Pure Exploration in Asynchronous Federated Bandits

Zichen Wang[1] **Chuanhao Li**[2] **Chenyu Song**[3] **Lianghui Wang**[3] **Quanquan Gu**[4] **Huazheng Wang**[3]

[1] University of Illinois Urbana-Champaign, zichenw6@illinois.edu
[2] Yale University, chuanhao.li.cl2637@yale.edu
[3] Oregon State University, {songchen,wangl9,huazheng.wang}@oregonstate.edu
[4] University of California, Los Angeles, qgu@cs.ucla.edu

Abstract

We study the federated pure exploration problem of multi-armed bandits and linear bandits, where M agents cooperatively identify the best arm via communicating with the central server. To enhance the robustness against latency and unavailability of agents that are common in practice, we propose the first federated asynchronous multi-armed bandit and linear bandit algorithms for pure exploration with fixed confidence. Our theoretical analysis shows the proposed algorithms achieve near-optimal sample complexities and efficient communication costs in a fully asynchronous environment. Moreover, experimental results based on synthetic and real-world data empirically elucidate the effectiveness and communication cost-efficiency of the proposed algorithms.

1 INTRODUCTION

Multi-Armed Bandits (MAB) [Auer et al., 2002, Lattimore and Szepesvári, 2020] is a classic sequential decision-making model that is characterized by the exploration-exploitation tradeoff. Pure exploration [Even-Dar et al., 2006, Soare et al., 2014, Bubeck et al., 2009], also known as best arm identification, is an important variant of the MAB problems where the objective is to identify the arm with the maximum expected reward. While most existing bandit solutions are designed under a centralized setting (i.e., data is readily available at a central server), there is increasing interest in federated bandits in terms of regret minimization [Wang et al., 2019, Li et al., 2022, He et al., 2022] and pure exploration [Hillel et al., 2013, Tao et al., 2019, Du et al., 2021] due to the increasing application scale and public concerns about privacy. Specifically, pure exploration for federated bandits considers M agents identifying the best arm collaboratively with limited communication bandwidth, while keeping each agent's raw data local. In federated bandits, the major challenge is the conflict between the need for timely data/model aggregation for low sample complexity and the need for communication efficiency with decentralized agents. Balancing model updates and communication is vital to efficiently solve the problem.

Prior works on distributed/federated pure exploration [Hillel et al., 2013, Tao et al., 2019, R'eda et al., 2022, Du et al., 2021] all focused on synchronous communication protocols, where all agents simultaneously participate in each communication round to exchange their latest observations with a central server (federated setting) or other agents (distributed setting). However, the synchronous setting cannot enjoy efficient communication in real-world applications due to 1) some agents may not interact with the environment in certain rounds and 2) the communication in a global synchronous setting needs to wait until the slowest agent responds to the server, which incurs a significant latency especially when the number of the agents is large and the communication is unstable.

To address the aforementioned challenges of model updates and communication, we study the *asynchronous* communication for federated pure exploration problem in this paper. We consider both stochastic multi-armed bandit and linear bandit settings. To reduce communication costs, we propose novel asynchronous event-triggered communication protocols where each agent sends local updates to and receives aggregated updates from the server independently from other agents, i.e., global synchronization is no longer needed. This improves the robustness against possible delays and unavailability of agents. Event-triggered communication only happens when the agent has a significant amount of new observations, which reduces communication costs while maintaining low sample complexity.

With the new communication protocols, we proposed two asynchronous federated pure exploration algorithms, Federated Asynchronous MAB Pure Exploration (`FAMABPE`) and Federated Asynchronous Linear Pure Exploration

Proceedings of the 40th Conference on Uncertainty in Artificial Intelligence (UAI 2024), PMLR 244:3562–3592.

(`FALinPE`) for MAB and linear bandits, respectively. We theoretically analyzed that these algorithms can return (ϵ, δ)-best arm with an *efficient communication cost*, *efficient switching cost* and *near-optimal sample complexity*, where the returned arm is ϵ close to the best arm with probability at least $1 - \delta$, known as fixed confidence setting [Gabillon et al., 2012, Soare et al., 2014, Xu et al., 2017]. Moreover, we empirically validated the theoretical results based on synthetic data and real-world data. Experimental results showed that our event-triggered communication strategy can achieve efficient communication cost, and would only moderately affect the sample complexity compared with the synchronous baselines.

2 RELATED WORK

Pure exploration The pure exploration problem in single-agent scenarios has been extensively explored in works like Mannor and Tsitsiklis [2004], Even-Dar et al. [2006], Bubeck et al. [2009], Gabillon et al. [2011, 2012], Jamieson et al. [2013], Garivier and Kaufmann [2016], Chen et al. [2016], primarily within the multi-armed bandit framework. Subsequently, Soare et al. [2014], Xu et al. [2017], Tao et al. [2018], Kazerouni and Wein [2019], Fiez et al. [2019], Degenne et al. [2020], Jedra and Proutière [2020] extended these investigations to linear bandits. Advancements by Scarlett et al. [2017], Vakili et al. [2021], Zhu et al. [2021a], Camilleri et al. [2021] further expanded the scope to kernelized bandits. However, these algorithms often suffer from prolonged learning processes and reduced efficacy in the face of limited sample budgets. Hence, our study focuses on the federated resolution of the pure exploration problem.

Distributed/federated pure exploration The exploration of pure exploration problem in distributed/federated bandits has become a focal point in recent research. Studies by Hillel et al. [2013], Tao et al. [2019], Karpov et al. [2020], Mitra et al. [2021], R'eda et al. [2022], Chen et al. [2022], Reddy et al. [2022] investigated MAB in a synchronous environment, while Du et al. [2021] explored kernelized bandits synchronously. Primarily designed for synchronous settings, these studies often rely on experimental design to extract exploration sequences. However, such algorithms encounter challenges in asynchronous environments, stemming from their reliance on 1) global synchronous communication rounds, 2) advance knowledge of the active agent for each round, and 3) the server and agents possessing prior knowledge of the time index t. Our solution addresses these challenges, presenting the inaugural purely asynchronous algorithms for federated pure exploration with fixed confidence.

Distributed/federated regret minimization In tandem with pure exploration, Auer et al. [2002], Abbasi-Yadkori et al. [2011], Filippi et al. [2010], Agrawal and Goyal [2012], Chowdhury and Gopalan [2017] pioneered the study of regret minimization in single-agent settings. This problem has recently expanded to distributed/federated bandits, with literature focusing on MAB Szörényi et al. [2013], Korda et al. [2016], Wang et al. [2019], Mahadik et al. [2020], Shi et al. [2021], Zhu et al. [2021b], Yang et al. [2021, 2022, 2023], Patel et al. [2023], linear bandits Wu et al. [2016], Wang et al. [2019], Dubey and Pentland [2020], Huang et al. [2021], Li and Wang [2022b], Amani et al. [2022], Huang et al. [2023], Zhou and Chowdhury [2023], kernelized bandits Li et al. [2022], and neural bandits Dai et al. [2022]. However, these works are confined to synchronous settings. In alignment with our approach, Chen et al. [2023], Li and Wang [2022a], He et al. [2022], Li et al. [2023] targeted regret minimization in an asynchronous environment. Despite this alignment, the primary objectives of regret minimization differ significantly from those of pure exploration, and none of the mentioned works directly addresses our specific problem.

3 PRELIMINARIES

In this paper, we let $[t] = \{1, ..., t\}$, $\|\mathbf{x}\|$ denotes the Euclidean norm, $\|\mathbf{x}\|_{\mathbf{V}} = \sqrt{\mathbf{x}^\top \mathbf{V} \mathbf{x}}$ denotes the matrix norm, $\log$ denotes the natural logarithm, $\log_2$ denotes the binary logarithm, $\mathbf{I} \in \mathcal{R}^{d \times d}$ denotes the identity matrix, $\mathbf{0}$ denotes the d-dimension zero vector or $d \times d$-dimension zero matrix, $\det(\mathbf{V})$ denotes the determinant of the matrix $\mathbf{V} \in \mathcal{R}^{d \times d}$ and $\mathbf{V}^\top$ denotes the transpose of $\mathbf{V}$. Besides, we utilize $x = \Omega(y)$ to denote that there exists some constant $C > 0$ such that $Cy \leq x$, $x = O(y)$ to denote that there exists some constant C' such that $C'y \geq x$, and $\tilde{O}$ to further hide poly-logarithmic terms.

3.1 FEDERATED BANDITS

MAB We consider the federated asynchronous MAB (similar to Li and Wang [2022a], He et al. [2022]) as follows. There exists a set $\mathcal{M} = \{m\}_{m=1}^{M}$ of M agents ($M \geq 2$), a central server and a environment $\mathcal{A} = \{k\}_{k=1}^{K}$ with K arms ($K \geq 2$). In each round t, an arbitrary agent $m_t \in \mathcal{M}$ becomes active, pulls an arm $k_{m_t,t} \in \mathcal{A}$, and receives reward $r_{m_t,t}$. The reward of each arm $k \in \mathcal{A}$ follows a σ-sub-Gaussian distribution with mean $\mu(k)$. Similar to the other papers that studied the pure exploration [Gabillon et al., 2012, Du et al., 2021], we suppose the best arm $k^* = \arg\max_{k \in \mathcal{A}} \mu(k)$ to be unique.

Linear bandits Different from the MAB, in the federated asynchronous linear bandits [Li and Wang, 2022a, He et al., 2022], every arm k is associated with a context $\mathbf{x}_k \in \mathcal{R}^d$. In round t, if the active agent $m_t \in \mathcal{M}$ pulls an arm $k_{m_t,t} \in \mathcal{A}$, it would receive reward $r_{m_t,t} = \mathbf{x}_{m_t,t}^\top \boldsymbol{\theta}^* + \eta_{m_t,t}$, where $\boldsymbol{\theta}^* \in \mathcal{R}^d$ is the unknown model parameter and $\eta_{m_t,t} \in \mathcal{R}$ denotes the conditionally σ-sub-Gaussian noise (more details are provided in Lemma 12 in the appendix). Without loss of generality, we suppose $\|\mathbf{x}_k\| \leq 1, \forall k \in \mathcal{A}$, $\|\boldsymbol{\theta}^*\| \leq 1$ and the best arm $k^* = \arg\max_{k\in\mathcal{A}} \mathbf{x}_k^\top \boldsymbol{\theta}^*$ to be unique.

3.2 LEARNING OBJECTIVE

This paper focuses on the fixed confidence (ϵ, δ)-pure exploration problem. The goal of the bandit algorithm is to find an estimated best arm $\hat{k}^* \in \mathcal{A}$ which satisfies

$$\mathcal{P}(\Delta(k^*, \hat{k}^*) \leq \epsilon) \geq 1 - \delta \tag{1}$$

with minimum sample complexity. The reward gap parameter satisfies $0 \leq \epsilon < 1$ and the probability parameter satisfies $0 < \delta < 1$. The expected reward gap between arm i and j in the MAB and linear bandits are denoted as $\Delta(i, j) = \mu(i) - \mu(j)$ and $\Delta(i, j) = \mathbf{y}(i, j)^\top \boldsymbol{\theta}^*$, respectively, where $\mathbf{y}(i, j) = \mathbf{x}_i - \mathbf{x}_j$ denotes the difference between contexts. The sample complexity is defined as the agents' total number of interactions with the environment, which is denotes as τ.

3.3 COMMUNICATION MODEL AND ASYNCHRONOUS ENVIRONMENT

Communication model In this paper, we consider a star-shaped communication network [Wang et al., 2019, Li and Wang, 2022a, He et al., 2022], where every agent can only communicate with the server and can not directly communicate with other agents. We define the communication cost $\mathcal{C}(\tau)$ as the *total number of times* that agents upload data to the server and download data from the server in total τ rounds [Dubey and Pentland, 2020, Li and Wang, 2022a, He et al., 2022], i.e.,

$$\begin{aligned}\mathcal{C}(\tau) = \sum_{t=1}^{\tau} & \mathbb{1}\{m_t \text{ uploads data to the server}\} \\ & + \mathbb{1}\{m_t \text{ downloads data from the server}\}.\end{aligned} \tag{2}$$

Asynchronous environment Similar to He et al. [2022], Li et al. [2023], in the asynchronous environment, there is only one active agent m_t (can be an arbitrary agent in $\mathcal{M}$) that interacts with the environment in each round t. Besides, except for the initialization steps, only the active agent is allowed to communicate with the server, i.e., independent from other offline agents.

In our setting, it's important to clarify that the variable t specifically represents the round index, indicating the sequence in which agents engage in the bandit problem. Importantly, it doesn't refer to the actual time of agent involvement. Even when multiple agents are involved, such as in data exchange with the server, there remains a discernible order among these participation events within a very short time frame. This means that even if two events occur very close together in time, a distinct sequence is maintained. As a result, agent participation happens sequentially, based on the index t. Our context has a broader scope compared to previous studies on pure exploration federated bandits [Hillel et al., 2013, Du et al., 2021, Reddy et al., 2022]. This difference arises because those settings require all agents to fully participate in each round, while our setting allows for partial participation, allowing any subset of agents to be involved.

Motivated example We here provide a piratical example for asynchronous federated pure exploration. Let's consider a sequential experimental design problem, e.g., for drug discovery or chemical synthesis, where our goal is to identify an arm that is ϵ-near optimal (i.e., chemical with desired properties) with high probability. In this problem, we are not concerned about cumulative regret (i.e., the quality of the chemicals tried during the online learning process); instead, we only care about whether we can find the optimal arm in the end, and the corresponding sample complexity and communication cost due to their expensive nature (see the introduction in Hillel et al. [2013], R'eda et al. [2022], Du et al. [2021] for details). Additionally, each laboratory lacks samples (i.e., funding for resource) to complete the task individually, so we need to involve multiple labs to collaborate on the learning task. These requirements motivate people to study federated pure exploration problems. Besides, previous synchronous federated pure exploration algorithms assume every agent (i.e., lab) should participate in the exploration (i.e., do the experiment) in each round and the server can force all the agents to upload their data in synchronization rounds. This is impractical due to some agents may get offline (e.g., they run out of resources), and all other agents should wait until they get online (e.g., collect enough resources), this will significantly reduce the learning speed (see the introduction in Li and Wang [2022a], He et al. [2022], Li et al. [2023] for details). In this paper, we propose two asynchronous federated pure exploration algorithms that do not rely on synchronous assumptions and are more practical for real-world applications.

4 ASYNCHRONOUS ALGORITHMS FOR FEDERATED MAB

In this section, we propose the first asynchronous algorithm for the pure exploration problem of federated MAB. As mentioned in Section 1, a key challenge in conducting pure exploration via asynchronous communication is the absence of dedicated synchronous communication rounds where the server can assign arms to explore each agent based on their latest observations. Moreover, there is no guarantee on when or whether an agent would become active again to execute the exploration and report its observations back. This severely hinders the applicability of all existing distributed/federated pure exploration algorithms, whose exploration strategies are based on experimental design [Hillel et al., 2013, Du et al., 2021, R'eda et al., 2022]. In order to address this challenge, we adopt a fully adaptive exploration strategy, such that each agent separately and asynchronously decides which arm to pull, based on the statistics received from the server in its latest communication. We name the resulting algorithm Federated Asynchronous MAB Pure Exploration (`FAMABPE`), and its description is given in Algorithm 1.

FAMABPE algorithm As illustrated in lines 2-7, Algorithm 1 begins with an initialization step for K rounds, where the K arms are pulled sequentially. Then the agents and the server update their local statistics accordingly. For round $t \geq K+1$, an agent m_t becomes active and computes its empirical best arm $i_{m_t,t}$ and the most ambiguous arm $j_{m_t,t}$, where

$$\begin{aligned} i_{m_t,t} =& \arg\max_{k \in \mathcal{A}} \hat{\mu}_{m_t,t}(k), \\ j_{m_t,t} =& \arg\max_{k \in \mathcal{A}/\{i_{m_t,t}\}} \hat{\Delta}_{m_t,t}(k, i_{m_t,t}) \\ &+ \alpha^M_{m_t,t}(i_{m_t,t}, k), \end{aligned} \tag{3}$$

based on which, it selects the most informative arm $k_{m_t,t} = \arg\max_{k \in \{i_{m_t,t}, j_{m_t,t}\}} \alpha^M_{m_t,t}(k)$ to pull in round t. We define the arm k's reward estimator of the agent as $\hat{\mu}_{m_t,t}(k)$, the estimated reward gap between arm i and j of the agent as $\hat{\Delta}_{m_t,t}(i,j) = \hat{\mu}_{m_t,t}(i) - \hat{\mu}_{m_t,t}(j)$ and the pair (i,j)'s exploration bonus of the agents as $\alpha^M_{m_t,t}(i,j) = \alpha^M_{m_t,t}(i) + \alpha^M_{m_t,t}(j)$ (the definition of $\alpha^M_{m_t,t}(k)$ would be provided in Theorem 1). Intuitively, pulling $k_{m_t,t}$ can most decrease $\alpha^M_{m_t,t}(i_{m_t,t}, j_{m_t,t})$ and thus help reduce sample complexity. After observing reward $r_{m_t,t}$ corresponding to $k_{m_t,t}$, m_t checks the communication event in line 11. If the event is true, agent m_t would upload its local reward sum $S^{loc}_{m_t,t}(k)$ and local observation number $T^{loc}_{m_t,t}(k)$, $\forall k \in \mathcal{A}$ to the server. The server then updates its data and estimation

$$\begin{aligned} \hat{\mu}_{ser,t}(k) &= \frac{\hat{\mu}_{ser,t-1}(k) T_{ser,t-1}(k) + S^{loc}_{m_t,t}(k)}{T^{ser}_{t-1}(k) + T^{loc}_{m_t,t}(k)}, \\ T_{ser,t}(k) &= T_{ser,t-1}(k) + T^{loc}_{m_t,t}(k), \ \forall k \in \mathcal{A} \end{aligned} \tag{4}$$

and

$$\begin{aligned} i_{ser,t} &= \arg\max_{k \in \mathcal{A}} \hat{\mu}_{ser,t}(k), \\ j_{ser,t} &= \arg\max_{k \in \mathcal{A}/\{i_{ser,t}\}} \hat{\Delta}_{ser,t}(k, i_{ser,t}) + \alpha^M_{ser,t}(i_{ser,t}, k), \\ B(t) &= \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}), \end{aligned} \tag{5}$$

where $\hat{\mu}_{ser,t}$ denotes the arm k's reward estimator of the server, $T_{ser,t}(k)$ denotes the arm k's observation number of the server, $\hat{\Delta}_{ser,t}(i,j) = \hat{\mu}_{ser,t}(i) - \hat{\mu}_{ser,t}(j)$ denotes the estimated reward gap of the server, and $\alpha^M_{ser,t}(i,j) = \alpha^M_{ser,t}(i) + \alpha^M_{ser,t}(j)$ (the setup of $\alpha^M_{ser,t}(k)$ is shown in Theorem 1) denotes the pair (i,j)'s exploration bonus of the server. If the breaking index $B(t) \leq \epsilon$, the server would set its estimated best arm $\hat{k}^* = i_{ser,t}$ and terminate the algorithm (which implies $\tau = t$). Otherwise, agent m_t would download $\hat{\mu}_{ser,t}(k)$ and $T_{ser,t}(k)$, $\forall k \in \mathcal{A}$ from the server and update its local data as shown in lines 18-19. More details are shown in the pseudo-code.

