A NEAR-OPTIMAL BEST-OF-BOTH-WORLDS ALGORITHM FOR FEDERATED BANDITS

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper studies federated multi-armed bandit (MAB) problems where multiple agents working together to solve a common MAB problem through a communication network. We focus on the heterogeneous setting in which no single agent can identify the global best arm using only local biased observations. In this setting, different agents may select the same arm at the same time step but receive varying rewards. We propose a novel algorithm called FEDFTRL for this problem, which is the first work to achieve near-optimal regret guarantees in both stochastic and adversarial environments. Notably, in the adversarial regime, our algorithm achieves $O(\sqrt{T})$ regret which is a significant improvement over the state-of-the-art regret of $O(T^{\frac{2}{3}})$ (Yi & Vojnovic, 2023). We also provide numerical evaluations comparing our algorithm with baseline methods, demonstrating the effectiveness of our approach on both synthetic and real-world datasets.

1 INTRODUCTION

The multi-armed bandit (MAB) problem is one of the most fundamental settings in online learning. Motivated by the emerging paradigm of federated learning where multiple heterogeneous agents collaboratively train a model without sharing their raw data (Kairouz et al., 2021), many recent studies have explored MAB problems in federated environments. In the federated bandit problem, the goal of all agents is to identify a globally optimal arm, while each agent can only observe locally biased rewards without disclosing any of the raw data from other agents.

Federated bandits arise in many real-world scenarios where each agent's sequence of arm pulls and outcomes remains local. For example, in a personalized online education system, optimizing a student's performance (i.e., rewards) often requires tailoring instructional methods (i.e., arms) to the student's individual characteristics (Cai et al., 2021). Given that educational software often operates locally on students' devices, it is essential for the central educational platform to personalize learning experiences effectively while maintaining strict privacy constraints. Specifically, the platform should adapt teaching strategies based on each student's unique context without directly accessing sensitive personal attributes or performance data.

In the federated bandit problem, there are V agents, each selecting one of K arms in each round. Each agent observes a heterogeneous (i.e., locally biased) reward for the chosen arm and communicates solely with its neighbors. The goal of each agent is to identify the global best arm and maximize the cumulative group reward while refraining from exchanging reward observations with other agents. Prior works on federated bandits mainly focus on two settings (i) *stochastic* settings (Dubey & Pentland, 2020; Zhu et al., 2021; Huang et al., 2021; Shi et al., 2021; Réda et al., 2022), where rewards are drawn from some underlying distributions, and (ii) *adversarial* settings (Yi & Vojnovic, 2023), where rewards are arbitrarily chosen by an adversary. However, in practice, environments are seldom purely stochastic or fully adversarial, and the precise nature of these environments is often unknown. Despite this, the existing literature on federated bandits continues to adhere to the traditional distinction between stochastic and adversarial settings.

In this paper, we study the so-called *best-of-both-worlds (BOBW)* algorithms for federated bandits, which means that our methods can achieve near-optimal regrets in both stochastic and adversarial regimes. We propose a variant of the Follow-The-Regularized-Leader (FTRL) framework for federated bandits, which incorporates a novel communication scheme. Since each agent can only com-

Settings	Algorithms	Individual regret
Stochastic	Gossip_UCB (Zhu et al., 2021)	$O\big(\sum_{k\neq k^*}\frac{V\log(T)}{\Delta_k}\big)$
	DRRB-bandit (Zhang et al., 2025)	$O\big(\sum_{k\neq k^*}\frac{\log(T)}{V\Delta_k}\big)$
	FEDFTRL (Ours)	$O\big(\sum_{k\neq k^*}\frac{\log(T)}{V\Delta_k}\big)$
	Lower bound (Zhu et al., 2021)	$\Omega\big(\sum_{k\neq k^*}\frac{\log(T)}{V\Delta_k}\big)$
Adversarial	FEDEXP3 (Yi & Vojnovic, 2023)	$O\big(\sqrt{C_T^P\log(K)}K^{\frac{2}{3}}T^{\frac{2}{3}}\big)$
	FEDFTRL (Ours)	$O\big(\sqrt{\frac{KT}{V}}+\sqrt{C_T^P\log(K)T}\big)$
	Lower bound (Yi & Vojnovic, 2023)	$\Omega\big(\max\big\{\sqrt{\frac{KT}{V}},\sqrt[4]{\frac{1+d_{\max}}{\lambda_{V-1}(M)}}\sqrt{\log(K)T}\big\}\big)$

