathbb{P}(H_{i,t},q_t\big|\theta),\mathbb{P}(H_{i,t},q_t\big|\theta')\right)\right)\right)\\
&\quad\geq \frac{1}{2}\left(1-\sqrt{1-e^{-\varepsilon'\|\theta-\theta'\|^2M(t-1)/d}}\right)\left(1-\sqrt{1-e^{-\|\theta-\theta'\|^2(t-1)/d}}\right).
\end{aligned}
$$

Therefore, if $t>1$, we have

$$
\mathbb{E}[\|\theta-\theta_{i,t}\|^2] \geq \frac{1}{2}\int_0^{r^2}\left(1-\sqrt{1-e^{-q\varepsilon'M(t-1)/d}}\right)\left(1-\sqrt{1-e^{-q(t-1)/d}}\right)dq
$$

$$
\begin{aligned}
&\geq \frac{1}{2}\int_0^{\min\{r^2, \frac{d}{\varepsilon' M(t-1)}, \frac{d}{t-1}\}} \left(1-\sqrt{q}\sqrt{\varepsilon' M(t-1)/d}\right)\left(1-\sqrt{q}\sqrt{(t-1)/d}\right) \\
&= \begin{cases} r^2\left(1+r^2/2-\frac{2}{3\sqrt{d}}r(1+M\varepsilon')\sqrt{t-1}\right), & \text{if } r^2 < \min\left\{1, \frac{1}{M\varepsilon'}\right\}\frac{d}{t-1}, \\ (2-\sqrt{M\varepsilon'})\frac{d}{t-1}, & \text{if } \varepsilon' M < 1, r^2 > \frac{d}{t-1}, \\ (2-1/\sqrt{M\varepsilon'})\frac{d}{M\varepsilon'(t-1)}, & \text{if } \varepsilon' M > 1, r^2 > \frac{d}{M\varepsilon'(t-1)}. \end{cases}
\end{aligned}
$$

Thus, the regret is bounded below by

$$
\begin{aligned}
\text{Regret}(M,T) &\geq \Omega\left(\frac{1}{r}\sum_{i,t}\mathbb{E}[\|\theta-\theta_{i,t}\|^2]\right) \\
&\geq \begin{cases} \Omega\left(rM\left(T+Tr^2/2-\frac{4}{9\sqrt{d}}r(1+M\varepsilon')T^{3/2}\right)\right), & \text{if } r^2 < \min\left\{1, \frac{1}{M\varepsilon'}\right\}\frac{d}{t-1}, \\ \Omega\left(C_0 dM\log T\right), & \text{if } \varepsilon' M < 1, r = O(1/C_0), \\ \Omega\left(\frac{C_0 d\log T}{\varepsilon'}\right), & \text{if } \varepsilon' M > 1, r = O(1/C_0). \end{cases}
\end{aligned}
$$

By selecting $r = O(\min\{1, 1/\sqrt{M\varepsilon'}\})\sqrt{d/T}$, and noting that $\varepsilon' = O(\varepsilon)$ when $\varepsilon < \log 2, \delta < 0.1$ we obtain two lower bounds:

$$
\text{Regret}(M,T) \geq \begin{cases} \Omega\left(\min\left\{M, \frac{\sqrt{M}}{\sqrt{\varepsilon}}\right\}\sqrt{dT}\right), & \text{without Assumption 3.2,} \\ \Omega\left(\min\left\{M, \frac{1}{\varepsilon}\right\}C_0 d\log T\right), & \text{with Assumption 3.2.} \end{cases}
$$

□

F.3. Proof of the Regret Lower Bounds Under User-level Pure-LDP Constraint

Corollary F.5 (Restatement of Corollary 5.5)**.** *For any $\varepsilon \in (0, \log 2)$, there exists a federated linear contextual bandits instance satisfying Assumptions 3.1 and 3.2 such that any* **with-memory** *federated algorithm satisfying ε-LDP must incur a regret lower bounded by*

$$
\Omega\left(\min\left\{M, 1/\varepsilon^2\right\}C_0 d\log T\right).
$$

If Assumption 3.2 is not satisfied, then the minimax regret lower bound becomes

$$
\Omega\left(\min\left\{M, \sqrt{M}/\varepsilon\right\}\sqrt{dT}\right).
$$

Proof. We follow nearly the same argument in Appendix F.2, except the upper bound of $d_{TV}\left(\mathbb{P}(\bar{q}_{\leq t}|\theta), \mathbb{P}(\bar{q}_{\leq}|\theta')\right)$. Since we are under the ε-LDP constraint, we have

$$
\begin{aligned}
&d_{TV}\left(\mathbb{P}(\bar{q}_{\leq t}|\theta), \mathbb{P}(\bar{q}_{\leq t}|\theta')\right) && (9) \\
&\leq \sqrt{1-\exp\left(-\mathtt{KL}\left[\mathbb{P}(\bar{q}_{\leq t}|\theta)\|\mathbb{P}(\bar{q}_{\leq t}|\theta')\right]\right)} && (10) \\
&= \sqrt{1-\exp\left(-\sum_{i\in[M]}\mathtt{KL}\left[\mathbb{P}(\bar{q}_{i,\leq t}|\theta)\|\mathbb{P}(\bar{q}_{i,\leq t}|\theta')\right]\right)} && (11) \\
&\overset{(a)}{\leq} \sqrt{1-\exp\left(-4\varepsilon^2\sum_{i\in[M]}d_{TV}^2\left[\mathbb{P}(\bar{H}_{i,\leq t}|\theta)\|\mathbb{P}(\bar{H}_{i,\leq t}|\theta')\right]\right)} && (12) \\
&\overset{(b)}{\leq} \sqrt{1-\exp\left(-2\varepsilon^2\sum_{i\in[M]}\mathtt{KL}\left[\mathbb{P}(\bar{H}_{i,\leq t}|\theta)\|\mathbb{P}(\bar{H}_{i,\leq t}|\theta')\right]\right)} && (13)
\end{aligned}
$$

$$\leq \sqrt{1 - \exp\Big(- 2\varepsilon^2 M \|\theta - \theta'\|^2 (t-1)/d \Big)}, \tag{14}$$

where (a) is due to Theorem 1 in Duchi et al. (2013), and (b) is due to the Pinsker's inequality.