Low switching cost Different from the previous distributed/federated pure exploration algorithms [Hillel et al., 2013, Du et al., 2021, R'eda et al., 2022], `FAMABPE` enjoys a low switching cost (i.e., $1/2\mathcal{C}(\tau)$). The definition of the switching cost is the number of times the agent $m \in \mathcal{M}$ updates $k_{m,t}$ [Abbasi-Yadkori et al., 2011, He et al., 2022, Li et al., 2023]. We suppose t_1 and t_2 are two neighborhood communication rounds of agent m, and $\hat{\mu}_{m,t}(k)$ and $T_{m,t}(k)$, $\forall k \in \mathcal{A}$, would remain unchanged from round $t_1 + 1$ to t_2 (line 20∼22 in Algorithm 1). This implies $k_{m,t}$, would also remain unchanged. Hence, the switching cost of `FAMABPE` equals the total communication number.

Design of communication event The event-triggered communication strategy of `FAMABPE` can control the amount of local data that each agent $m \in \mathcal{M}$ hasn't uploaded, i.e., $\sum_{k=1}^{K} T^{loc}_{m,t}(k)$ and the size of the exploration bonuses simultaneously. Note that in our setting, neither the agents nor the server knows the total number of observations in the system, i.e., time index t. Therefore, we utilize $\sum_{k=1}^{K} T_{m_t,t}(k)$ and $\sum_{k=1}^{K} T_{ser,t}(k)$ to establish the exploration bonuses of agents and server, respectively. This requires $\sum_{k=1}^{K} T_{m_t,t}(k)$ and $\sum_{k=1}^{K} T_{ser,t}(k)$ to be in a desired proportion to t (which is different from Li and Wang [2022a], He et al. [2022], Li et al. [2023]). Besides, when the server terminates the algorithm, some agents may possess data that has not been uploaded to the server. We wish the amount of these data to be small compared with the sample complexity τ since they have no contribution to identifying $\hat{k}^*$. Our event-triggered communication protocol can efficiently limit the number of the useless samples.

We can show that our proposed `FAMABPE` algorithm can attain near-optimal sample complexity τ, with a low communication cost $\mathcal{C}(\tau)$, which is given in the following theorem.

Theorem 1. *With $\gamma = 1/(2MK)$ and exploration bonuses*

$$\begin{aligned} &\alpha^M_{m_t,t}(k) = \\ &\sigma \sqrt{\frac{2}{T_{m_t,t}(k)} \log\left(\frac{4K}{\delta}\left((1+\gamma M)\sum_{k=1}^{K} T_{m_t,t}(k)\right)^2\right)} \\ &\alpha^M_{ser,t}(k) = \\ &\sigma \sqrt{\frac{2}{T_{ser,t}(k)} \log\left(\frac{4K}{\delta}\left((1+\gamma M)\sum_{k=1}^{K} T_{ser,t}(k)\right)^2\right)}, \end{aligned} \tag{6}$$

the estimated best arm $\hat{k}^*$ of `FAMABPE` *can satisfy condition*

Algorithm 1 Federated Asynchronous MAB Pure Exploration (FAMABPE)

```
Inputs: Arm set $\mathcal{A}$, agent set $\mathcal{M}$, triggered parameter $\gamma$ and $(\delta, \epsilon)$
Initialization:
From round 1 to $K$ sequentially pulls arm from 1 to $K$ and receives reward $r_t$, $\forall t \in [K]$
Server sets $\hat{\mu}_{ser,K}(t) = r_t$ and $T_{ser,K}(t) = 1$    ▷ Server initialization
for $m = 1 : M$ do
    Agent $m$ sets $\hat{\mu}_{m,K+1}(k) = r_t$, $T_{m,K+1}(k) = 1$ and $T^{loc}_{m,K}(k) = S^{loc}_{m,K}(k) = 0, \forall k \in \mathcal{A}$    ▷ Agents initialization
end for
for $t = K + 1 : \infty$ do
    Agent $m_t$ sets $i_{m_t,t}$ and $j_{m_t,t}$ based on (3), pulls arm $k_{m_t,t}$ and receives reward $r_{m_t,t}$    ▷ Sampling rule
    Agent $m_t$ sets $S^{loc}_{m_t,t}(k_{m_t,t}) = S^{loc}_{m_t,t-1}(k_{m_t,t}) + r_{m_t,t}$ and $T^{loc}_{m_t,t}(k_{m_t,t}) = T^{loc}_{m_t,t-1}(k_{m_t,t}) + 1$
    if $\sum_{k=1}^{K}(T_{m_t,t}(k) + T^{loc}_{m_t,t}(k)) > (1+\gamma)\sum_{k=1}^{K} T_{m_t,t}(k)$ then
        [Agent $m_t \to$ Server] Send $S^{loc}_{m_t,t}(k)$ and $T^{loc}_{m_t,t}(k), \forall k \in \mathcal{A}$ to the server    ▷ Upload data to server
        Server updates $\hat{\mu}_{ser,t}(k), T_{ser,t}(k), \forall k \in \mathcal{A}$, $i_{ser,t}$, $j_{ser,t}$ and $B(t)$ based on (4) and (5)
        if $B(t) \leq \epsilon$ then
            Server returns $i_{ser,t}$ as the estimated best arm $\hat{k}^*$ and break    ▷ Stopping rule and decision rule
        end if
        [Server $\to$ Agent $m_t$] Send $T_{ser,t}(k)$ and $\hat{\mu}_{ser,t}(k), \forall k \in \mathcal{A}$ to agent $m_t$    ▷ Download data from server
        Agent $m_t$ sets $T_{m_t,t+1}(k) = T_{ser,t}(k)$ and $\hat{\mu}_{m_t,t+1}(k) = \hat{\mu}_{ser,t}(k), \forall k \in \mathcal{A}$
        Agent $m_t$ sets $T^{loc}_{m_t,t}(k) = 0$ and $S^{loc}_{m_t,t}(k) = 0, \forall k \in \mathcal{A}$
    else
        Agent $m_t$ sets $T_{m_t,t+1}(k) = T_{m_t,t}(k)$ and $\hat{\mu}_{m_t,t+1}(k) = \hat{\mu}_{m_t,t}(k), \forall k \in \mathcal{A}$
    end if
    Inactive agent $m \neq m_t$ sets $T_{m,t+1}(k) = T_{m,t}(k)$ and $\hat{\mu}_{m,t+1}(k) = \hat{\mu}_{m,t}(k), \forall k \in \mathcal{A}$
end for
```

(1) and with probability at least $1 - \delta$ the sample complexity can be bounded by

$$\tau \leq \frac{M + 1/(2K)}{M - 1/2} H^M_\epsilon 2 \log\left(\frac{4K}{\delta}\left(\left(1 + 1/(2K)\right)\Lambda\right)^2\right),$$

where

$$H^M_\epsilon = \sum_{k=1}^{K} \frac{\sigma^2}{\max\left(\frac{\Delta(k^*,k)+\epsilon}{3}, \epsilon\right)^2} = O\left(\sum_{k=1}^{K} \frac{1}{(\Delta(k^*,k)+\epsilon)^2}\right)$$

is the problem complexity in the MAB [Gabillon et al., 2012] and

$$\Lambda = \left(\frac{M + 1/(2K)}{M - 1/2} H^M_\epsilon 4\right)^2 \frac{4K(1 + 1/(2K))}{\delta^{1/2}}.$$

The communication cost satisfies $\mathcal{C}(\tau) = \tilde{O}(KM)$.

Proof sketch of Theorem 1 Proof of Theorem 1 consists of three main components: a) the communication cost $\mathcal{C}(\tau)$; b) the sample complexity τ; c) the estimated best arm satisfies Eq (1). Specifically, to upper bound the total communications cost $\mathcal{C}(\tau)$, we utilize the property of the event-trigger that controls when the agents would communicate with the server (Lemma 1 in the Appendix). To upper bound the sample complexity τ, we first need to establish the relation between $\sum_{k=1}^{K} T_{ser,t}(k)$ and $\sum_{k=1}^{K} T^{loc}_{m,t}(k)$ based on the event triggered strategy (Lemma 4 in the Appendix). Then, we establish exploration bonuses by $\sum_{k=1}^{K} T_{ser,t}(k)$ and $\sum_{k=1}^{K} T_{m,t}(k), \forall k \in \mathcal{A},\ m \in \mathcal{M}$ and bound $T_{ser,\tau}(k)$, $\forall k \in \mathcal{A}$ accordingly (Lemma 2, 5 and 3 in the Appendix). Finally, utilizing the relations of $T_{ser,t}(k)$ and $T^{all}_t(k) = T_{ser,t}(k) + \sum_{m=1}^{M} T^{loc}_{m,t}(k) = \sum_{s=1}^{t} \mathbb{1}\{k_{m_t,t} = k\}$, we can bound $T^{all}_\tau(k), \forall k \in \mathcal{A}$, and $\tau = \sum_{k=1}^{K} T^{all}_\tau(k)$. The guarantee of finding the best arm, i.e., Eq (1), directly follows the property of the breaking index, i.e., if $B(\tau) \leq \epsilon$, then $\Delta(k^*, \hat{k}^*) \leq \epsilon$ with probability at least $1 - \delta$.

Remark 1. *The sample complexity of FAMABPE (i.e., $\tau = O(H^M_\epsilon \log(H^M_\epsilon/\delta))$) can match the lower bound of (ϵ, δ) pure exploration problem $\Omega(\sum_{k=1}^{K} \log(1/\delta)/(\Delta(k^*,k) + \epsilon)^2)$ (see details in Lemma 1 of Kaufmann et al. [2014]) up to a constant factor. It implies if we run (ϵ, δ) pure exploration algorithms on M agents independently with no communication, the sample complexity is $O(M \sum_{k=1}^{K} \log(1/\delta)/(\Delta(k^*,k)+\epsilon)^2)$ and FAMABPE can accelerate the learning process $O(M)$ times. In terms of communication cost, FAMABPE's linear dependence on M matches that attained by previous works studying distributed/federated pure exploration under the less challenging synchronous communication environment [Hillel et al., 2013, Karpov et al., 2020, R'eda et al., 2022, Reddy et al., 2022]. Moreover, the factor K is due to the communication* *event that ensures $\sum_{k=1}^{K} T_{m,t}(k)$ and $\sum_{k=1}^{K} T_{m,t}^{loc}(k)$ are in a desired proportion to t. As mentioned in our previous discussion on its design, this is necessary for the asynchronous communication studied in this paper.*

5 ASYNCHRONOUS ALGORITHM FOR FEDERATED LINEAR BANDITS

In this section, we further consider the pure exploration problem of federated linear bandits. We propose an algorithm called Federated Asynchronous Linear Pure Exploration (FALinPE), and its description is given in Algorithm 2.

FALinPE algorithm Similar to Algorithm 1, FALinPE starts with an initialization step (line 2-7), where each arm is pulled once. Then in each round $t \geq K+1$, the active agent m_t sets its estimated model parameter $\hat{\boldsymbol{\theta}}_{m_t,t}$, empirical best arm $i_{m_t,t}$ and most ambiguous arm $j_{m_t,t}$ as

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{m_t,t} =& \mathbf{V}_{m_t,t}^{-1}\mathbf{b}_{m_t,t}, \quad i_{m_t,t} = \arg\max_{k\in\mathcal{A}} \mathbf{x}_k^\top \hat{\boldsymbol{\theta}}_{m_t,t}, \\ j_{m_t,t} =& \arg \max_{k\in\mathcal{A}/\{i_{m_t,t}\}} \hat{\Delta}_{m_t,t}(k, i_{m_t,t}) \\ &+ \alpha_{m_t,t}^{L}(i_{m_t,t}, k) \end{aligned} \tag{7}$$

and pulls the most informative arm $k_{m_t,t}$ (context denotes as $\mathbf{x}_{m_t,t}$). The exploration bonuses of pair (i,j) in the linear case are defined as $\alpha_{m_t,t}^{L}(i,j) = \|\mathbf{y}(i,j)\|_{\mathbf{V}_{m_t,t}^{-1}} C_{m_t,t}$ and $\alpha_{ser,t}^{L}(i,j) = \|\mathbf{y}(i,j)\|_{\mathbf{V}_{ser,t}^{-1}} C_{ser,t}$, where the definitions of the scalers $C_{m_t,t}$ and $C_{ser,t}$ are provided in Theorem 2. Besides, the estimated reward gaps between arm i and j are defined as $\hat{\Delta}_{m_t,t}(i,j) = \mathbf{y}(i,j)^\top \hat{\boldsymbol{\theta}}_{m_t,t}$ and $\hat{\Delta}_{ser,t}(i,j) = \mathbf{y}(i,j)^\top \hat{\boldsymbol{\theta}}_{ser,t}$. Agent m_t would update its covariance matrix $\mathbf{V}_{m_t,t}^{loc}$, $\mathbf{b}_{m_t,t}^{loc}$ and $T_{m_t,t}^{loc}(k_{m_t,t})$ as

$$\begin{aligned} \mathbf{V}_{m_t,t}^{loc} &= \mathbf{V}_{m_t,t-1}^{loc} + \mathbf{x}_{m_t,t}\mathbf{x}_{m_t,t}^\top, \\ \mathbf{b}_{m_t,t}^{loc} &= \mathbf{b}_{m_t,t-1}^{loc} + r_{m_t,t}\mathbf{x}_{m_t,t}, \\ T_{m_t,t}^{loc}(k_{m_t,t}) &= T_{m_t,t-1}^{loc}(k_{m_t,t}) + 1. \end{aligned} \tag{8}$$

FALinPE utilizes a hybrid event-triggered strategy to control the size of the exploration bonus $\alpha_{ser,t}^{L}(i,j)$ and $\alpha_{m_t,t}^{L}(i,j)$, and the observation number $\sum_{k=1}^{K} T_{ser,t}(k)$ and $\sum_{k=1}^{K} T_{ser,t}(k)$. If at least one of the two events is triggered, then agent m_t would upload its collected data $\mathbf{V}_{m_t,t}^{loc}$, $\mathbf{b}_{m_t,t}^{loc}$ and $T_{m_t,t}^{loc}(k), \forall k \in \mathcal{A}$ to the server. The server would update its collected data and estimation

$$\begin{aligned} &\mathbf{V}_{ser,t} = \mathbf{V}_{ser,t-1} + \mathbf{V}_{m_t,t}^{loc}, \\ &\mathbf{b}_{ser,t} = \mathbf{b}_{ser,t-1} + \mathbf{b}_{m_t,t}^{loc}, \\ &T_{ser,t}(k) = T_{ser,t-1}(k) + T_{m_t,t}^{loc}(k),\ \forall k \in \mathcal{A}, \\ &\hat{\boldsymbol{\theta}}_{ser,t} = V_{ser,t}^{-1} b_{ser,t} \end{aligned} \tag{9}$$

and set $i_{ser,t}$, $j_{ser,t}$, and the breaking index $B(t)$ as

$$\begin{aligned} &i_{ser,t} = \arg\max_{k\in\mathcal{A}} \mathbf{x}_k^\top \hat{\boldsymbol{\theta}}_{ser,t}, \\ &j_{ser,t} = \arg \max_{k=\mathcal{A}/\{i_{ser,t}\}} \hat{\Delta}_{ser,t}(k, i_{ser,t}) + \alpha_{ser,t}^{L}(i_{ser,t}, k) \\ &B(t) = \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha_{ser,t}^{L}(i_{ser,t}, j_{ser,t}). \end{aligned} \tag{10}$$

If the breaking index $B(t) > \epsilon$, the server would return $\mathbf{V}_{ser,t}$, $\mathbf{b}_{ser,t}$ and $T_{ser,t}(k), \forall k \in \mathcal{A}$ to the user. FALinPE would repeat the above steps until $B(\tau) \leq \epsilon$.

Design of communication events of FALinPE Similar to FAMABPE, FALinPE also enjoys a low switching cost (i.e, $1/2\mathcal{C}(\tau)$). Besides, the hybrid event-triggered strategy can simultaneously control the size of $\mathbf{V}_{m,t}^{loc}$, $\sum_{k=1}^{K} T_{m,t}^{loc}(k)$ and the exploration bonuses. Note that Min et al. [2021] also utilize a hybrid event-triggered communication protocol to achieve a similar goal, but for learning stochastic shortest path with linear function approximation. The exploration bonuses in the linear case are not only related to t but also related to covariance matrices. Therefore, different from the communication protocol in the MAB, the event-triggered communication protocol in the linear bandits is additionally required to keep $\mathbf{V}_{m_t,t}$ and $\mathbf{V}_{ser,t}$ in a desired proportion to the global covariance matrix $\lambda\mathbf{I} + \sum_{s=1}^{t} \mathbf{x}_{m_t,t}\mathbf{x}_{m_t,t}^\top$.