Table 1: Overview of state-of-the-art regret bounds for federated bandits. P denotes a doubly stochastic matrix representing the communication pattern over the network G and $\sigma_2(P)$ is its second-largest singular value. $C_T^P=\frac{\min\{\log(VT),\sqrt{V}\}}{1-\sigma_2(P)}+2+D$ captures the dependence on the network topology, where D is the diameter of G. M denotes the Laplacian matrix of G, $\lambda_{V-1}(M)$ is its second-smallest eigenvalue, and $d_{\max}$ is the maximum degree among all nodes in G.

municate with its neighbors in each round, information from agents beyond the immediate neighborhood will only be received after multiple rounds. We regard the resulting latency as a form of feedback delay. Based on this idea, we develop our algorithm by adopting a hybrid regularizer (Zimmert & Seldin, 2020; Masoudian et al., 2022) for bandits with delay feedback, while introducing a novel learning rate. Additionally, to address the heterogeneous feedback, we introduce a novel truncated loss estimator that ensures the action probabilities of each agent remain nearly aligned, while keeping the aggregate loss estimate at each time step closer to the average loss.

Another technical contribution of this work is a novel analysis of individual regret. Unlike other studies on multi-agent bandits that directly analyze the individual regret of each agent, we first establish an upper bound for the group regret. Given that the action probabilities of each agent are nearly aligned, we can approximately divide the group regret by the number of agents V to derive the individual regret for each agent. This approach enables us to achieve near-optimal regret bounds in both stochastic and adversarial settings.

To keep the presentation simple, we assume that there exists a unique best arm k^*. Our method can be generalized to the environments with multiple best arms by leveraging the techniques in Ito (2021b). The regret bounds of our method along with comparisons to recent works are presented in Table 1. Our contributions are summarized as follows.

- We provide an anytime near-optimal federated bandit algorithm, called FEDFTRL, which achieves an $O\big(\sum_{k\neq k^*}(\frac{\log(T)}{V\Delta_k}+\frac{C_T^P}{\Delta_k\log(K)})\big)$ individual regret bound in the stochastic regime and simultaneously achieves an $O(\sqrt{KT/V}+\sqrt{C_T^PT\log(K)})$ individual regret bound in the adversarial regime. Here C_T^P defined in eq. (2) captures the topology of the communication graph. Our FEDFTRL algorithm is the first method to achieve BOBW regret guarantee, and the individual regret bound of our method matches the lower bound up to small polynomial gaps.

- In the adversarial regime, existing works (Yi & Vojnovic, 2023) only achieve a regret bound of $O(T^{2/3})$. In contrast, our method achieves a significantly tighter regret bound of $O(T^{1/2})$.

- We conduct experiments on both synthetic and real-world datasets to validate the effectiveness of our method. The empirical results show that our algorithm significantly outperforms prior approaches.

2 Related Work

Federated bandits. In the stochastic setting, Dubey & Pentland (2020) and Huang et al. (2021) first considered linear contextual federated bandits and extended the classical LinUCB algorithm (Li et al., 2010) to the federated environment. Shi et al. (2021) formally defined the federated bandit problems and proposed an optimal algorithm for a centralized communication network. Zhu et al. (2021) were the first to study federated bandits under a decentralized system, applying efficient gossip-based communication to achieve a near-optimal regret bound. Recently, Zhang et al. (2025) proposed a fully distributed online consensus estimation approach and integrated it into a distributed successive elimination bandit algorithm to achieve an optimal regret. In the adversarial setting, Yi & Vojnovic (2023) were the first formalize federated bandits without stochastic assumptions on the losses, called doubly adversarial bandit problems. They also proposed a federated bandit algorithm FedExp3 for such setting, which achieves a sub-linear regret of order $O(T^{2/3})$.