Following the same argument, we can conlude that under the user-level ε-LDP, we have

$$\text{Regret}(M,T) \geq \begin{cases} \Omega\left(\min\left\{M, \frac{\sqrt{M}}{\varepsilon}\right\}\sqrt{dT}\right), & \text{without Assumption 3.2,} \\ \Omega\left(\min\left\{M, \frac{1}{\varepsilon^2}\right\} C_0 d \log T\right), & \text{with Assumption 3.2.} \end{cases}$$

□

G. Auxiliary Lemmas

This section present lemmas that are commonly used in both bandits literature and differential privacy works, including concentration inequality, composition rule and elliptical potential lemma.

The first is the advanced composition rule, which allows us to reduce the dependency on k for a k-fold composition mechanism.

Lemma G.1 (Advanced composition rule, Theorem 3.20 in (Dwork et al., 2014))**.** *For all $\varepsilon, \delta, \delta' > 0$, the class of (ε, δ)-differentially private mechanisms satisfies $(\varepsilon', k\delta + \delta')$-differential privacy under k-fold composition for*

$$\varepsilon' = \varepsilon\sqrt{2k\log(1/\delta')} + k\varepsilon(e^{\varepsilon} - 1).$$

By noting that $e^{\varepsilon} - 1 < \varepsilon$ when $\varepsilon < \log 2$, we have the following corollary.

Corollary G.2. *Under the same setting in the advanced composition rule, when $\varepsilon < 1/\sqrt{k} < \log 2$, we must have $\varepsilon' \leq \sqrt{6k\log(1/\delta')}$.*

Then, we provide several probability bound random vector and random matrices.

Lemma G.3. *Let $X_1, \ldots, X_d$ be d IID random variables following distribution Laplace$(0, b)$. Then, for any $\beta > 0$, we have*

$$\mathbb{P}\left(\sqrt{\sum_{s=1}^{d} X_s^2} \geq b\sqrt{d}\log(d/\beta)\right) \leq \beta.$$

Proof. We note that

$$\begin{aligned} \mathbb{P}\left(\sqrt{\sum_{s=1}^{d} X_s^2} \geq t\right) &= \mathbb{P}\left(\sum_{s=1}^{d} X_s^2 \geq t^2\right) \\ &\leq \mathbb{P}(\max_s X_s^2 \geq t^2/d) \\ &\overset{(a)}{\leq} \sum_{s=1}^{d} \mathbb{P}(|X_s| \geq t/\sqrt{d}) \\ &\overset{(b)}{\leq} d e^{-\frac{t}{b\sqrt{d}}}, \end{aligned}$$

where (a) is due to union bound, and (b) follows from the fact that the density of Lap$(0, b)$ is $e^{-|x|/b}/(2b)$. Setting $t = b\sqrt{d}\log(d/\beta)$, we complete the proof. □

Lemma G.4 (Matrix concentration lemma, Theorem 1.2 in (Tropp, 2011))**.** *Consider a martingale difference sequence of symmetric random matrices $\{X_t\}_t$ with filtration $\{\mathcal{H}_t\}_t$. Suppose $\mathbb{E}[X_t|\mathcal{H}_t] = 0$ and $\lambda_{\max}(X_t) \leq R$ almost surely. Then,*

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{t=1}^{T} X_t\right) \geq n, \text{ and } \left\|\sum_{t\in[T]} \mathbb{E}\left[X_t^2\right]\right\| \leq \sigma^2\right) \leq d\exp\left(\frac{-n^2/2}{\sigma^2 + Rn/3}\right),$$

where $\|X\|$ is the spectral norm of a matrix X.

The following lemma is widely used in the linear bandits literature.

Lemma G.5 (Elliptical potential lemma, Proposition 1 in Carpentier et al. (2020))**.** *Let $\{x_t\}_{t\geq 1} \subset \mathbb{R}^d$ be an arbitrary sequence of d-dimensional vectors such that $\|x_t\| \leq 1$ for all $t \geq 1$. If $V_t = \lambda I_d + \sum_{\tau=1}^{t-1} x_\tau x_\tau^\intercal$, then*

$$\sum_{t=1}^{T} \|x_t\|_{V_t^{-1}} \leq \sqrt{dT \log \frac{T + d\lambda}{d\lambda}}. \tag{15}$$

Finally, we provide the regret lower bound under the non-private setting and the margin condition for completeness.

Due the margin condition in Assumption 3.2, we choose $r = O(1/C_0)$ in Proposition 4.1 in He et al. (2022b). Then, we have the following non-private regret lower bound under the margin condition.

Proposition G.6. *There exists a federated linear contextual bandits instance satisfying the diversity (Assumption 3.1) and the margin (Assumption 3.2) conditions such that any non-private federated learning algorithm must incur a regret lower bounded by $\Omega(C_0 d \log T)$.*