Arm selection strategy To minimize the sample complexity τ. We hope every agent can pull the most informative arm $k_{m_t,t}$ to reduce the exploration bonus $\alpha_{m_t,t}^{L}(i_{m_t,t}, j_{m_t,t})$ as fast as possible. The arm selection strategy of Algorithm 2 ensures active agent m_t to pull $k_{m_t,t}$ to most decrease the matrix norm $\|\mathbf{y}(i_{m_t,t}, j_{m_t,t})\|_{\mathbf{V}_{m_t,t}^{-1}}$ (and also $\alpha_{m_t,t}^{L}(i_{m_t,t}, j_{m_t,t})$). Different from the MAB, in the linear case we can not directly find $k_{m_t,t}$, and need to derive it with a linear programming [Xu et al., 2017], it yields

$$k_{m_t,t} = \arg\min_{k\in\mathcal{A}} \frac{T_{m_t,t}(k)}{p_k^*(\mathbf{y}(i_{m_t,t}, j_{m_t,t}))}, \tag{11}$$

where $p^*(\cdot)$ is defined as follows

$$p_k^*(\mathbf{y}(i_{m_t,t}, j_{m_t,t})) = \frac{w_k^*(\mathbf{y}(i_{m_t,t}, j_{m_t,t}))}{\sum_{s=1}^{K} |w_s^*(y(i_{m_t,t}, j_{m_t,t}))|} \tag{12}$$

and

$$\begin{aligned} \mathbf{w}^*(\mathbf{y}(i_{m_t,t},j_{m_t,t})) = \arg\min_{\mathbf{w}\in\mathcal{R}^d} \sum_{k=1}^{K} |w_k| \\ s.t. \quad \mathbf{y}(i_{m_t,t}, j_{m_t,t}) = \sum_{k=1}^{K} w_k \mathbf{x}_k. \end{aligned} \tag{13}$$

The notation $w_k^*(\mathbf{y}(i_{m_t,t}, j_{m_t,t}))$ denotes the k-th element of vector $\mathbf{w}^*(\mathbf{y}(i_{m_t,t}, j_{m_t,t}))$. Besides, the optimal value of programming (13) denotes as $\rho(\mathbf{y}(i,j)) =$

Algorithm 2 Federated Asynchronous Linear Pure Exploration (FALinPE)

```
Inputs: Arm set $\mathcal{A}$, agent set $\mathcal{M}$, regularization parameter $\lambda > 0$, triggered parameter $\gamma_1, \gamma_2$, and $(\delta, \epsilon)$
Initialization:
From round $1$ to $K$ sequentially pulls arm from $1$ to $K$ and receives reward $r_t$, $\forall t \in |K|$
Server sets $\mathbf{V}_{ser,K} = \lambda\mathbf{I} + \sum_{t=1}^{K} \mathbf{x}_t\mathbf{x}_t^\top$, $\mathbf{b}_{ser,K} = \sum_{t=1}^{K} \mathbf{x}_t r_t$, $T_{ser,K}(k) = 1$, $\forall k \in \mathcal{A}$    ▷ Server initialization
for $m = 1 : M$ do
    Agent $m$ sets $\mathbf{V}_{m,K+1} = \lambda\mathbf{I} + \sum_{t=1}^{K} \mathbf{x}_t\mathbf{x}_t^\top$, $\mathbf{b}_{m,K+1} = \sum_{t=1}^{K} \mathbf{x}_t r_t$, $T_{m,K+1}(k) = 1$, $\mathbf{V}_{m,K}^{loc} = \mathbf{0}$, $\mathbf{b}_{m,K}^{loc} = \mathbf{0}$ and $T_{m,K}^{loc}(k) = 0, \forall k \in \mathcal{A}$    ▷ Agents initialization
end for
for $t = K+1 : \infty$ do
    Agent $m_t$ sets $\hat{\boldsymbol{\theta}}_{m_t,t}$, $i_{m_t,t}$ and $j_{m_t,t}$ by (7), pulls $k_{m_t,t}$ by (11) or (14) and receive $r_{m_t,t}$    ▷ Sampling rule
    Agent $m_t$ updates $\mathbf{V}_{m_t,t}^{loc}$, $\mathbf{b}_{m_t,t}^{loc}$ and $T_{m_t,t}^{loc}(k_{m_t,t})$ based on (8)
    if $\frac{\det(\mathbf{V}_{m_t,t}+\mathbf{V}_{m_t,t}^{loc})}{\det(\mathbf{V}_{m_t,t})} > (1+\gamma_1)$ or $\frac{\sum_{k=1}^{K}(T_{m_t,t}(k)+T_{m_t,t}^{loc}(k))}{\sum_{k=1}^{K} T_{m_t,t}(k)} > (1+\gamma_2)$ then
        [Agent $m_t \to$ Server] Send $\mathbf{V}_{m_t,t}^{loc}$, $\mathbf{b}_{m_t,t}^{loc}$ and $T_{m_t,t}^{loc}(k)$, $\forall k \in \mathcal{A}$ to the server    ▷ Upload data to server
        Server sets $\mathbf{V}_{ser,t}$, $\mathbf{b}_{ser,t}$, $T_{ser,t}(k)$, $\forall k \in \mathcal{A}$, $\hat{\boldsymbol{\theta}}_{ser,t}$, $i_{ser,t}$, $j_{ser,t}$ and $B(t)$ based on (9) and (10)
        if $B(t) \le \epsilon$ then
            Server returns $i_{ser,t}$ as the estimated best arm $\hat{k}^*$ and break the loop    ▷ Stopping rule and decision rule
        end if
        [Server $\to$ Agent $m_t$] Send $\mathbf{V}_{ser,t}$, $\mathbf{b}_{ser,t}$ and $T_{ser,t}(k)$, $\forall k \in \mathcal{A}$ to agent $m_t$    ▷ Download data
        Agent $m_t$ sets $\mathbf{V}_{m_t,t+1} = \mathbf{V}_{ser,t}$, $\mathbf{b}_{m_t,t+1} = \mathbf{b}_{ser,t}$ and $T_{m_t,t+1}(k) = T_{ser,t}(k)$, $\forall k \in \mathcal{A}$
        Agent $m_t$ sets $\mathbf{V}_{m_t,t}^{loc} = \mathbf{0}$, $\mathbf{b}_{m_t,t}^{loc} = \mathbf{0}$ and $T_{m_t,t}^{loc}(k) = 0$, $\forall k \in \mathcal{A}$
    else
        Agent $m_t$ sets $\mathbf{V}_{m_t,t+1} = \mathbf{V}_{m_t,t}$, $\mathbf{b}_{m_t,t+1} = \mathbf{b}_{m_t,t}$ and $T_{m_t,t+1}(k) = T_{m_t,t}(k)$, $\forall k \in \mathcal{A}$
    end if
    Inactive agent $m \neq m_t$ sets $\mathbf{V}_{m,t+1} = \mathbf{V}_{m,t}$, $\mathbf{b}_{m,t+1} = \mathbf{b}_{m,t}$ and $T_{m,t+1}(k) = T_{m,t}(k)$, $\forall k \in \mathcal{A}$
end for
```

$\sum_{k=1}^{K} w_k^*(\mathbf{y}(i,j))$, $\forall i,j \in \mathcal{A}$. However, the programming is computationally inefficient and we also propose to select the arm greedily similar to [Xu et al., 2017]

$$k_{m_t,t} = \arg\max_{k\in\mathcal{A}} \mathbf{y}(i_{m_t,t}, j_{m_t,t})^\top \times (\mathbf{V}_{m_t,t} + \mathbf{x}_k\mathbf{x}_k^\top)^{-1}\mathbf{y}(i_{m_t,t}, j_{m_t,t}). \tag{14}$$

Although we did not analyze the theoretical property of the greedy arm selection strategy, in the experiment section, we empirically validate that it performs well.

Theorem 2. *With $0 < \lambda \le \sigma^2\big(\sqrt{1+\gamma_1 M} + \sqrt{2\gamma_1}M\big)^2 \log(2/\delta)$, $\gamma_1 = 1/M^2$, $\gamma_2 = 1/(2MK)$, arm selection strategy (11) and 5*

$$C_{m_t,t} = \sqrt{\lambda} + \left(\sqrt{2\gamma_1}M + \sqrt{1+\gamma_1 M}\right) \times \left(\sigma\sqrt{d\log\left(\frac{2}{\delta}\left(1 + \frac{(1+\gamma_2 M)\sum_{k=1}^{K} T_{m_t,t}(k)}{\min(\gamma_1,1)\lambda}\right)\right)}\right)$$

$$C_{ser,t} = \sqrt{\lambda} + \left(\sqrt{2\gamma_1}M + \sqrt{1+\gamma_1 M}\right) \times \left(\sigma\sqrt{d\log\left(\frac{2}{\delta}\left(1 + \frac{(1+\gamma_2 M)\sum_{k=1}^{K} T_{ser,t}(k)}{\min(\gamma_1,1)\lambda}\right)\right)}\right),$$

the estimated best arm $\hat{k}^*$ of FALinPE can satisfy condition (1) and with probability at least $1-\delta$ the sample complexity can be bounded by

$$\tau \le \frac{M+1/(2K)}{M-1/2}\left(\sqrt{2} + \sqrt{1+\frac{1}{M}}\right)^2 H_\epsilon^L 4\sigma^2 d \times \log\left(1 + \frac{(1+1/(2K))\Lambda^2}{\lambda/M^2}\right) + \Gamma,$$

where

$$H_\epsilon^L = \sum_{k=1}^{K} \max_{i,j\in\mathcal{A}} \frac{\rho(\mathbf{y}(i,j))p_k^*(\mathbf{y}(i,j))}{\max\left(\frac{\Delta(k^*,i)+\epsilon}{3}, \frac{\Delta(k^*,j)+\epsilon}{3}, \epsilon\right)^2}$$

is the problem complexity in the linear bandits [Xu et al., 2017],

$$\Lambda = \frac{M+1/(2K)}{M-1/2}\left(\sqrt{2} + \sqrt{1+\frac{1}{M}}\right)^2 H_\epsilon^L 4\sigma^2 d \times \sqrt{1 + \frac{1+1/(2K)}{\lambda/M^2}} + \Gamma$$

and

$$\Gamma = \frac{M+1/(2K)}{M-1/2}\left(\sqrt{2} + \sqrt{1+\frac{1}{M}}\right)^2 H_\epsilon^L 4\sigma^2 d\log\left(\frac{2}{\delta}\right).$$

The communication cost satisfies $\mathcal{C}(\tau) = \tilde{O}\big(\max(M^2 d, MK)\big)$.

Remark 2. *As mention in [Xu et al., 2017], the sample complexity of the LinGapE runs by a single agent (i.e., $\tau = \tilde{O}(H_\epsilon^L d)$) can match the lower bound in Soare et al. [2014] up to a constant factor. As shown in the Theorem 2, the sample complexity of* `FALinPE` *can also satisfies $\tau = \tilde{O}(H_\epsilon^L d)$ when we select the proper λ, γ_1 and γ_2. Besides, the communication cost of the* `FALinPE` *satisfies $\mathcal{C}(\tau) = \tilde{O}(dM^2)$ when $M \geq K/d$, which is the same as the communication cost of Async-LinUCB [Li and Wang, 2022a] and FedLinUCB [He et al., 2022] (both are $\tilde{O}(dM^2)$) in the regret minimization setting. It is worth noting that in our and He et al. [2022]'s setting, the communication between the active agent and server is independent to the offline agent, while in Li and Wang [2022a]'s setting, the algorithm requires a global download section. In addition, the guarantee in Li and Wang [2022a] relies on a stringent regularity assumption on the contexts, while ours and He et al. [2022]'s do not. We claim that the* `FALinPE` *can achieve a near-optimal sample complexity and efficient communication cost.*

6 EXPERIMENTS

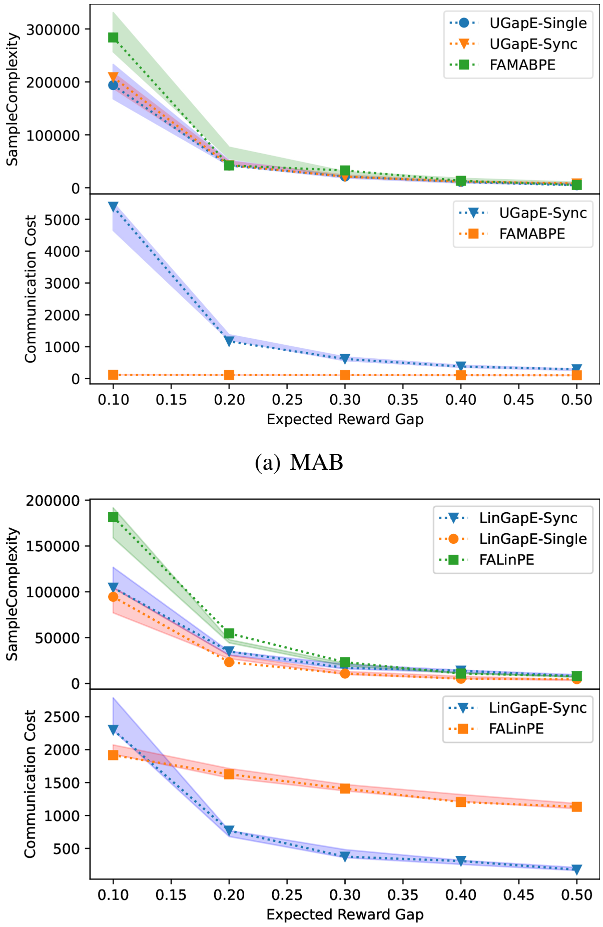
To empirically validate the communication and sample efficiency of `FAMABPE` and `FALinPE`, we conduct experiments on both synthetic and real-world dataset. Our algorithms are compared with some baseline algorithms, in which the active agent would share its data with other agents via the server in every round. We would compare `FAMABPE` with single agent UGapEc and synchronous UGapEc [Gabillon et al., 2012] in the MAB setting. Besides, we would also compare `FALinPE` with single agent LinGapE and synchronous LinGapE [Xu et al., 2017]. We want to clarify that the asynchronous algorithms typically incur larger communication cost than synchronous ones under the same regret/sample complexity guarantee, which is also acknowledged in prior works studying regret minimization [Li and Wang, 2022a, He et al., 2022, Li et al., 2023]. Therefore, the inclusion of synchronous algorithms' mainly serves as a reference showing the performance under the easier synchronous setting. We run the algorithms 10 times and plot their average results.

6.1 EXPERIMENTS ON SYNTHETIC DATA

In this section, we report experiments on synthetic dataset for federated MAB and linear bandits.

6.1.1 Experiment Setup

MAB We simulate the federated MAB in Section 3.1, with $\sigma = 0.3$, $\delta = 0.05$, $\epsilon = 0$, $K = 5$ and $M = 10$. We sample the optimal arm from the uniform distribution and selectively sample the non-optimal arm to guarantee the reward gap. For synchronous UGapEc, we set the communication frequency as 100 rounds. At the end of every 100 rounds, the agents would upload their exploration results to the server and download other agents' exploration results from the server. The communication cost of this naive synchronous algorithm is just $C(\tau) = \tau/50$, this is due to there are $\tau/(100M)$ communication episodes and in each episode agents would upload and download data for $2M$ times. The setup of `FAMABPE` follows Theorem 1 and in each round, the active agent m_t is uniformly sampled from $\mathcal{M}$.



(a) MAB

(b) Linear

Figure 1: Synthetic data: Experimental results for federated MAB and federated linear bandits.

Linear bandits Similar to the MAB case, we simulate the federated linear bandits with $d = 5$ and other parameters are the same as the MAB setting. We first sample the model parameter $\boldsymbol{\theta}^*$ from a uniform distribution. Then, we sample the context of the optimal arm and selectively sample non-optimal arms to guarantee the reward gap. The synchronous LinGapE is similar to synchronous UGapEc. The setup of `FALinPE` follows Theorem 2 and the active agent in the linear case is also uniformly sampled from $\mathcal{M}$.

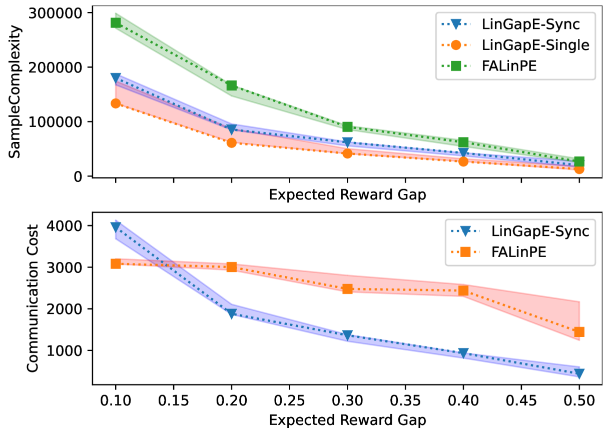


Figure 2: Experimental results on MovieLens for federated linear bandits.

6.1.2 Experiment Results

MAB The results of federated MAB are shown in Figure 1(a). All algorithms output their estimated best arms $\hat{k}^* = k^*$. We report the sample complexity and communication cost for the reward gap from 0.1 to 0.5. We can observe that the single agent which runs UGapEc achieved the smallest sample complexity. In comparison, the synchronous UGapEc would spend a slightly larger sample complexity when the gap equals 0.1 and spend an almost identical cost when the gap equals 0.2 to 0.5. Compared with these baseline algorithms, our `FAMABPE` had a slightly larger sample complexity and can achieve the lowest communication cost (only took a communication cost of 100 to 120 to go from a gap of 0.5 to a gap of 0.1). `FAMABPE` is the only algorithm that can achieve near-optimal sample complexity and efficient communication cost in a fully asynchronous environment.

Linear bandits The results of federated linear bandits are provided in Figures 1(b). All algorithms output their estimated best arm $\hat{k}^* = k^*$. Similar to the MAB, we can observe that a single agent which runs LinGapE and synchronous LinGapE achieved the lowest sample complexity. In comparison, `FALinPE` required a relatively large sample complexity, especially when the gap equals 0.1. Furthermore, the communication cost of synchronous LinGapE is larger than `FALinPE` when the gap equals 0.1. Otherwise, smaller than `FALinPE`.

6.2 EXPERIMENTS ON REAL-WORLD DATA

In this section, we report an additional experiment on real-world dataset for federated linear bandits setting.

6.2.1 Experiment Setup

We use the MovieLens 20M dataset [Harper and Konstan, 2016] for the experiment. We follow [Li and Wang, 2022a] to preprocess the data and extract item features. Specifically, we keep users with over $3,000$ observations, which results in a dataset with 54 users, 26567 items (movies), and 214729 interactions. For each item, we extract TF-IDF features from its associated tags and apply PCA to obtain item features with dimension $d = 25$. We consider all items with non-zero ratings as positive feedback (reward $r = 1$), and use ridge regression to learn $\boldsymbol{\theta}^*$ from extracted item features and their 0/1 rewards. To construct an arm set, we follow the same procedure as the simulation in Section 6 by first sampling an optimal arm and then selectively sampling non-optimal arms to guarantee the reward gap.

The baseline algorithms considered in this section are the same as those in the synthetic data case (Section 6). Besides, we set $d = 25$, $K = 10$, $\epsilon = 0.05$, and other parameters of the federated linear bandits are also identical to Section 6. We report the average results of 10 runs.

6.2.2 Experiment Results

The results of the federated linear bandits are shown in Figures 2. In each run, every algorithm could derive the best arm k^*. Similar to the results based on synthetic data, single agent LinGapE and synchronous LinGapE enjoyed the lowest sample complexity, and `FALinPE` spent a relatively large sample complexity. Besides, according to the tendency, `FALinPE`'s communication cost would be smaller than the synchronous LinGapE's communication cost when the expected reward gap is smaller equals 0.15. Note that synchronous LinGapE can only work in a synchronous environment, hence, `FALinPE` is the *only* known federated linear bandit algorithm that can simultaneously achieve near-optimal sample complexity and efficient communication cost in the fully asynchronous environment.

7 CONCLUSION

In this paper, we propose the first study on the pure exploration problem of both federated MAB and federated linear bandits in an asynchronous environment. First, we proposed an algorithm named `FAMABPE`, which can complete the (ϵ, δ)-pure exploration object of the federated MAB with $\tilde{O}(H_\epsilon^M)$ sample complexity and $\tilde{O}(MK)$ communication cost using a novel event-triggered communication protocol. Then, we improved `FAMABPE` to `FALinPE`, which can finish the same object in the linear case with $\tilde{O}(H_\epsilon^L d)$ sample complexity and $\tilde{O}\big(\max(M^2 d, MK)\big)$ communication cost. At the end of the paper, the effectiveness of the offered algorithms was further examined by the numerical simulation based on synthetic data and real-world data. In our future work, a potential direction is to investigate federated asynchronous pure exploration algorithms with a fixed budget.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvari. Improved algorithms for linear stochastic bandits. In *NIPS*, 2011.

Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, 2012.

Sanae Amani, Tor Lattimore, Andr'as Gyorgy, and Lin F. Yang. Distributed contextual linear bandits with minimax optimal communication cost. In *International Conference on Machine Learning*, 2022. URL `https://api.semanticscholar.org/CorpusID:249097947`.

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 2002.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *International Conference on Algorithmic Learning Theory*, 2009.

Romain Camilleri, Julian Katz-Samuels, and Kevin G. Jamieson. High-dimensional experimental design and kernel bandits. In *International Conference on Machine Learning*, 2021.

Lijie Chen, Jian Li, and Mingda Qiao. Towards instance optimal bounds for best arm identification. In *Annual Conference Computational Learning Theory*, 2016. URL `https://api.semanticscholar.org/CorpusID:11275420`.

Yu-Zhen Janice Chen, L. Yang, Xuchuang Wang, Xutong Liu, Mohammad Hassan Hajiesmaili, John C.S. Lui, and Donald F. Towsley. On-demand communication for asynchronous multi-agent bandits. In *International Conference on Artificial Intelligence and Statistics*, 2023. URL `https://api.semanticscholar.org/CorpusID:256868900`.

Zhirui Chen, P. N. Karthik, Vincent Yan Fu Tan, and Yeow Meng Chee. Federated best arm identification with heterogeneous clients. *ArXiv*, abs/2210.07780, 2022.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *International Conference on Machine Learning*, 2017.

Zhongxiang Dai, Yao Shu, Arun Verma, Flint Xiaofeng Fan, Bryan Kian Hsiang Low, and Patrick Jaillet. Federated neural bandit. *ArXiv*, abs/2205.14309, 2022.

Rémy Degenne, Pierre M'enard, Xuedong Shang, and Michal Valko. Gamification of pure exploration for linear bandits. In *International Conference on Machine Learning*, 2020. URL `https://api.semanticscholar.org/CorpusID:220302498`.

Yihan Du, Wei Chen, Yuko Kuroki, and Longbo Huang. Collaborative pure exploration in kernel bandit. *ArXiv*, abs/2110.15771, 2021.

Abhimanyu Dubey and Alex 'Sandy' Pentland. Differentially-private federated linear bandits. *ArXiv*, abs/2010.11425, 2020.

Eyal Even-Dar, Shie Mannor, and Y. Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *J. Mach. Learn. Res.*, 7:1079–1105, 2006.

Tanner Fiez, Lalit P. Jain, Kevin G. Jamieson, and Lillian J. Ratliff. Sequential experimental design for transductive linear bandits. In *Neural Information Processing Systems*, 2019. URL `https://api.semanticscholar.org/CorpusID:195218369`.

Sarah Filippi, Olivier Cappé, Aurélien Garivier, and Csaba Szepesvari. Parametric bandits: The generalized linear case. In *NIPS*, 2010.

Victor Gabillon, Mohammad Ghavamzadeh, Alessandro Lazaric, and Sébastien Bubeck. Multi-bandit best arm identification. In *NIPS*, 2011. URL `https://api.semanticscholar.org/CorpusID:13982470`.

Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. In *NIPS*, 2012.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Annual Conference Computational Learning Theory*, 2016. URL `https://api.semanticscholar.org/CorpusID:1278907`.

F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. *ACM Trans. Interact. Intell. Syst.*, 5:19:1–19:19, 2016.

Jiafan He, Tianhao Wang, Yifei Min, and Quanquan Gu. A simple and provably efficient algorithm for asynchronous federated contextual linear bandits. *ArXiv*, abs/2207.03106, 2022.

Eshcar Hillel, Zohar S. Karnin, Tomer Koren, Ronny Lempel, and Oren Somekh. Distributed exploration in multi-armed bandits. In *NIPS*, 2013.

Ruiquan Huang, Weiqiang Wu, Jing Yang, and Cong Shen. Federated linear contextual bandits. *ArXiv*, abs/2110.14177, 2021.

Ruiquan Huang, Huanyu Zhang, Luca Melis, Milan Shen, Meisam Hajzinia, and J. Yang. Federated linear contextual bandits with user-level differential privacy. In *International Conference on Machine Learning*, 2023. URL https://api.semanticscholar.org/CorpusID:259108568.

Kevin G. Jamieson, Matthew Malloy, Robert D. Nowak, and Sébastien Bubeck. lil' ucb : An optimal exploration algorithm for multi-armed bandits. *ArXiv*, abs/1312.7308, 2013. URL https://api.semanticscholar.org/CorpusID:2606438.

Yassir Jedra and Alexandre Proutière. Optimal best-arm identification in linear bandits. *ArXiv*, abs/2006.16073, 2020. URL https://api.semanticscholar.org/CorpusID:220249870.

Nikolai Karpov, Qin Zhang, and Yuanshuo Zhou. Collaborative top distribution identifications with limited interaction. *ArXiv*, abs/2004.09454, 2020.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *J. Mach. Learn. Res.*, 17:1:1–1:42, 2014. URL https://api.semanticscholar.org/CorpusID:12309216.

Abbas Kazerouni and Lawrence M. Wein. Best arm identification in generalized linear bandits. *Oper. Res. Lett.*, 49: 365–371, 2019.

Nathaniel Korda, Balázs Szörényi, and Shuai Li. Distributed clustering of linear bandits in peer to peer networks. *ArXiv*, abs/1604.07706, 2016.

Tor Lattimore and Csaba Szepesvári. Bandit algorithms. 2020.

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

Chuanhao Li and Hongning Wang. Communication efficient federated learning for generalized linear bandits. *ArXiv*, abs/2202.01087, 2022b.

Chuanhao Li, Huazheng Wang, Mengdi Wang, and Hongning Wang. Communication efficient distributed learning for kernelized contextual bandits. *ArXiv*, abs/2206.04835, 2022.

Chuanhao Li, Huazheng Wang, Mengdi Wang, and Hongning Wang. Learning kernelized contextual bandits in a distributed and asynchronous environment. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=-G1kjTFsSs.

Kanak Mahadik, Qingyun Wu, Shuai Li, and Amit Sabne. Fast distributed bandits for online recommendation systems. *Proceedings of the 34th ACM International Conference on Supercomputing*, 2020. URL https://api.semanticscholar.org/CorpusID:220252162.

Shie Mannor and John N. Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. In *Journal of machine learning research*, 2004.

Yifei Min, Jiafan He, Tianhao Wang, and Quanquan Gu. Learning stochastic shortest path with linear function approximation. In *International Conference on Machine Learning*, 2021. URL https://api.semanticscholar.org/CorpusID:239768454.

Aritra Mitra, Hamed Hassani, and George J. Pappas. Exploiting heterogeneity in robust federated best-arm identification. *ArXiv*, abs/2109.05700, 2021. URL https://api.semanticscholar.org/CorpusID:237513934.

Kumar Kshitij Patel, Lingxiao Wang, Aadirupa Saha, and Nathan Srebro. Federated online and bandit convex optimization. *ArXiv*, abs/2311.17586, 2023. URL https://api.semanticscholar.org/CorpusID:260927382.

Cl'emence R'eda, Sattar Vakili, and Emilie Kaufmann. Near-optimal collaborative learning in bandits. *ArXiv*, abs/2206.00121, 2022.

Kota Srinivas Reddy, P. N. Karthik, and Vincent Yan Fu Tan. Almost cost-free communication in federated best arm identification. *ArXiv*, abs/2208.09215, 2022.

Jonathan Scarlett, Ilija Bogunovic, and Volkan Cevher. Lower bounds on regret for noisy gaussian process bandit optimization. *ArXiv*, abs/1706.00090, 2017.

Chengshuai Shi, Cong Shen, and Jing Yang. Federated multi-armed bandits with personalization. In *International Conference on Artificial Intelligence and Statistics*, 2021.

Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. *ArXiv*, abs/1409.6110, 2014.

Balázs Szörényi, Róbert Busa-Fekete, István Hegedüs, Róbert Ormándi, Márk Jelasity, and Balázs Kégl. Gossip-based distributed stochastic bandit algorithms. In *International Conference on Machine Learning*, 2013.

Chao Tao, Saúl A. Blanco, and Yuanshuo Zhou. Best arm identification in linear bandits with linear dimension dependency. In *International Conference on Machine Learning*, 2018.

Chao Tao, Qin Zhang, and Yuanshuo Zhou. Collaborative learning with limited interaction: Tight bounds for distributed exploration in multi-armed bandits. *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 126–146, 2019.

Lin Tie, Kai-Yuan Cai, and Yan Lin. Rearrangement inequalities for hermitian matrices. *Linear Algebra and its Applications*, 434:443–456, 2011.

Sattar Vakili, Nacime Bouziani, Sepehr Jalali, Alberto Bernacchia, and Da shan Shiu. Optimal order simple regret for gaussian process bandits. In *Neural Information Processing Systems*, 2021.

Yuanhao Wang, Jiachen Hu, Xiaoyu Chen, and Liwei Wang. Distributed bandit learning: Near-optimal regret with efficient communication. *arXiv: Learning*, 2019.

Qingyun Wu, Huazheng Wang, Quanquan Gu, and Hongning Wang. Contextual bandits in a collaborative environment. *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 2016. URL `https://api.semanticscholar.org/CorpusID:18898947`.

Liyuan Xu, Junya Honda, and Masashi Sugiyama. A fully adaptive algorithm for pure exploration in linear bandits. In *International Conference on Artificial Intelligence and Statistics*, 2017.

L. Yang, Xuchuang Wang, Mohammad Hassan Hajiesmaili, Lijun Zhang, John C.S. Lui, and Don Towsley. Cooperative multi-agent bandits: Distributed algorithms with optimal individual regret and constant communication costs. *ArXiv*, abs/2308.04314, 2023. URL `https://api.semanticscholar.org/CorpusID:260704623`.

Lin Yang, Yu-Zhen Janice Chen, Stephen Pasteris, Mohammad Hassan Hajiesmaili, John C.S. Lui, and Donald F. Towsley. Cooperative stochastic bandits with asynchronous agents and constrained feedback. In *Neural Information Processing Systems*, 2021.

Lin Yang, Yu-Zhen Janice Chen, Mohammad Hassan Hajiesmaili, John C.S. Lui, and Donald F. Towsley. Distributed bandits with heterogeneous agents. *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, pages 200–209, 2022. URL `https://api.semanticscholar.org/CorpusID:246240664`.

Xingyu Zhou and Sayak Ray Chowdhury. On differentially private federated linear contextual bandits. *ArXiv*, abs/2302.13945, 2023. URL `https://api.semanticscholar.org/CorpusID:257219732`.

Yinglun Zhu, Dongruo Zhou, Ruoxi Jiang, Quanquan Gu, Rebecca M. Willett, and Robert D. Nowak. Pure exploration in kernel and neural bandits. *ArXiv*, abs/2106.12034, 2021a.

Zhaowei Zhu, Jingxuan Zhu, Ji Liu, and Yang Liu. Federated bandit: A gossiping approach. *ACM SIGMETRICS Performance Evaluation Review*, 49:3 – 4, 2021b.

A NOTATIONS

$\mathcal{A}$	Arm set
$\mathcal{M}$	Agent set
d	Dimension of the model parameter and context
τ	Sample complexity (stopping time)
$\mathcal{C}(\tau)$	Communication cost
k^*	Best arm
$\hat{k}^*$	Estimated best arm
m_t	Active agent in round t
$\eta_{m_t,t}$	σ-sub-Gaussian noise
$r_{m_t,t}$	Received reward of agent m_t in round t
$k_{m_t,t}$	Arm pulled by agent m_t in round t
$\mu(k)$	Expected reward of arm k in MAB
$\mathbf{x}_k$	Context of arm k
$\mathbf{y}(i,j)$	Context difference between $\mathbf{x}_i$ and $\mathbf{x}_j$
$\boldsymbol{\theta}^*$	Model parameter
δ	Probability parameter of the fixed-confidence pure exploration problem
ϵ	Reward gap parameter of the fixed-confidence pure exploration problem
$\gamma/\gamma_1/\gamma_2$	Triggered parameters
λ	Regularization parameter of the covariance matrix
$B(t)$	Breaking index
$\alpha^M_{m_t,t}(k)$	Exploration bonus of arm k of the agent m_t in FAMABPE
$\alpha^M_{ser,t}(k)$	Exploration bonus of arm k of the server in FAMABPE
$\alpha^L_{m_t,t}(i,j)$	Exploration bonus of pair (i,j) of the agent m_t in FALinPE
$\alpha^L_{ser,t}(i,j)$	Exploration bonus of pair (i,j) of the server in FALinPE
$\hat{\boldsymbol{\theta}}_{m_t,t}$	Estimated model parameter of the agent m_t in FALinPE
$\hat{\boldsymbol{\theta}}_{ser,t}$	Estimated model parameter of the server in FALinPE
$i_{m_t,t}$	Empirical best arm of agent m_t in round t
$j_{m_t,t}$	Most ambiguous arm of agent m_t in round t
$\Delta(i,j)$	Expected reward gap between arms i and j
$\hat{\Delta}_{m_t,t}(i,j)$	Estimated reward gap between arms i and j of the agent m_t
$\hat{\Delta}_{ser,t}(i,j)$	Estimated reward gap between arms i and j of the server
$\hat{\mu}_{m_t,t}(k)$	Estimated reward of arm k of the agent m_t in FAMABPE
$\hat{\mu}_{ser,t}(k)$	Estimated reward of arm k of the server in FAMABPE
$T_{m_t,t}(k)$	Number of observations on arm k that is available to the agent m_t at τ
$T_{ser,t}(k)$	Number of observations on arm k that is available to the server at τ
$T^{loc}_{m_t,t}(k)$	Number of observations on arm k has not been uploaded to the server by the agent m_t at t
$T^{all}_t(k)$	Total number of observations on arm k
$\mathbf{V}_{m_t,t}$	Covariance matrix of the agent m_t
$\mathbf{V}_{ser,t}$	Covariance matrix of the server
$\mathbf{V}^{loc}_{m_t,t}$	Covariance matrix has not been uploaded to the server by the agent m_t
$\mathbf{V}^{all}_t$	Global covariance matrix
H^M_ϵ	Problem complexity of the MAB
H^L_ϵ	Problem complexity of the linear bandits

B PROOF OF THEOREM 1

For clarity, we here reintroduce the notations used in the proof. Recall that $T_t^{all}(k)$ denotes the total number of arm k be pulled till round t, $T_{ser,t}(k)$ denotes the number of observations on arm k that is available to the server at t, $T_{m,t}(k)$ denotes the the number of observations on arm k that is available to the agent m at t and $T_{m,t}^{loc}(k)$ denotes the observations on arm k of agent m has not been uploaded to the server at t. Besides, $\hat{\Delta}_{ser,t}(i,j)$ and $\hat{\Delta}_{m,t}(i,j)$ denote the estimated reward gap between arm i and j of the agent m and server in round t, respectively. Furthermore, we let $\alpha_{m,t}^{M}(k)$ and $\alpha_{ser,t}^{M}(k)$ denoting the exploration bonuses of the agent m and server, respectively. Moreover, we define the reward estimator of arm k of the agent m and server as $\hat{\mu}_{m,t}(k)$ and $\hat{\mu}_{ser,t}(k)$, respectively.

Remark 3 (Global and local data in the federated MAB). *By the design of Algorithm 1, the total number of times arm k has been pulled till round t satisfies $T_t^{all}(k) = T_{ser,t}(k) + \sum_{m=1}^{M} T_{m,t}^{loc}(k)$, where $T_{ser,t}(k)$ denotes the number of observations on arm k that has been uploaded to the server and $\sum_{m=1}^{M} T_{m,t}^{loc}(k)$ denotes the total number of observations on arm k that agents $m = 1, 2, \dots, M$ have not uploaded to the server. Besides, as $T_{m,t}(k)$, $\forall m \in \mathcal{M}, k \in \mathcal{A}$ is downloaded from the server in some round earlier than t, we have $T_{m,t}(k) \leq T_{ser,t}(k)$, $\forall k \in \mathcal{A}$ and $\sum_{k=1}^{K} T_{m,t}(k) \leq \sum_{k=1}^{K} T_{ser,t}(k)$.*

Detailed proof for the first two components are given in the following two subsections.

B.1 UPPER BOUND COMMUNICATION COST $\mathcal{C}(\tau)$

Lemma 1 (Communication cost). *Following the setting of Theorem 1, the total communication cost of* `FAMABPE` *can be bounded by*

$$\mathcal{C}(\tau) \leq 2\Big((M + 1/\gamma)\log_2 \tau\Big).$$

Proof of Lemma 1. The proof of this Lemma can be divided into two sections, in the first section, we would divide the sample complexity τ into $\log_2 \tau$ episodes, then we would analyze the upper bound of the communication number of all agents in each episode. We define

$$\mathcal{T}_i = \min\Big\{t \in [K+1, \tau],\ \sum_{k=1}^{K} T_{ser,t}(k) \geq 2^i\Big\}$$

and the set of all rounds into episodes i as $\{\mathcal{T}_i, \mathcal{T}_i{+}1, ..., \mathcal{T}_{i+1}{-}1\}$. According to the definition, we have $\sum_{k=1}^{K} T_{ser,\tau}(k) \leq \tau$, and thus

$$\max\{i \geq 0\} = \log_2\left(\sum_{k=1}^{K} T_{ser,\tau}(k)\right) \leq \log_2 \tau.$$

We then prove $\forall i \geq 0$, from round $\mathcal{T}_i$ to $\mathcal{T}_{i+1} - 1$, the communication number of all agents can be bounded by $M + 1/\gamma$. We first define the communication number of agent m from $\mathcal{T}_i$ to $\mathcal{T}_{i+1} - 1$ as $\mathcal{N}_m$, the sequence of communication round of agent m from round $\mathcal{T}_i$ to $\mathcal{T}_{i+1} - 1$ as $t_1^m, ..., t_{\mathcal{N}_m}^m$, the communication number of all agents as L and the sequence of all communication rounds from round $\mathcal{T}_i$ to $\mathcal{T}_{i+1} - 1$ as $t_{i,1}, ..., t_{i,L}$. According to the communication condition (line 11 of Algorithm 1), we have $\forall m \in \mathcal{M},\ j \in |\mathcal{N}_m|$

$$\begin{gathered}\sum_{k=1}^{K}\Big(T_{m,t_j^m}(k) + T_{m,t_j^m}^{loc}(k)\Big) > (1+\gamma)\sum_{k=1}^{K} T_{m,t_j^m}(k)\\ \sum_{k=1}^{K} T_{m,t_j^m}^{loc}(k) > \gamma \sum_{k=1}^{K} T_{m,t_j^m}(k).\end{gathered} \tag{15}$$

Then, $\forall m \in \mathcal{M},\ j \in |\mathcal{N}_m|/\{1\}$, we have

$$\sum_{k=1}^{K} T_{m,t_j^m}^{loc}(k) > \gamma \sum_{k=1}^{K} T_{m,t_j^m}(k) \geq \gamma \sum_{k=1}^{K} T_{ser,\mathcal{T}_i}(k). \tag{16}$$

The inequality holds due to $T_{m,t_j^m}(k) = T^{ser}_{t^m_{j-1}}(k)$ and $t^m_{j-1} \geq \mathcal{T}_i$ when $j \in |\mathcal{N}_m|/\{1\}$. The above inequality implies $\forall t_{i,l} \geq t_2^{m_{t_{i,l}}}$

$$\begin{aligned}\sum_{k=1}^{K}\Big(T_{ser,t_{i,l}}(k) - T_{ser,t_{i,l-1}}(k)\Big) &\geq \sum_{k=1}^{K}\Big(T_{ser,t_{i,l-1}}(k) + T^{loc}_{m_{t_{i,l}},t_{i,l}}(k)\Big) - \sum_{k=1}^{K} T_{ser,t_{i,l-1}}(k)\\ &= \sum_{k=1}^{K} T^{loc}_{m_{t_{i,l}},t_{i,l}}(k)\\ &> \gamma\sum_{k=1}^{K} T_{m_{t_{i,l}},t_{i,l}}(k)\\ &\geq \gamma\sum_{k=1}^{K} T_{ser,\mathcal{T}_i}(k).\end{aligned}$$

Finally we can bound $L = \sum_{m=1}^{M}\mathcal{N}_m$

$$\begin{aligned}\sum_{k=1}^{K}\Big(T_{ser,\mathcal{T}_{i+1}-1}(k) - T_{ser,\mathcal{T}_i}(k)\Big) &= \sum_{l=1}^{L-1}\sum_{k=1}^{K}\Big(T_{ser,t_{i,l+1}}(k) - T_{ser,t_{i,l}}(k)\Big)\\ &\geq \gamma\sum_{m=1}^{M}(\mathcal{N}_m - 1)\sum_{k=1}^{K} T_{ser,\mathcal{T}_i}(k).\end{aligned} \tag{17}$$

The last inequality holds owing to (15) and (16). With the definition of the episode, we have $\sum_{k=1}^{K} T_{ser,\mathcal{T}_{i+1}-1}(k) \leq 2\sum_{k=1}^{K} T_{ser,\mathcal{T}_i}(k)$. We can then rewrite equation (17) as

$$M + 1/\gamma \geq \sum_{m=1}^{M}\mathcal{N}_m = L.$$

Due to one communication including one upload and one download, the communication cost in one episode is at most $2(M + 1/\gamma)$ (following the definition of (2)). We can then bound the total communication cost

$$\mathcal{C}(\tau) \leq 2\Big((M + 1/\gamma)\log_2(\tau)\Big). \tag{18}$$

In the light of (18), setting of the Theorem 1 and (22) (the upper bound of the sample complexity τ), we can bound the communication cost

$$\mathcal{C}(\tau) = \tilde{O}(KM). \tag{19}$$

□

B.2 UPPER BOUND SAMPLE COMPLEXITY τ

Combining the breaking condition in Algorithm 1 (line 14∼16), and the definition of $B(\tau)$, we have

$$\epsilon \geq B(\tau) = \hat{\Delta}^{ser}_{\tau}(j^{ser}_{\tau}, i^{ser}_{\tau}) + \beta^{ser}_{\tau}(i^{ser}_{\tau}, j^{ser}_{\tau}).$$

Let's first consider the case when the empirically best arm on the server side is not the optimal arm, i.e., $i_{ser,\tau} \neq k^*$. By the definition of $j_{ser,\tau}$, we have

$$\hat{\Delta}_{ser,\tau}(j_{ser,\tau}, i_{ser,\tau}) + \beta_{ser,\tau}(i_{ser,\tau}, j_{ser,\tau}) \geq \hat{\Delta}_{ser,\tau}(k^*, i_{ser,\tau}) + \alpha^{M}_{ser,\tau}(i_{ser,\tau}, k^*).$$

Recall that $\hat{k}^* = i_{ser,\tau}$ is the empirically best arm. Therefore, we have

$$\epsilon \geq \hat{\Delta}_{ser,\tau}(k^*, \hat{k}^*) + \alpha^{M}_{ser,\tau}(\hat{k}^*, k^*) \geq \Delta(k^*, \hat{k}^*),$$

where the second inequality is due to Lemma 2 below (proof of Lemma 2 is at the end of this section).

Lemma 2. *Following the setting of Theorem 1, we can establish the exploration bonuses*

$$\alpha^M_{m_t,t}(k) = \sigma\sqrt{\frac{2}{T_{m_t,t}(k)}\log\left(\frac{4K}{\delta}\Big((1+\gamma M)\sum_{k=1}^{K} T_{m_t,t}(k)\Big)^2\right)}$$

$$\alpha^M_{ser,t}(k) = \sigma\sqrt{\frac{2}{T_{ser,t}(k)}\log\left(\frac{4K}{\delta}\Big((1+\gamma M)\sum_{k=1}^{K} T_{ser,t}(k)\Big)^2\right)},$$

for all $t \in [K+1, \tau]$, $k \in \mathcal{A}$, and the event

$$\mathcal{I} = \left\{\forall k \in \mathcal{A}, \forall t \in [K+1,\tau],\ |\hat{\mu}_{m_t,t}(k) - \mu(k)| \le \beta_{m_t,t}(k),\ |\hat{\mu}_{ser,t}(k) - \mu(k)| \le \alpha^M_{ser,t}\right\}.$$

We have $\mathcal{P}(\mathcal{I}) \ge 1-\delta$.

Furthermore, if $i_{ser,t} = k^*$, we have $\epsilon \ge \Delta(k^*, \hat{k}^*) = 0$. The discussion above implies $\hat{k}^*$ output by `FAMABPE` satisfies the (ϵ,δ)-condition (1).

We now continue to bound the sample complexity τ. First, we need to establish Lemma 3 below, which upper bounds $T_{ser,\tau}(k)$, the number of observations on arm k that is available to the server at τ.

Lemma 3. *Under the setting of Theorem 1, if event $\mathcal{I}$ happens, we can bound*

$$T_{ser,\tau}(k) \le \frac{2\sigma^2\log\Big(4K\big((1+M\gamma)\sum_{s=1}^{K} T_{ser,\tau}(s)\big)^2/\delta\Big)}{\max\left(\frac{\Delta(k^*,k)+\epsilon}{3},\epsilon\right)^2} + \gamma M\sum_{s=1}^{K} T_{ser,\tau}(s),\ \forall k \in \mathcal{A}.$$

With Lemma 3, we have

$$\begin{aligned}
\sum_{k=1}^{K} T_{ser,\tau}(k) &\le \sum_{k=1}^{K}\frac{2\sigma^2\log\Big(4K\big((1+\gamma M)\sum_{s=1}^{K} T_{ser,\tau}(s)\big)^2/\delta\Big)}{\max\left(\frac{\Delta(k^*,k)+\epsilon}{3},\epsilon\right)^2} + \gamma KM\sum_{s=1}^{K} T_{ser,\tau}(s)\\
&\le \frac{1}{1-\gamma KM}\sum_{k=1}^{K}\frac{2\sigma^2\log\Big(4K\big((1+\gamma M)\sum_{s=1}^{K} T_{ser,\tau}(s)\big)^2/\delta\Big)}{\max\left(\frac{\Delta(k^*,k)+\epsilon}{3},\epsilon\right)^2}\\
&\le \frac{1}{1-\gamma KM}\sum_{k=1}^{K}\frac{2\sigma^2\log\Big(4K\big((1+\gamma M)\tau\big)^2/\delta\Big)}{\max\left(\frac{\Delta(k^*,k)+\epsilon}{3},\epsilon\right)^2}.
\end{aligned}$$

The last inequality is due to $\sum_{k=1}^{K} T_{ser,t}(k) \le \tau$ (Remark 3). Based on the relation between τ, $\sum_{m=1}^{M}\sum_{k=1}^{K} T^{loc}_{m,\tau}(k)$, and $\sum_{k=1}^{K} T_{ser,\tau}(k)$ (Remark 3), we have

$$\begin{aligned}
\tau &= \sum_{k=1}^{K} T^{all}_{\tau}(k) = \sum_{k=1}^{K} T_{ser,\tau}(k) + \sum_{m=1}^{M}\sum_{k=1}^{K} T^{loc}_{m,\tau}(k)\\
&\le \big(1+\gamma M\big)\sum_{k=1}^{K} T_{ser,\tau}(k)\\
&\le \frac{1+\gamma M}{1-\gamma KM}\sum_{k=1}^{K}\frac{2\sigma^2\log\Big(4K\big((1+\gamma M)\tau\big)^2/\delta\Big)}{\max\left(\frac{\Delta(k^*,k)+\epsilon}{3},\epsilon\right)^2}\\
&= \frac{1+\gamma M}{1-\gamma KM}H^M_\epsilon 2\log\left(\frac{4K\big((1+\gamma M)\tau\big)^2}{\delta}\right),
\end{aligned}\tag{20}$$

where the first inequality is due to Lemma 4, the second inequality is due to the inequality we established above, and the third is by definition of H_ϵ^M.

Recalling that in Theorem 1 we suppose $\gamma = 1/(2MK)$. Let τ' be a parameter that satisfies

$$\tau' \leq \tau = \frac{M + 1/(2K)}{M - 1/2} H_\epsilon^M 2 \log\left(\frac{4K\Big((1 + 1/(2K))\tau'\Big)^2}{\delta}\right),$$

where the equality is owing to the definition of the γ and (20). Due to the fact that $\sqrt{x} \geq \log(x)$ holds when $x > 0$, we have

$$\begin{aligned}
\tau' &\leq \frac{M + 1/(2K)}{M - 1/2} H_\epsilon^M 2 \log\left(\Big(\frac{4K(1 + 1/(2K))\tau'}{\delta^{1/2}}\Big)^2\right) \\
&\leq \frac{M + 1/(2K)}{M - 1/2} H_\epsilon^M 4 \sqrt{\frac{4K(1 + 1/(2K))\tau'}{\delta^{1/2}}} \\
&\leq \left(\frac{M + 1/(2K)}{M - 1/2} H_\epsilon^M 4\right)^2 \frac{4K(1 + 1/(2K))}{\delta^{1/2}}.
\end{aligned} \tag{21}$$

We define the last term of (21) as Λ. Then based on (20) and (21), we can finally get

$$\tau \leq \frac{M + 1/(2K)}{M - 1/2} H_\epsilon^M 2 \log\left(\frac{4K\Big((1 + 1/(2K))\Lambda\Big)^2}{\delta}\right), \tag{22}$$

where

$$\Lambda = \left(\frac{M + 1/(2K)}{M - 1/2} H_\epsilon^M 4\right)^2 \frac{4K(1 + 1/(2K))}{\delta^{1/2}}.$$

Lemma 4. *Following the setting of Theorem 1, we have* $\forall t \in [K + 1, \tau]$

$$\sum_{k=1}^{K} T_{ser,t}(k) > 1/\gamma \sum_{k=1}^{K} T_{m_t,t}^{loc}(k).$$

Proof for Lemma 2, Lemma 3, and Lemma 4 In the following paragraphs, we provide the detailed proof for the lemmas used above.

Proof of Lemma 4. For every round $t \in [K + 1, \tau]$, if the agent m_t communicates with the server at round t, we have

$$\sum_{k=1}^{K} T_{ser,t}(k) \geq 1/\gamma \sum_{k=1}^{K} T_{m_t,t}^{loc}(k) = 0.$$

The inequality holds by line 19 of Algorithm 1.

Else, according to the triggered condition of Algorithm 1, if agent m_t does not communicate with the server in round t, we have

$$\begin{aligned}
\sum_{k=1}^{K} (T_{m_t,t}(k) + T_{m_t,t}^{loc}(k)) \leq (1 + \gamma) \sum_{k=1}^{K} T_{m_t,t}(k) \\
\sum_{k=1}^{K} T_{m_t,t}^{loc}(k) \leq \gamma \sum_{k=1}^{K} T_{m_t,t}(k).
\end{aligned}$$

With the fact that $\sum_{k=1}^{K} T_{m_t,t}(k) \leq \sum_{k=1}^{K} T_{ser,t}(k)$ (Remark 3), we can finally get

$$\sum_{k=1}^{K} T_{m_t,t}^{loc}(k) \leq \gamma \sum_{k=1}^{K} T_{ser,t}(k).$$

Here we finish the proof of Lemma 4. □

Based on Lemma 4, we can prove Lemma 2 as shown below.

Proof of Lemma 2. Due to $\hat{\mu}_{m_t,t}(k)$ and $T_{m_t,t}(k)$, $\forall k \in \mathcal{A},\ t \in [K+1, \tau]$ are all downloaded from the server and they would remain unchanged until the next round agent m_t communicates with the server. This implies $\forall t_1 \in [K+1, \tau]$, there exists a $t_2 \in [K+1, \tau]$ which satisfies

$$\alpha^M_{m_{t_1},t_1}(k) = \alpha^M_{ser,t_2}(k) \text{ and } \hat{\mu}_{m_{t_1},t_1}(k) = \hat{\mu}_{ser,t_2}(k),\ \forall k \in \mathcal{A}.$$

Hence, we can derive

$$\mathcal{P}(\mathcal{I}) = \mathcal{P}\Big(\forall k \in \mathcal{A}, \forall t \in [K+1, \tau],\ |\hat{\mu}_{ser,t}(k) - \mu(k)| \leq \alpha^M_{ser,t}(k)\Big). \tag{23}$$

We define $\mathcal{I}^c$ as the contradicted event of $\mathcal{I}$. Utilizing the union bound, it can be decomposed by

$$\mathcal{P}(\mathcal{I}^c) \leq \sum_{k=1}^{K} \sum_{t=K+1}^{\tau} \mathcal{P}\Big(|\hat{\mu}_{ser,t}(k) - \mu(k)| > \alpha^M_{ser,t}(k)\Big). \tag{24}$$

With the help of the Hoeffeding inequality (Lemma 13), it has

$$\begin{aligned}
\mathcal{P}\Big(|\hat{\mu}_{ser,t}(k) - \mu(k)| > \alpha^M_{ser,t}(k)\Big) &\leq e^{-\alpha^M_{ser,t}(k)^2 T_{ser,t}(k)/2\sigma^2} \\
&= e^{-\log\left(\frac{4K}{\delta}\left((1+\gamma M)\sum_{k=1}^{K} T_{ser,t}(k)\right)^2\right)} \\
&= \frac{\delta}{4K\big((1+\gamma M)\sum_{k=1}^{K} T_{ser,t}(k)\big)^2} \\
&\leq \frac{\delta}{4Kt^2}.
\end{aligned} \tag{25}$$

The first equality is owing to the definition of $\alpha_{ser,t}(k)$ and the last inequality is owing to $t = \sum_{k=1}^{K} T_{ser,t}(k) + \sum_{m=1}^{M}\sum_{k=1}^{K} T^{loc}_{m,t}(k) \leq (1+\gamma M)\sum_{k=1}^{K} T_{ser,t}(k)$ (Lemma 4). Substituting the last term of (25) into (24), we can finally bound

$$\sum_{k=1}^{K} \sum_{t=K+1}^{\tau} \mathcal{P}\Big(|\hat{\mu}^{ser}_t(k) - \mu(k)| > \alpha^M_{ser,t}(k)\Big) \leq \delta$$

and $\mathcal{P}(\mathcal{I}) = 1 - \mathcal{P}(\mathcal{I}^c) \geq 1 - \delta$. Here we finish the proof of Lemma 2. □

Before proving Lemma 3, we first need to establish Lemma 5 below.

Lemma 5. *We define $k_{ser,t}$ in round $t \in [K+1, \tau]$ as $k_{ser,t} = \arg\max_{k \in \{i_{ser,t}, j_{ser,t}\}} \alpha^M_{ser,t}(k)$. Following the setting of Theorem 1, if event $\mathcal{I}$ happens, we can bound the index $B(t)$ as*

$$B(t) \leq \min\Big(0, -\Delta(k^*, k^{ser}_t) + 4\beta_{ser,t}(k_{ser,t})\Big) + 2\alpha^M_{ser,t}(k_{ser,t}). \tag{26}$$

Proof of Lemma 5. This proof follows the idea of Gabillon et al. [2012]. Consider the case when $i_{ser,t} = k^*$, we have

$$\begin{aligned}
B(t) &= \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}) \\
&\leq \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + 2\alpha^M_{ser,t}(k_{ser,t}) \\
&\leq \Delta(j_{ser,t}, i_{ser,t}) + 4\alpha^M_{ser,t}(k_{ser,t}) \\
&= -\Delta(k^*, j_{ser,t}) + 4\alpha^M_{ser,t}(k_{ser,t}) \\
&\leq -\Delta(k^*, k_{ser,t}) + 4\alpha^M_{ser,t}(k_{ser,t}),
\end{aligned} \tag{27}$$

where the first inequality is owing to the definition of $k_{ser,t}$ and the second inequality is owing to the definition of the event $\mathcal{I}$. Then, consider the case when $j_{ser,t} = k^*$, we have

$$\begin{aligned} B(t) &= \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\le \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + 2\alpha^M_{ser,t}(k_{ser,t}) \\ &\le -\hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + 2\alpha^M_{ser,t}(k_{ser,t}) \\ &\le -\Delta(j_{ser,t}, i_{ser,t}) + 4\alpha^M_{ser,t}(k_{ser,t}) \\ &= -\Delta(k^*, i_{ser,t}) + 4\alpha^M_{ser,t}(k_{ser,t}) \\ &\le -\Delta(k^*, k_{ser,t}) + 4\alpha^M_{ser,t}(k_{ser,t}), \end{aligned} \tag{28}$$

where the second inequality is owing to the definition of the $i_{ser,t}$ (line 13 of Algorithm 1).

Combine (27) and (28), it yields

$$B(t) \le -\Delta(k^*, k_{ser,t}) + 4\alpha^M_{ser,t}(k_{ser,t}), \tag{29}$$

when $i_{ser,t} = k^*$ or $j_{ser,t} = k^*$. Furthermore, due to $B(t) = \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^M_{ser,t}(j_{ser,t}, i_{ser,t})$ and $\hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) \le 0$, we can derive $B(t) \le \alpha^M_{ser,t}(j_{ser,t}, i_{ser,t}) \le 2\alpha^M_{ser,t}(k_{ser,t})$. In the light of (29), we can finally get

$$B(t) \le \min\Big(0, \Delta(k^*, k_{ser,t}) + 2\alpha^M_{ser,t}(k_{ser,t})\Big) + 2\alpha^M_{ser,t}(k_{ser,t}). \tag{30}$$

We further consider the case when $i_{ser,t} \ne k^*$ and $j_{ser,t} \ne k^*$, then we can derive

$$\begin{aligned} B(t) &= \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\le \Delta(j_{ser,t}, i_{ser,t}) + 2\alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &= \Delta(j_{ser,t}, k^*) + \Delta(k^*, i_{ser,t}) + 2\alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\le \Delta(j_{ser,t}, k^*) + 3\alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\le \Delta(j_{ser,t}, k^*) + 6\alpha^M_{ser,t}(k_{ser,t}) \\ &= -\Delta(k^*, j_{ser,t}) + 6\alpha^M_{ser,t}(k_{ser,t}), \end{aligned} \tag{31}$$

where the second equality is owing to $\Delta(i_{ser,t}, j_{ser,t}) = \Delta(i_{ser,t}, k^*) + \Delta(k^*, j_{ser,t})$ and second inequality is owing to

$$\begin{aligned} \alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}) &\ge \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\ge \hat{\Delta}_{ser,t}(k^*, i_{ser,t}) + \alpha^M_{ser,t}(i_{ser,t}, k^*) \\ &\ge \Delta(k^*, i_{ser,t}). \end{aligned}$$

Similar to (31), we also can show

$$\begin{aligned} B(t) &= \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\le \alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\le -\Delta(k^*, i_{ser,t}) + \hat{\Delta}_{ser,t}(k^*, i_{ser,t}) + \alpha^M_{ser,t}(k^*, i_{ser,t}) + \alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\le -\Delta(k^*, i_{ser,t}) + \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^M_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^M_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\le -\Delta(k^*, i_{sert}) + 4\alpha^M_{ser,t}(k_{ser,t}). \end{aligned} \tag{32}$$

The second inequality is due to the definition of the even $\mathcal{I}$ and the third inequality is due to the definition of the $j_{ser,t}$.

Combine (31) and (32), it yields

$$B(t) \le -\Delta(k^*, k_{ser,t}) + 6\alpha^M_{ser,t}(k_{ser,t}) \tag{33}$$

when $i_{ser,t} \ne k^*$ and $j_{ser,t} \ne k^*$. In the light of (33), we can finally get

$$B(t) \le \min\Big(0, -\Delta(k^*, k_{ser,t}) + 4\alpha^M_{ser,t}(k_{ser,t})\Big) + 2\alpha^M_{ser,t}(k_{ser,t}). \tag{34}$$

Combine (30) and (34), then we can finish the proof of Lemma 2. □

Proof of Lemma 3. Suppose agent m communicates in round t_1 and t_2, then from round $t \in [t_1+1, t_2]$, owing to $\hat{\mu}_{m,t}(k)$ and $T_{m,t}(k)$ remain unchanged, $k_{m,t}$ would not change either (we highlight this knowledge in Remark **??**). We define at round $t_k \neq \tau$, an agent m communicates with the server and $k_{ser,t_k} = k, \forall k \in \mathcal{A}$ (the definition of the $k_{ser,t}$ is provided in Lemma 5). And after round t_k, when any agent communicating with the server, $k_{ser,t} \neq k$. This implies

$$\begin{aligned} T_{ser,\tau}(k) \leq & T_{m,t_k+1}(k) + (\gamma M)\sum_{s=1}^{K} T_{ser,\tau}(s) \\ = & T_{ser,t_k}(k) + (\gamma M)\sum_{s=1}^{K} T_{ser,\tau}(s). \end{aligned} \tag{35}$$

The inequality holds due to for $t \in [t_k+1, \tau], \forall m \in \mathcal{M}$ would upload $T^{loc}_{m,t}(k) > 0$ to the server at most one time (according to the definition of t_k) and $T^{loc}_{m,t}(k) \leq \sum_{s=1}^{K} T^{loc}_{m,t}(s) \leq \gamma \sum_{s=1}^{K} T_{ser,\tau}(s)$ according to $t \leq \tau$ and the Lemma 4.

We would further bound $T_{ser,t_k}(k) = T_{m,t_k+1}(k)$. According to agent m sets $k_{ser,t_k} = k$ and the definition of the breaking condition (line 14∼16 in Algorithm 1), with Lemma 5, we can derive

$$\begin{aligned} &\epsilon \leq B(t_k) \\ &\leq \min\left(0, -\Delta(k^*, k) + 4\alpha^M_{ser,t_k}(k)\right) + 2\alpha^M_{ser,t_k}(k). \end{aligned} \tag{36}$$

Substituting (6) (the definition of $\alpha^M_{ser,t_k}(k)$) into (36), we have

$$T_{ser,t_k}(k) \leq \frac{2\sigma^2 \log\left(4K\left((1+\gamma M)\sum_{s=1}^{K} T_{ser,t_k}(s)\right)^2/\delta\right)}{\max\left(\frac{\Delta(k^*,k)+\epsilon}{3}, \epsilon\right)^2}.$$

With (35) and $t_k < \tau$, we can finally bound $T_{ser,\tau}(k)$, i.e.

$$T_{ser,\tau}(k) \leq \frac{2\sigma^2 \log\left(4K\left((1+\gamma M)\sum_{s=1}^{K} T_{ser,\tau}(s)\right)^2/\delta\right)}{\max\left(\frac{\Delta(k^*,k)+\epsilon}{3}, \epsilon\right)^2} + \gamma M \sum_{s=1}^{K} T_{ser,\tau}(s).$$

Here we finish the proof of Lemma 3. □

C PROOF OF THEOREM 2

Recall that the upper confidence bounds of the agent m and server are defined as $\alpha_{m,t}(i,j) = \|\mathbf{y}(i,j)\|_{\mathbf{V}^{-1}_{m_t,t}} C_{m,t}$ and $\alpha^L_{ser,t}(i,j) = \|\mathbf{y}(i,j)\|_{\mathbf{V}^{-1}_{ser,t}} C_{ser,t}$, respectively. The estimated model parameters of the agent m and server are denoted as $\hat{\boldsymbol{\theta}}_{m,t}$ and $\hat{\boldsymbol{\theta}}_{ser,t}$, respectively. Besides, we also provide the Remark 4 to specifically illustrate the relationship between some most important data used in the proof of Theorem 2.

Remark 4 (Global and local data of the linear case)**.** *Due to the transmitted data in the linear case being more complicated than the data in MAB. We here provide new notations to clarify the relations between global data and local data. The following matrix and vector denote the global data*

$$\begin{aligned} &\mathbf{V}^{all}_t = \lambda \mathbf{I} + \sum_{s=1}^{t} \mathbf{x}_{m_s,s}\mathbf{x}^\top_{m_s,s} = \mathbf{V}_{ser,t} + \sum_{m=1}^{M} \mathbf{V}^{loc}_{m,t}, \\ &\mathbf{b}^{all}_t = \sum_{s=1}^{t} \mathbf{x}_{m_s,s} r_{m_s,s} = \mathbf{b}_{ser,t} + \sum_{m=1}^{M} \mathbf{b}^{loc}_{m,t}, \\ &T^{all}_t(k) = \sum_{s=1}^{t} \mathbb{1}\{k = k_{m_s,s}\} = T_{ser,t}(k) + \sum_{m=1}^{M} T^{loc}_{m,t}(k), \; \forall k \in \mathcal{A}. \end{aligned}$$

We define $N_{m,t}$ as the final round when agent m communicates with the server at the end of round t. The collected data of agent m that has not been uploaded to the server is provided as follows

$$\begin{aligned}
\mathbf{V}_{m,t}^{loc} &= \sum_{s=N_{m,t}+1}^{t} \mathbb{1}\{m_s = m\}\mathbf{x}_{m_s,s}\mathbf{x}_{m_s,s}^{\top},\\
\mathbf{b}_{m,t}^{loc} &= \sum_{s=N_{m,t}+1}^{t} \mathbb{1}\{m_s = m\}\mathbf{x}_{m_s,s}r_{m_s,s}\\
T_{m,t}^{loc}(k) &= \sum_{s=N_{m,t}+1}^{t} \mathbb{1}\{m_s = m, k = k_{m_s,s}\},\ \forall k \in \mathcal{A}.
\end{aligned}$$

Similarly, the data that has been uploaded to the server yields

$$\begin{aligned}
\mathbf{V}_{ser,t} &= \lambda\mathbf{I} + \sum_{m=1}^{M}\sum_{s=1}^{N_{m,t}} \mathbb{1}\{m_s = m\}\mathbf{x}_{m_s,s}\mathbf{x}_{m_s,s}^{\top},\\
\mathbf{b}_{ser,t} &= \sum_{m=1}^{M}\sum_{s=1}^{N_{m,t}} \mathbb{1}\{m_s = m\}\mathbf{x}_{m_s,s}r_{m_s,s}\\
T_{ser,t}(k) &= \sum_{m=1}^{M}\sum_{s=1}^{N_{m,t}} \mathbb{1}\{m_s = m, k = k_{m_s,s}\},\ \forall k \in \mathcal{A}.
\end{aligned}$$

According to the communication protocol, the local data of every agent $m \in \mathcal{M}$ can be represented by

$$\mathbf{V}_{m,t} = \mathbf{V}_{ser,N_{m,t}},\ \mathbf{b}_{m,t} = \mathbf{b}_{ser,N_{m,t}},\ T_{m,t}(k) = T_{ser,N_{m,t}}(k),\ \forall k \in \mathcal{A}.$$

Accordingly, we have $V_{m,t} \preceq V_{ser,t}$ and $\sum_{k=1}^{K} T_{m,t}(k) \leq \sum_{k=1}^{K} T_{ser,t}(k)$, $\forall t \in [K+1, \tau]$.

Proof sketch of Theorem 2 The proof of Theorem 2 also consists of three main components, i.e., a) the sample complexity τ; b) the communication cost $\mathcal{C}(\tau)$; c) the estimated best arm satisfies Eq (1). Specifically, to upper bound the total communication cost $C(\tau)$, we utilize the property that the agents communicating with the server when at least one of the two events (line 11 of Algorithm 2) is triggered (Lemma 6). To upper bound the sample complexity τ, we first establish the relationship between $\sum_{k=1}^{K} T_{ser,t}(k)$ and $\sum_{k=1}^{K} T_{m,t}^{loc}(k)$ and the relationship between $\mathbf{V}_{ser,t}$ and $\mathbf{V}_{m,t}^{loc}$ based on the hybrid event triggered strategy (Lemma 9). Then, we design the exploration bonuses by $\sum_{k=1}^{K} T_{m_t,t}(k)$, $\mathbf{V}_{m_t,t}$, $\sum_{k=1}^{K} T_{ser,t}(k)$ and $\mathbf{V}_{ser,t}$ (Lemma 7). Furthermore, we bound the matrix norms $\|\mathbf{y}(i,j)\|_{\mathbf{V}_{m_t,t}^{-1}}$ and $\|\mathbf{y}(i,j)\|_{\mathbf{V}_{ser,t}^{-1}}$, $\forall i,j \in \mathcal{A}$ based on the arm selection strategy (Lemma 10). Combine these knowledge, we can bound $T_{ser,\tau}(k)$, $\forall k \in \mathcal{A}$ (Lemma 11 and 8). Finally, utilizing the knowledge of Remark 4, we can bound $T_{\tau}^{all}(k)$, $\forall k \in \mathcal{A}$, and $\tau = \sum_{k=1}^{K} T_{\tau}^{all}(k)$. Similar to the MAB setting, the guarantee on finding the best arm directly follows the property of the breaking index, i.e., if $B(\tau) \leq \epsilon$, then $\Delta(k^*, \hat{k}^*) \leq \epsilon$ with probability at least $1 - \delta$.

C.1 UPPER BOUND COMMUNICATION COST $\mathcal{C}(\tau)$

Lemma 6 (Communication cost of the hybrid event-triggered strategy in Algorithm 2)**.** *Under the setting of Theorem 2, the total triggered number of the first event (line 11 of Algorithm 2) can be bounded by*

$$(M + 1/\gamma_1)d\log_2\left(1 + \frac{\tau}{\lambda d}\right) \tag{37}$$

and the total triggered number of the second event can be bounded by

$$(M + 1/\gamma_2)\log_2(\tau).$$

The total communication cost can be bounded by

$$\mathcal{C}(\tau) \leq 2\Bigg((M + 1/\gamma_1)d\log_2\left(1 + \frac{\tau}{\lambda d}\right) + (M + 1/\gamma_2)\log_2(\tau)\Bigg).$$

Proof of Lemma 6. The triggered number of the second event can be bounded by Lemma 1. Besides, we can bound the triggered number of the first event similar to He et al. [2022]. The proof of (37) also can be divided into two sections, in the first section, we would divide the sample complexity τ into $\log_2(1+\tau/\lambda d)$ episodes, then we would analysis the upper bound of the triggered number of the first event in each episode. We define

$$\mathcal{T}_i = \min\Big\{t \in |\tau|,\ \det(\mathbf{V}_{ser,t}) \geq 2^i \lambda^d\Big\}$$

and the set of all rounds into episodes $\{\mathcal{T}_i, \mathcal{T}_i + 1, ..., \mathcal{T}_{i+1} - 1\}, \forall i \geq 0$. By the Lemma 14, we can bound

$$\det(\mathbf{V}_{ser,\tau}) \leq \lambda^d \Big(1 + \frac{\tau}{\lambda d}\Big)^d.$$

Accordingly, the number of the episode can be bounded by

$$\max\{i \geq 0\} = \log_2\left(\frac{\det(\mathbf{V}_{ser,\tau})}{\lambda^d}\right) \leq d \log_2\left(1 + \frac{\tau}{\lambda d}\right).$$

We then prove $\forall i \geq 0$, from round $\mathcal{T}_i$ to $\mathcal{T}_{i+1} - 1$, the triggered number of the first event can be bounded by $M + 1/\gamma_1$. We first define the number of agents m triggers the first event in $\mathcal{T}_i$ to $\mathcal{T}_{i+1} - 1$ as $\mathcal{N}_m$, the sequence of agent m triggers the first event in round $\mathcal{T}_i$ to $\mathcal{T}_{i+1} - 1$ as $t_1^m, ..., t_{\mathcal{N}_m}^m$, the number of every agent triggers the first event in $\mathcal{T}_i$ to $\mathcal{T}_{i+1} - 1$ as L and the sequence of the first event be triggered in $\mathcal{T}_i$ to $\mathcal{T}_{i+1} - 1$ as $t_{i,1}, ..., t_{i,L}$. According to the definition of the first event, we have

$$\det(\mathbf{V}_{m_t,t} + \mathbf{V}^{loc}_{m_t,t}) > (1 + \gamma_1)\det(\mathbf{V}_{m_t,t}).$$

Then, $\forall m \in \mathcal{M},\ j \in |\mathcal{N}_m|/\{1\}$, we have

$$\det(\mathbf{V}_{ser,\mathcal{T}_i} + \mathbf{V}^{loc}_{m,t_j^m}) \geq \frac{\det(\mathbf{V}_{ser,\mathcal{T}_i})}{\det(\mathbf{V}_{m,t_j^m})}\det(\mathbf{V}_{m,t_j^m} + \mathbf{V}^{loc}_{m,t_j^m}) \geq (1 + \gamma_1)\det(\mathbf{V}_{ser,\mathcal{T}_i}) \tag{38}$$

The inequality holds due to $\forall j \in |\mathcal{N}_m|/\{1\}$, $\mathbf{V}_{ser,\mathcal{T}_i} \preceq \mathbf{V}_{m,t_j^m}$ and Lemma 15. The above inequality implies $\forall t_{i,l} \geq t_2^{m_{t_{i,l}}}$

$$\begin{aligned}
\det(\mathbf{V}_{ser,t_{i,l}} - \mathbf{V}_{ser,t_{i,l-1}}) &= \det(\mathbf{V}_{ser,t_{i,l-1}} + \mathbf{V}^{loc}_{m_{t_{i,l}},t_{i,l}}) - \det(\mathbf{V}_{ser,t_{i,l-1}}) \\
&\geq \det(\mathbf{V}_{ser,\mathcal{T}_i} + \mathbf{V}^{loc}_{m_{t_{i,l}},t_{i,l}}) - \det(\mathbf{V}_{ser,\mathcal{T}_i}) \\
&\geq \gamma_1 \det(\mathbf{V}_{ser,\mathcal{T}_i}).
\end{aligned}$$

The first inequality holds is owing to Lemma 16 and the last inequality holds is owing to (38).

Finally we can bound $L = \sum_{m=1}^{M} \mathcal{N}_m$

$$\begin{aligned}
\det(\mathbf{V}_{ser,\mathcal{T}_{i+1}-1}) - \det(\mathbf{V}_{ser,\mathcal{T}_i}) &= \sum_{l=1}^{L-1}\Big(\det(\mathbf{V}_{ser,t_{i,l+1}}) - \det(\mathbf{V}_{ser,t_{i,l}})\Big) \\
&\geq \gamma_1 \sum_{m=1}^{M} (\mathcal{N}_m - 1)\det(\mathbf{V}_{ser,\mathcal{T}_i}).
\end{aligned} \tag{39}$$

Due to the definition of the episode, it has $2\det(\mathbf{V}_{ser,\mathcal{T}_i}) \geq \det(\mathbf{V}_{ser,\mathcal{T}_{i+1}-1})$. We can rewrite equation (39) as

$$M + 1/\gamma_1 \geq \sum_{m=1}^{M} \mathcal{N}_m.$$

We can then bound the total triggered number of the first event by

$$(M + 1/\gamma_2)\log_2\left(1 + \frac{\tau}{\lambda d}\right).$$

Due to the communication would happen when at least one of the events is triggered, the total communication round is smaller or equal to the triggered number of two events. Hence, the total communication number from $t = K + 1$ to $t = \tau$ can be bounded by

$$(M + 1/\gamma_1) d \log_2 \left(1 + \frac{\tau}{\lambda d} + \tau^{1/d}\right).$$

Furthermore, due to one communication includes one upload and one download, the total communication cost can be bounded by

$$\mathcal{C}(\tau) \leq 2\Bigg((M + 1/\gamma_1) d \log_2 \left(1 + \frac{\tau}{\lambda d}\right) + (M + 1/\gamma_2) \log_2(\tau)\Bigg). \tag{40}$$

With (40), (44) (the upper bound of the sample complexity), and the setting of Theorem (2), we can bound the communication cost

$$\mathcal{C}(\tau) = O\Big(\max(2MK, M^2)\Big).$$

Here we finish the proof of Lemma 6. □

C.2 UPPER BOUND SAMPLE COMPLEXITY τ

Combine the breaking condition of the Algorithm 2 (line 14∼16) and definition of $B(\tau)$, we have

$$\epsilon \geq \hat{\Delta}_{ser,\tau}(j_{ser,\tau}, i_{ser,\tau}) + \alpha^L_{ser,\tau}(i_{ser,\tau}, j_{ser,\tau}) = B(\tau).$$

Let's first consider the case when the empirically best arm on the server side is not the optimal arm, i.e., $i_{ser,\tau} \neq k^*$. By the definition of the $j_{ser,t}$, we have

$$\hat{\Delta}_{ser,\tau}(j_{ser,\tau}, i_{ser,\tau}) + \alpha^L_{ser,\tau}(i_{ser,\tau}, j_{ser,\tau}) \geq \hat{\Delta}_{ser,\tau}(k^*, i_{ser,\tau}) + \alpha^L_{ser,\tau}(i_{ser,\tau}, k^*).$$

Recall that $\hat{k}^* = i_{ser,\tau}$ is the estimated best arm. Therefore, we have

$$\epsilon \geq \hat{\Delta}_{ser,\tau}(k^*, \hat{k}^*) + \alpha^L_{ser,\tau}(\hat{k}^*, k^*) \geq \Delta(k^*, \hat{k}^*),$$

where the second inequality is due to Lemma 7 below (proof of Lemma 7 is at the end of the section).

Lemma 7. *Following the setting of Theorem 2, we define event*

$$\mathcal{I} = \left\{\forall i, j \in \mathcal{A}, \forall t \in [K+1, \tau],\ |\hat{\Delta}_{m_t,t}(i,j) - \Delta(i,j)| \leq \alpha^L_{m_t,t}(i,j),\ |\hat{\Delta}_{ser,t}(i,j) - \Delta(i,j)| \leq \alpha^L_{ser,t}(i,j)\right\}$$

where

$$\alpha^L_{m_t,t}(i,j) = \left(\sqrt{\lambda} + \left(\sqrt{2\gamma_1} M + \sqrt{1 + \gamma_1 M}\right)\left(\sigma\sqrt{d \log\left(\frac{2}{\delta}\left(1 + \frac{(1+\gamma_2 M)\sum_{k=1}^K T_{m_t,t}(k)}{\min(\gamma_1, 1)\lambda}\right)\right)}\right)\right) \|\mathbf{y}(i,j)\|_{\mathbf{V}^{-1}_{m_t,t}}$$

$$\alpha^L_{ser,t}(i,j) = \left(\sqrt{\lambda} + \left(\sqrt{2\gamma_1} M + \sqrt{1 + \gamma_1 M}\right)\left(\sigma\sqrt{d \log\left(\frac{2}{\delta}\left(1 + \frac{(1+\gamma_2 M)\sum_{k=1}^K T_{ser,t}(k)}{\min(\gamma_1, 1)\lambda}\right)\right)}\right)\right) \|\mathbf{y}(i,j)\|_{\mathbf{V}^{-1}_{ser,t}}.$$

We have $\mathcal{P}(\mathcal{I}) \geq 1 - \delta$.

Moreover, when $i_{ser,\tau} = k^*$, we can trivially derive $\Delta(k^*, \hat{k}^*) = 0 \leq \epsilon$. The above discussion implies $\hat{k}^*$ output by `FALinPE` satisfies the (ϵ, δ)-condition (1).

We now continue to bound the sample complexity τ. First, we need to establish Lemma 8 below, which upper bounds $T_{ser,\tau}(k)$, the number of observation on arm k that is available to the server at τ.

Lemma 8. *Under the setting of Theorem 2 and event $\mathcal{I}$, we can bound*

$$T_{ser,\tau}(k) \le \max_{i,j\in\mathcal{A}} \frac{\rho(\mathbf{y}(i,j))p_k^*(\mathbf{y}(i,j))}{\max\left(\frac{\Delta(k^*,i)+\epsilon}{3}, \frac{\Delta(k^*,j)+\epsilon}{3}, \epsilon\right)^2} C_{ser,\tau}^2 + \gamma_2 M \sum_{s=1}^K T_{ser,\tau}(s),\ \forall k \in \mathcal{A}.$$

With Lemma 8, we can derive

$$\begin{aligned}\sum_{k=1}^K T_{ser,\tau}(k) &\le \sum_{k=1}^K \max_{i,j\in\mathcal{A}} \frac{\rho(\mathbf{y}(i,j))p_k^*(\mathbf{y}(i,j))}{\max\left(\frac{\Delta(k^*,i)+\epsilon}{3}, \frac{\Delta(k^*,j)+\epsilon}{3}, \epsilon\right)^2} C_{ser,\tau}^2 + \gamma_2 KM \sum_{s=1}^K T_{ser,\tau}(s) \\ &\le \frac{1}{1-\gamma_2 KM} \sum_{k=1}^K \max_{i,j\in\mathcal{A}} \frac{\rho(\mathbf{y}(i,j))p_k^*(\mathbf{y}(i,j))}{\max\left(\frac{\Delta(k^*,i)+\epsilon}{3}, \frac{\Delta(k^*,j)+\epsilon}{3}, \epsilon\right)^2} C_{ser,\tau}^2.\end{aligned}$$

Furthermore, based on the relationship between τ and $\sum_{k=1}^K T_{ser,\tau}(k)$ (Remark 4), we have

$$\begin{aligned}\tau &= \sum_{k=1}^K T_{ser,\tau}(k) + \sum_{m=1}^M \sum_{k=1}^K T_{m,\tau}^{loc}(k) \\ &\le \left(1+\gamma_2 M\right) \sum_{k=1}^K T_{ser,\tau}(k) \\ &\le \frac{1+\gamma_2 M}{1-\gamma_2 KM} \sum_{k=1}^K \max_{i,j\in\mathcal{A}} \frac{\rho(\mathbf{y}(i,j))p_k^*(\mathbf{y}(i,j))}{\max\left(\frac{\Delta(k^*,i)+\epsilon}{3}, \frac{\Delta(k^*,j)+\epsilon}{3}, \epsilon\right)^2} C_{ser,\tau}^2 \\ &= \frac{1+\gamma_2 M}{1-\gamma_2 KM} H_\epsilon^L C_{ser,\tau}^2\end{aligned} \tag{41}$$

where the second inequality is owing to the inequality we establish above and the last equality is owing to the definition of H_ϵ^L.

Recalling that we suppose $\gamma_1 = 1/(M^2)$, $\gamma_2 = 1/(2MK)$ and $0 < \lambda \le \sigma^2\left(\sqrt{1+\gamma_1 M} + \sqrt{2\gamma_1}M\right)^2 \log(2/\delta)$. We first need to decompose $C_{ser,\tau}$

$$\begin{aligned}C_{ser,\tau} &= \sqrt{\lambda} + \left(\sqrt{2}+\sqrt{1+\frac{1}{M}}\right)^2 \left(\sigma\sqrt{d\log\left(\frac{2}{\delta}\left(1+\frac{(1+1/(2K))\sum_{k=1}^K T_{ser,\tau}(k)}{\lambda/M^2}\right)\right)}\right) \\ &= \sqrt{\lambda} + \left(\sqrt{2}+\sqrt{1+\frac{1}{M}}\right)^2 \left(\sigma\sqrt{d\log\left(\frac{2}{\delta}\right) + d\log\left(1+\frac{(1+1/(2K))\sum_{k=1}^K T_{ser,\tau}(k)}{\lambda/M^2}\right)}\right) \\ &\le 2\left(\sqrt{2}+\sqrt{1+\frac{1}{M}}\right)^2 \left(\sigma\sqrt{d\log\left(\frac{2}{\delta}\right) + d\log\left(1+\frac{(1+1/(2K))\sum_{k=1}^K T_{ser,\tau}(k)}{\lambda/M^2}\right)}\right) \\ &\le 2\left(\sqrt{2}+\sqrt{1+\frac{1}{M}}\right)^2 \left(\sigma\sqrt{d\log\left(\frac{2}{\delta}\right) + d\log\left(1+\frac{(1+1/(2K))\tau}{\lambda/M^2}\right)}\right).\end{aligned} \tag{42}$$

The first inequality is owing to the definition of λ and the last inequality is owing to $\sum_{k=1}^K T_{ser,\tau}(k) \le \tau$. Substituting the last term of (42) into (41), we have

$$\tau \le \frac{M+1/(2K)}{M-1/2}\left(\sqrt{2}+\sqrt{1+\frac{1}{M}}\right)^2 H_\epsilon^L 4\sigma^2 d\left(\log\left(\frac{2}{\delta}\right) + \log\left(1+\frac{(1+1/(2K))\tau}{\lambda/M^2}\right)\right).$$

We define

$$\Gamma = \frac{M+1/(2K)}{M-1/2}\left(\sqrt{2}+\sqrt{1+\frac{1}{M}}\right)^2 H_\epsilon^L 4\sigma^2 d\log\left(\frac{2}{\delta}\right).$$

Let τ' be a parameter satisfies

$$\tau' \leq \tau = \frac{M + 1/(2K)}{M - 1/2}\left(\sqrt{2} + \sqrt{1 + \frac{1}{M}}\right)^2 H_\epsilon^L 4\sigma^2 d \log\left(1 + \frac{(1 + 1/(2K))\tau'}{\lambda/M^2}\right) + \Gamma$$
$$\leq \frac{M + 1/(2K)}{M - 1/2}\left(\sqrt{2} + \sqrt{1 + \frac{1}{M}}\right)^2 H_\epsilon^L 4\sigma^2 d \sqrt{1 + \frac{(1 + 1/(2K))\tau'}{\lambda/M^2}} + \Gamma.$$

We can further derive

$$\sqrt{\tau'} \leq \frac{M + 1/(2K)}{M - 1/2}\left(\sqrt{2} + \sqrt{1 + \frac{1}{M}}\right)^2 H_\epsilon^L 4\sigma^2 d \sqrt{1 + \frac{(1 + 1/(2K))\tau'}{\lambda/M^2}} + \Gamma = \Lambda. \tag{43}$$

In the light of (43), we can finally bound τ by

$$\tau \leq \frac{M + 1/(2K)}{M - 1/2}\left(\sqrt{2} + \sqrt{1 + \frac{1}{M}}\right)^2 H_\epsilon^L 4\sigma^2 d \sqrt{1 + \frac{(1 + 1/(2K))\Lambda^2}{\lambda/M^2}} + \Gamma. \tag{44}$$

Lemma 9. *Under the setting of Thoerem 2 and the communication strategy of line 11 in Algorithm 2, we can derive*

$$\mathbf{V}_{ser,t} \succeq (1/\gamma_1)\mathbf{V}_{m,t}^{loc},\ \forall m \in \mathcal{M},\ t \in [K+1, \tau]. \tag{45}$$

Furthermore, following the results of Lemma 4, we can also derive

$$\sum_{k=1}^{K} T_{ser,t}(k) \geq (1/\gamma_2) \sum_{k=1}^{K} T_{m,t}^{loc}(k),\ \forall m \in \mathcal{M},\ t \in [K+1, \tau]. \tag{46}$$

Proof for Lemma 7, Lemma 8, and Lemma 9 In the following paragraphs, we provide the detailed proof for the lemmas used above.

Proof of Lemma 9. The proof of (45) is similar to the proof in He et al. [2022]. Suppose the last round of agent m communicates with the server is t_1 and the first event is triggered. Then, we can trivially derive

$$\mathbf{V}_{ser,t} \succ \mathbf{0} = \mathbf{V}_{m,t}^{loc},\ \forall t \in [K+1, \tau].$$

Otherwise, according to the definition of the first event, we have

$$\text{det}(\mathbf{V}_{m,t} + \mathbf{V}_{m,t}^{loc}) \leq (1 + \gamma_1)\text{det}(\mathbf{V}_{m,t}).$$

Based on the Lemma 17, we have

$$1 + \gamma_1 \geq \frac{\text{det}(\mathbf{V}_{m,t} + \mathbf{V}_{m,t}^{loc})}{\text{det}(\mathbf{V}_{m,t})} \geq \frac{\|\mathbf{x}\|^2_{\mathbf{V}_{m,t} + \mathbf{V}_{m,t}^{loc}}}{\|\mathbf{x}\|^2_{\mathbf{V}_{m,t}}} = \frac{\|\mathbf{x}\|^2_{\mathbf{V}_{m,t}} + \|\mathbf{x}\|^2_{\mathbf{V}_{m,t}^{loc}}}{\|\mathbf{x}\|^2_{\mathbf{V}_{m,t}}}$$

and

$$\mathbf{V}_{m,t} \succeq (1/\gamma_1)\mathbf{V}_{m,t}^{loc}.$$

With the fact that $\mathbf{V}_{ser,t} \succeq \mathbf{V}_{m,t}, \forall t \in [K+1, \tau]$, we can finish the proof of (45).

The proof of (46) is similar to the proof of Lemma 4. Combine the two results and we can finish the whole proof of Lemma 9. □

Based on Lemma 9, we can prove Lemma 7 as shown below.

Proof of Lemma 7. Following the same argument of Lemma 5, we only need to proof

$$\mathcal{P}(\mathcal{I}) = \mathcal{P}\bigg(\forall i, j \in \mathcal{A}, \forall t \in [K+1, \tau],\ |\hat{\Delta}_{ser,t}(i,j) - \Delta(i,j)| \le \alpha^{L}_{ser,t}(i,j)\bigg) \ge 1 - \delta.$$

Decompose $|\hat{\Delta}_{ser,t}(i,j) - \Delta(i,j)|$, we have

$$\begin{aligned}|\Delta(i,j) - \hat{\Delta}_{ser,t}(i,j)| &= |\mathbf{y}(i,j)^\top \boldsymbol{\theta}^* - \mathbf{y}(i,j)^\top \hat{\boldsymbol{\theta}}_{ser,t}| \\ &= |\mathbf{y}(i,j)^\top (\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}_{ser,t})| \\ &\le \|\mathbf{y}(i,j)\|_{\mathbf{V}^{-1}_{ser,t}} \|\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}_{ser,t}\|_{\mathbf{V}_{ser,t}}.\end{aligned}$$

Hence, according to the definition of the $\alpha_{ser,t}(i,j)$, we can derive

$$\mathcal{P}(\mathcal{I}) \ge \mathcal{P}\Big(\forall t \in [K+1, \tau],\ \|\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}_{ser,t}\|_{\mathbf{V}_{ser,t}} \le C_{ser,t}\Big) \tag{47}$$

and only need to proof

$$\mathcal{P}\Big(\forall t \in [K+1, \tau],\ \|\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}_{ser,t}\|_{\mathbf{V}_{ser,t}} \le C_{ser,t}\Big) \ge 1 - \delta. \tag{48}$$

We first discompose $\|\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}_{ser,t}\|_{\mathbf{V}_{ser,t}}$

$$\begin{aligned}\|\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}_{ser,t}\|_{\mathbf{V}_{ser,t}} &= \|\boldsymbol{\theta}^* - \mathbf{V}^{-1}_{ser,t}\mathbf{b}_{ser,t}\|_{\mathbf{V}_{ser,t}} \\ &= \Big\|\boldsymbol{\theta}^* - \mathbf{V}^{-1}_{ser,t}\Big((\mathbf{V}_{ser,t} - \lambda\mathbf{I})\boldsymbol{\theta}^* + \sum_{m=1}^{M}\sum_{s=1}^{N_{m,t}} \mathbb{1}\{m_s = m\}\mathbf{x}_{m,s}\eta_{m,s}\Big)\Big\|_{\mathbf{V}_{ser,t}} \\ &= \Big\|\boldsymbol{\theta}^* - \boldsymbol{\theta}^* + \lambda\mathbf{V}^{-1}_{ser,t}\boldsymbol{\theta}^* + \mathbf{V}^{-1}_{ser,t}\sum_{m=1}^{M}\sum_{s=1}^{N_{m,t}} \mathbb{1}\{m_s = m\}\mathbf{x}_{m,s}\eta_{m,s}\Big\|_{\mathbf{V}_{ser,t}} \\ &= \|\lambda\mathbf{V}^{-1}_{ser,t}\boldsymbol{\theta}^*\|_{\mathbf{V}_{ser,t}} + \Big\|\mathbf{V}^{-1}_{ser,t}\sum_{m=1}^{M}\sum_{s=1}^{N_{m,t}} \mathbb{1}\{m_s = m\}\mathbf{x}_{m,s}\eta_{m,s}\Big\|_{\mathbf{V}_{ser,t}} \\ &= \lambda\|\boldsymbol{\theta}^*\|_{\mathbf{V}^{-1}_{ser,t}} + \Big\|\mathbf{V}^{-1}_{ser,t}\Big(\sum_{s=1}^{t}\mathbf{x}_{m_s,s}\eta_{m_s,s} - \sum_{m=1}^{M}\sum_{s=N_{m,t}+1}^{t} \mathbb{1}\{m_s = m\}\mathbf{x}_{m,s}\eta_{m,s}\Big)\Big\|_{\mathbf{V}_{ser,t}} \\ &\le \sqrt{\lambda} + \underbrace{\Big\|\mathbf{V}^{-1}_{ser,t}\sum_{s=1}^{t}\mathbf{x}_{m_s,s}\eta_{m_s,s}\Big\|_{\mathbf{V}_{ser,t}}}_{\Lambda} + \underbrace{\Big\|\mathbf{V}^{-1}_{ser,t}\sum_{m=1}^{M}\sum_{s=N_{m,t}+1}^{t} \mathbb{1}\{m_s = m\}\mathbf{x}_{m,s}\eta_{m,s}\Big\|_{\mathbf{V}_{ser,t}}}_{\Gamma},\end{aligned} \tag{49}$$

where the last inequality is owing to $\mathbf{V}_{ser,t} \succeq \lambda\mathbf{I}$. We further decompose term Λ and Γ. Based on the Lemma 18, we have $\forall t \in [K+1, \tau]$

$$\Big\|\sum_{s=1}^{t}\mathbf{x}_{m_s,s}\eta_{m_s,s}\Big\|_{\mathbf{V}^{all-1}_{t}} \le \sigma\sqrt{d\log\Big(\frac{2(1+t/\lambda)}{\delta}\Big)} \tag{50}$$

holds with probability at least $1 - \delta/2$. With a union bound and utilize the self normalized martingale again, it holds that for each $m \in \mathcal{M}$ and $\forall t \in [K+1, \tau]$

$$\Big\|\sum_{s=N_{m,t}+1}^{t}\mathbb{1}\{m_s = m\}\mathbf{x}_{m_s,s}\eta_{m_s,s}\Big\|_{(\gamma_1\lambda\mathbf{I} + \mathbf{V}^{loc}_{m,t})^{-1}} \le \sigma\sqrt{d\log\Big(\frac{2(1+t/(\gamma_1\lambda))}{\delta}\Big)} \tag{51}$$

holds with probability at least $1 - \delta/2$.

According to the Lemma 9 and (50), $\forall t \in [K+1, \tau]$, Λ can be bounded by

$$\begin{aligned}
&\Big\|\mathbf{V}_{ser,t}^{-1}\sum_{s=1}^{t}\mathbf{x}_{m_s,s}\eta_{m_s,s}\Big\|_{\mathbf{V}_{ser,t}} \\
=&\Big\|\sum_{s=1}^{t}\mathbf{x}_{m_s,s}\eta_{m_s,s}\Big\|_{\mathbf{V}_{ser,t}^{-1}} \\
\leq&\sqrt{1+\gamma_1 M}\Big\|\sum_{s=1}^{t}\mathbf{x}_{m_s,s}\eta_{m_s,s}\Big\|_{\mathbf{V}_t^{all-1}} \\
\leq&\sqrt{1+\gamma_1 M}\Big(\sigma\sqrt{d\log\Big(\frac{2(1+t/\lambda)}{\delta}\Big)}\Big)
\end{aligned} \tag{52}$$

with probability at least $1-\delta/2$. The first inequality holds due to Lemma 9 and the last inequality holds according to (50). Besides, in the light of (51), $\forall t \in [K+1, \tau]$, Γ can be bounded by

$$\begin{aligned}
&\Big\|\mathbf{V}_{ser,t}^{-1}\sum_{m=1}^{M}\sum_{s=N_{m,t}+1}^{t}\mathbb{1}\{m_s=m\}\mathbf{x}_{m,s}\eta_{m,s}\Big\|_{\mathbf{V}_{ser,t}} \\
=&\Big\|\sum_{m=1}^{M}\sum_{s=N_{m,t}+1}^{t}\mathbb{1}\{m_s=m\}\mathbf{x}_{m_s,s}\eta_{m_s,s}\Big\|_{\mathbf{V}_{ser,t}^{-1}} \\
\leq&\sum_{m=1}^{M}\Big\|\sum_{s=N_{m,t}+1}^{t}\mathbb{1}\{m_s=m\}\mathbf{x}_{m_s,s}\eta_{m_s,s}\Big\|_{\mathbf{V}_{ser,t}^{-1}} \\
\leq&\sqrt{2\gamma_1}\sum_{m=1}^{M}\Big\|\sum_{s=N_{m,t}+1}^{t}\mathbb{1}\{m_s=m\}\mathbf{x}_{m_s,s}\eta_{m_s,s}\Big\|_{(\mathbf{V}_{m,t}^{loc}+\gamma_1\lambda\mathbf{I})^{-1}} \\
\leq&\sqrt{2\gamma_1}M\Big(\sigma\sqrt{d\log\Big(\frac{2(1+t/(\gamma_1\lambda))}{\delta}\Big)}\Big)
\end{aligned} \tag{53}$$

with probability at least $1-\delta/2$. The second inequality holds is due to $\forall m \in \mathcal{M}$

$$\begin{aligned}
\mathbf{V}_{ser,t} &\succeq \frac{1}{\gamma_1}\mathbf{V}_{m,t}^{loc} \\
\frac{1}{2}\mathbf{V}_{ser,t}+\frac{1}{2}\mathbf{V}_{ser,t} &\succeq \frac{1}{2}\lambda\mathbf{I}+\frac{1}{2\gamma_1}\mathbf{V}_{m,t}^{loc}.
\end{aligned}$$

Combine (49), (52) and (53), due to the server or the agents can not directly derive t, we can utilize $(1+\gamma_2 M)\sum_{k=1}^{K}T_{ser,t}(k)$ to replace t and the above inequalities still hold. We can finally get $\forall t \in [K+1, \tau]$

$$\|\boldsymbol{\theta}^*-\hat{\boldsymbol{\theta}}_{ser,t}\|_{\mathbf{V}_{ser,t}} \leq \sqrt{\lambda}+\Big(\sqrt{2\gamma_1}M+\sqrt{1+\gamma_1 M}\Big)\Big(\sigma\sqrt{d\log\Big(\frac{2}{\delta}\Big(1+\frac{(1+\gamma_2 M)\sum_{k=1}^{K}T_{ser,t}(k)}{\min(\gamma_1,1)\lambda}\Big)\Big)}\Big)$$

holds with probability at least $1-\delta$. Combine this with (48), here we finish the proof of Lemma 7. □

Before proving Lemma 8, we first need to establish Lemma 10 and Lemma 11 below.

Lemma 10. *Following the setting of Theorem 2, $\forall t \in [K+1, \tau]$, The matrix norm $\|\mathbf{y}(i,j)\|_{\mathbf{V}_{m_t,t}^{-1}}$ can be bounded by*

$$\|\mathbf{y}(i,j)\|_{V_{m_t,t}^{-1}} \leq \sqrt{\frac{\rho(\mathbf{y}(i,j))}{T_{m_t,t}(i,j)}},\ \textit{and}\ \|\mathbf{y}(i,j)\|_{V_{ser,t}^{-1}} \leq \sqrt{\frac{\rho(\mathbf{y}(i,j))}{T_{ser,t}(i,j)}},\ \forall i,j \in \mathcal{A}, \tag{54}$$

where

$$\begin{aligned}
T_{m_t,t}(i,j) &= \min_{k\in\mathcal{A},\ p_k^*(\mathbf{y}(i,j))>0} T_{m_t,t}(k)/p_k^*(\mathbf{y}(i,j)) \\
T_{ser,t}(i,j) &= \min_{k\in\mathcal{A},\ p_k^*(\mathbf{y}(i,j))>0} T_{ser,t}(k)/p_k^*(\mathbf{y}(i,j)).
\end{aligned} \tag{55}$$

Proof of Lemma 10. According to the Lemma 2 of Xu et al. [2017], the optimal value of (13) (i.e., $\rho(\mathbf{y}(i,j))$) is equal to the optimal value of

$$\begin{aligned} &\min_{p_k, w_k} \sum_{k=1}^{K} \frac{w_k^2}{p_k} \\ s.t. \quad &\mathbf{y}(i,j) = \sum_{k=1}^{K} w_k \mathbf{x}_k \\ &\sum_{k=1}^{K} p_k = 1,\ p_k > 0,\ w_k \in \mathcal{R}, \end{aligned} \tag{56}$$

for all $i, j \in \mathcal{A}$.

Due to $\mathbf{V}_{m_t,t}$ and $T_{m_t,t}(k)$, $\forall k \in \mathcal{A},\ t \in [K+1, \tau]$ are all downloaded from the server. This implies $\forall t_1 \in [K+1, \tau]$, there exists a $t_2 \in [K+1, \tau]$ which satisfies

$$\mathbf{V}_{m_{t_1},t_1} = \mathbf{V}_{ser,t_2} \text{ and } T_{m_{t_1},t_1}(k) = T_{ser,t_2}(k),\ \forall k \in \mathcal{A}.$$

Therefore, we only need to prove the second inequality of (54). We can decompose the covariance matrix $\mathbf{V}_{ser,t} = \lambda \mathbf{I} + \sum_{k=1}^{K} T_{ser,t}(k) \mathbf{x}_k \mathbf{x}_k^\top$. We define the auxiliary covariance matrix as $\tilde{\mathbf{V}}_{ser,t} = \lambda \mathbf{I} + \sum_{k=1}^{K} T_{ser,t}(i,j) p_k^*(\mathbf{y}(i,j)) \mathbf{x}_k \mathbf{x}_k^\top$. From (55), we have

$$T_{ser,t}(i,j) p_k^*(\mathbf{y}(i,j)) \leq T_{ser,t}(k),\ \forall k, i, j \in \mathcal{A}$$

which implies $\tilde{\mathbf{V}}_{ser,t} \preceq \mathbf{V}_{ser,t}$ and

$$\mathbf{y}(i,j)^\top \mathbf{V}_{ser,t}^{-1} \mathbf{y}(i,j) \leq \mathbf{y}(i,j)^\top \tilde{\mathbf{V}}_{ser,t}^{-1} \mathbf{y}(i,j),\ i, j \in \mathcal{A}.$$

We then bound $\mathbf{y}(i,j)^\top \tilde{\mathbf{V}}_{ser,t}^{-1} \mathbf{y}(i,j)$, according to the KKT condition of (56), we have the following formulas

$$\begin{aligned} w_k^*(\mathbf{y}(i,j)) &= \frac{1}{2} p_k^*(\mathbf{y}(i,j)) \mathbf{x}_k^\top \varepsilon,\ \forall k, i, j \in \mathcal{A} \\ \mathbf{y}(i,j) &= \frac{1}{2} \sum_{k=1}^{K} p_k^*(\mathbf{y}(i,j)) \mathbf{x}_k \mathbf{x}_k^\top \varepsilon,\ \forall i, j \in \mathcal{A}, \end{aligned}$$

where $\varepsilon \in \mathcal{R}^d$ corresponds to the Lagrange multiplier. Hence, we can rewrite $\mathbf{y}(i,j)^\top \tilde{\mathbf{V}}_{ser,t}^{-1} \mathbf{y}(i,j)$ as

$$\mathbf{y}(i,j)^\top \tilde{\mathbf{V}}_{ser,t}^{-1} \mathbf{y}(i,j) = \frac{1}{4} \bigg(\sum_{k=1}^{K} p_k^*(\mathbf{y}(i,j)) \mathbf{x}_k \mathbf{x}_k^\top \varepsilon \bigg)^\top \tilde{\mathbf{V}}_{ser,t}^{-1} \bigg(\sum_{k=1}^{K} p_k^*(\mathbf{y}(i,j)) \mathbf{x}_k \mathbf{x}_k^\top \varepsilon \bigg). \tag{57}$$

Besides, based on (56), we can rewrite $\rho(\mathbf{y}(i,j))$ as

$$\rho(\mathbf{y}(i,j)) = \sum_{k=1}^{K} \frac{w_k^{*2}(\mathbf{y}(i,j))}{p_k^*(\mathbf{y}(i,j))} = \frac{1}{4} \varepsilon^\top \bigg(\sum_{k=1}^{K} p_k^*(\mathbf{y}(i,j)) \mathbf{x}_k \mathbf{x}_k^\top \bigg) \varepsilon. \tag{58}$$

In the light of (57) and (58), we can bound $\mathbf{y}(i,j)^\top \tilde{\mathbf{V}}_{ser,t}^{-1} \mathbf{y}(i,j) - \rho(\mathbf{y}(i,j))/T_{ser,t}(i,j)$ with 0

$$\begin{aligned} &\mathbf{y}(i,j)^\top \tilde{\mathbf{V}}_{ser,t}^{-1} \mathbf{y}(i,j) - \frac{\rho(\mathbf{y}(i,j))}{T_{ser,t}(i,j)} \\ =&\frac{1}{4} \varepsilon^\top \Bigg(\bigg(\sum_{k=1}^{K} p_k^*(\mathbf{y}(i,j)) \mathbf{x}_k \mathbf{x}_k^\top \bigg) - \frac{\tilde{\mathbf{V}}_{ser,t}}{T_{ser,t}(i,j)} \Bigg) \tilde{\mathbf{V}}_{ser,t}^{-1} \bigg(\sum_{k=1}^{K} p_k^*(\mathbf{y}(i,j)) \mathbf{x}_k \mathbf{x}_k^\top \bigg) \varepsilon \\ =&-\frac{\lambda}{4} \varepsilon^\top \tilde{\mathbf{V}}_{ser,t}^{-1} \bigg(\sum_{k=1}^{K} p_k^*(\mathbf{y}(i,j)) \mathbf{x}_k \mathbf{x}_k^\top \bigg) \varepsilon \\ \leq&0. \end{aligned}$$

The second equality holds due to the definition of the $\tilde{\mathbf{V}}_{ser,t}$, and the last inequality holds due to $\lambda > 0$ and the definition of the positive definite matrix. Here we finish the proof of Lemma 10. □

Lemma 11. *Under the setting of Theorem 2 and event $\mathcal{I}$, $\forall t \in [K+1, \tau]$, $B(t)$ can be bounded as follows*

$$B(t) \leq \min\Big(0, -\max\Big(\Delta(k^*, i_{ser,t}), \Delta(k^*, j_{ser,t})\Big) + 2\alpha^L_{ser,t}(i_{ser,t}, j_{ser,t})\Big) + \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}).$$

Proof of Lemma 11. This proof is similar to the proof of Lemma 5. According to the definition of the event $\mathcal{I}$, consider the case when $i_{ser,t} = k^*$, we have

$$\begin{aligned} B(t) &= \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\leq \Delta(j_{ser,t}, i_{ser,t}) + 2\alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &= -\Delta(k^*, j_{ser,t}) + 2\alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}). \end{aligned} \tag{59}$$

Consider the case when $j_{ser,t} = k^*$, we have

$$\begin{aligned} B(t) &= \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\leq -\hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^L_{ser,t}(i^L_{ser,t}, j_{ser,t}) \\ &\leq -\Delta(j_{ser,t}, i_{ser,t}) + 2\alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &= -\Delta(k^*, i_{ser,t}) + 2\alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}), \end{aligned} \tag{60}$$

where the first inequality is owing to $\hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) \leq 0$.

Combine (59) and (60), it yields

$$B(t) \leq \min\Big(0, -\max\Big(\Delta(k^*, i_{ser,t}), \Delta(k^*, j_{ser,t})\Big) + \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t})\Big) + \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \tag{61}$$

when $i_{ser,t} = k^*$ or $j_{ser,t} = k^*$.

Consider the case when $i_{ser,t} \neq k^*$ and $j_{ser,t} \neq k^*$, then we can derive

$$\begin{aligned} B(t) &= \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\leq \Delta(j_{ser,t}, k^*) + \Delta(k^*, i_{ser,t}) + 2\alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\leq \Delta(j_{ser,t}, k^*) + 3\alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &= -\Delta(k^*, j_{ser,t}) + 3\alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \end{aligned} \tag{62}$$

where the second inequality holds is owing to

$$\begin{aligned} \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) &\geq \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\geq \hat{\Delta}_{ser,t}(k^*, i_{ser,t}) + \alpha^L_{ser,t}(i_{ser,t}, k^*) \\ &\geq \Delta(k^*, i_{ser,t}). \end{aligned}$$

We also can show

$$\begin{aligned} B(t) &= \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\leq \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\leq -\Delta(k^*, i_{ser,t}) + \hat{\Delta}_{ser,t}(k^*, i_{ser,t}) + \alpha^L_{ser,t}(k^*, i_{ser,t}) + \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\leq -\Delta(k^*, i_{ser,t}) + \hat{\Delta}_{ser,t}(j_{ser,t}, i_{ser,t}) + \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) + \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}) \\ &\leq -\Delta(k^*, i_{ser,t}) + 2\alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}). \end{aligned} \tag{63}$$

The second inequality is due to the definition of the even $\mathcal{I}$ and the third inequality is due to the definition of $j_{ser,t}$.

Combine (62) and (63), it yields

$$B(t) \leq \min\Big(0, -\Delta(k^*, k_{ser,t}) + 2\alpha^L_{ser,t}(i_{ser,t}, j_{ser,t})\Big) + \alpha^L_{ser,t}(i_{ser,t}, j_{ser,t}). \tag{64}$$

Combine the (61) and (64), then we can finish the proof of Lemma 11. □

Proof of Lemma 8. The difference between this proof and Lemma 3 is Algorithm 2 employs a different arm selection strategy. We define at round $t_k \neq \tau$, an agent m communicates with the server and $k_{ser,t_k} = k$, where

$$k_{ser,t_k} = \arg\min_{k \in \mathcal{A}} \frac{T_{ser,t_k}(k)}{p_k^*(\mathbf{y}(i_{ser,t}, j_{ser,t}))}. \tag{65}$$

And from round $t \in [t_k + 1, \tau]$, when any agent communicating with the server, $k_{ser,t} \neq k$. This implies

$$\begin{aligned} T_{ser,\tau}(k) \leq & T_{m,t_k+1}(k) + (\gamma_2 M) \sum_{s=1}^{K} T_{ser,\tau}(s) \\ = & T_{ser,t_k}(k) + (\gamma_2 M) \sum_{s=1}^{K} T_{ser,\tau}(s). \end{aligned}$$

The inequality holds due to for $t \in [t_k + 1, \tau]$, $\forall m \in \mathcal{M}$ would upload $T_{m,t}^{loc}(k) > 0$ to the server at most one time and $T_{m,t}^{loc}(k) \leq \gamma_2 \sum_{s=1}^{K} T_{ser,\tau}(s)$ according to the Lemma 9.

With Lemma 11, we can derive

$$\begin{aligned} \epsilon \leq & B(t_k) \\ \leq & \min\Big(0, -\max(\Delta(k^*, i_{ser,t_k}), \Delta(k^*, j_{ser,t_k})) + 2\alpha_{ser,t_k}^{L}(i_{ser,t_k}, j_{ser,t_k})\Big) + \alpha_{ser,t_k}^{L}(i_{ser,t_k}, j_{ser,t_k}). \end{aligned} \tag{66}$$

We would further bound $T_{ser,t}(k)$. Recalling the arm selection strategy of Algorithm 2, when k is chosen by agent m in round $t_k + 1$, this implies

$$T_{ser,t_k}(i_{ser,t_k}, j_{ser,t_k}) = T_{ser,t_k}(k) / p_k^*(i_{ser,t_k}, j_{ser,t_k}). \tag{67}$$

Recalling the definition of $\alpha_{ser,t_k}(i_{ser,t_k}, j_{ser,t_k})$ and substituting (54) and (66) into (67), we can derive

$$\begin{aligned} T_{ser,t_k}(k) \leq & \frac{\rho(\mathbf{y}(i_{ser,t_k}, j_{ser,t_k})) p_k^*(\mathbf{y}(i_{ser,t_k}, j_{ser,t_k}))}{\max\big(\frac{\Delta(k^*, i_{ser,t_k}) + \epsilon}{3}, \frac{\Delta(k^*, j_{ser,t_k}) + \epsilon}{3}, \epsilon\big)^2} C_{ser,t_k}^2 \\ \leq & \max_{i,j \in \mathcal{A}} \frac{\rho(\mathbf{y}(i,j)) p_k^*(\mathbf{y}(i,j))}{\max\big(\frac{\Delta(k^*, i) + \epsilon}{3}, \frac{\Delta(k^*, j) + \epsilon}{3}, \epsilon\big)^2} C_{ser,\tau}^2. \end{aligned}$$

We can finally bound $T_{ser,\tau}(k)$, i.e.

$$T_{ser,\tau}(k) \leq \max_{i,j \in \mathcal{A}} \frac{\rho(\mathbf{y}(i,j)) p_k^*(\mathbf{y}(i,j))}{\max\big(\frac{\Delta(k^*, i) + \epsilon}{3}, \frac{\Delta(k^*, j) + \epsilon}{3}, \epsilon\big)^2} C_{ser,\tau}^2 + \gamma_2 M \sum_{s=1}^{K} T_{ser,\tau}(s).$$

Here we finish the proof of Lemma 8. □

D AUXILIARY LEMMAS

Lemma 12 (Conditionally σ-sub-Gaussian noise [Abbasi-Yadkori et al., 2011, Lattimore and Szepesvári, 2020, Li and Wang, 2022a, He et al., 2022])**.** *The noise $\eta_{m_t,t}$ of the linear case is drawn from a conditionally σ-sub-Gaussian distribution, which satisfies*

$$\mathbb{E}\Big[e^{\lambda \eta_{m_t,t}} \Big| \mathbf{x}_{m_1,1}, ..., \mathbf{x}_{m_t,t}, m_1, ..., m_t, \eta_{m_1,1}, ..., \eta_{m_t,t}\Big] \leq e^{\sigma^2 \lambda^2 / 2}, \quad \forall \lambda \in \mathcal{R}. \tag{68}$$

Lemma 13 (Hoffeding inequality)**.** *Suppose $X_1, X_2, ..., X_n$ are i.i.d drawn from a σ-sub-Gaussian distribution and $\bar{X} = (1/n) \sum_{s=1}^{n} X_s$ represents the mean, then*

$$\mathcal{P}(|\mathbb{E}[X] - \bar{X}| \geq -a) \leq e^{-a^2 n / 2\sigma^2}.$$

Lemma 14 (Lemma 10 of Abbasi-Yadkori et al. [2011]). *The matrix norm can be bounded by*

$$det\Big(\lambda\mathbf{I} + \sum_{s=1}^{t} \mathbf{x}_{m_s,s}\mathbf{x}_{m_s,s}^{\top}\Big) \leq \Big(\lambda + \frac{t}{d}\Big)^{d}.$$

Lemma 15 (Lemma 2.3 of Tie et al. [2011]). *For arbitrary positive definitive matrices A, B and C, it has*

$$det(A + B + C)det(A) \leq det(A + B)det(A + C).$$

Lemma 16 (Lemma 2.2 of Tie et al. [2011]). *For arbitrary positive definitive matrices A, B and C, it has*

$$det(A + B + C) + det(A) \geq det(A + B) + det(A + C).$$

Lemma 17 (Lemma 12 of Abbasi-Yadkori et al. [2011]). *For arbitrary positive definitive matrices A and B satisfies $A \succ B$, it has*

$$\frac{\|\mathbf{x}\|_A^2}{\|\mathbf{x}\|_B^2} \leq \frac{det(A)}{det(B)}.$$

Lemma 18 (Theorem 1 of Abbasi-Yadkori et al. [2011]). *For $t \in |t|$, it has*

$$\Big\|\sum_{s=1}^{t} \mathbf{x}_{m_s,s}\eta_{m_s,s}\Big\|_{\lambda\mathbf{I}+\sum_{s=1}^{t}\mathbf{x}_{m_s,s}\mathbf{x}_{m_s,s}^{\top}} \leq \sigma\sqrt{d\log\Big(\frac{1+t/\lambda}{\delta}\Big)}$$

holds with probability at least $1 - \delta$.