Best-of-Both-Worlds. For a long time, stochastic and adversarial environments have been studied independently. However, in practice, the nature of the environment is often unknown or may vary over time. This has motivated increasing interest in algorithms that perform well simultaneously in both stochastic and adversarial settings, a paradigm commonly referred to as BOBW (Bubeck & Slivkins, 2012; Auer & Chiang, 2016; Seldin & Lugosi, 2017; Wei & Luo, 2018). Zimmert & Seldin (2021) applied a Tsallis-INF regularizer within the FTRL framework to achieve BoBW guarantees not only for purely stochastic and adversarial regimes but also for a continuum of intermediate regimes. Leveraging FTRL's flexibility and strong theoretical properties, subsequent work has extended BOBW results to more complex settings, including combinatorial bandits (Zimmert et al., 2019; Ito, 2021a; Tsuchiya et al., 2023b), linear bandits (Lee et al., 2021; Dann et al., 2023), graph bandits (Rouyer et al., 2022; Ito et al., 2022), partial monitoring (Tsuchiya et al., 2023a), and delayed feedback (Masoudian et al., 2022). Among these, FTRL variants addressing delayed feedback are particularly relevant to our work, as federated bandits inherently involve implicit delays due to decentralized communication. Our algorithm builds on this line of research, adapting the FTRL paradigm to accommodate both heterogeneous rewards and decentralized communication while preserving a best-of-both-worlds guarantee.

3 Preliminaries

Let $[V] = \{1, 2, \ldots, V\}$ be the set of V agents and $[K] = \{1, 2, \ldots, K\}$ be the set of K arms. The network of V agents is represented by a simple undirected connected graph $G = ([V], E)$, where E is the set of edges. The diameter D is the maximum shortest-path distance between any pair of nodes in G.

We consider a heterogeneous multi-agent system in which all agents collaboratively solve a common K-armed bandit problem over a horizon of T round. At each time step $t \in [T]$, each agent v selects an arm $k_{v,t}$ according to its own strategy, then observes a local biased feedback $\ell_{v,t}(k_{v,t}) \in [0, 1]$. The *average loss* is defined as the average of the losses of arm k across all agents:

$$\bar{\ell}_t(k) = \frac{1}{V} \sum_{v=1}^{V} \ell_{v,t}(k).$$

At the end of each time step t, each agent v can exchange information with its neighbors $\mathcal{N}(v) = \{u \in [V] : (v, u) \in E\}$. The received information can be used in the next round if desired. The communication process is characterized by a communication matrix $P \in [0, 1]^{V \times V}$ where $P_{u,v} = 0$ only holds for $(u, v) \notin E$. We assume P is doubly stochastic, i.e., it satisfies:

$$\sum_{u \in [V]} P_{u,v} = \sum_{v \in [V]} P_{u,v} = 1, \quad P_{u,v} \geq 0.$$

We consider both adversarial and stochastic regimes with heterogeneous feedback across agents. In the adversarial regime, for each round t and agent v, the losses $\{\ell_{v,t}(k)\}_{k \in [K]}$ are arbitrarily chosen by an adversary before the game starts and may differ across agents even for the same arm. In the stochastic regime, for each agent–arm pair (v, k), the sequence $\{\ell_{v,t}(k)\}_{t=1}^{T}$ is drawn i.i.d. over time from an unknown fixed distribution with different means $\mu_{v,k}$. The performance of each agent v is evaluated by its individual pseudo-regret:

$$R_T(v) = \mathbb{E}\left[\sum_{t=1}^{T} \bar{\ell}_t(k_{v,t})\right] - \min_{k\in[K]} \mathbb{E}\left[\sum_{t=1}^{T} \bar{\ell}_t(k)\right].$$

We define the globally optimal arm in hindsight as $k^* \in \arg\min_{k\in[K]} \mathbb{E}\Big[\sum_{t=1}^{T} \bar{\ell}_t(k)\Big]$,

Notations. We denote the n-simplex by $\Delta^{n-1} = \{x \in \mathbb{R}_+^n \mid \|x\|_1 = 1\}$. For a convex function F, let F^* denote its convex conjugate (Fenchel conjugate) and $\overline{F}^*$ its conjugate constrained to the simplex. That is,

$$F^*(y) = \max_{x\in\mathbb{R}^K}\{\langle x,\, y\rangle - F(x)\}, \quad \overline{F}^*(y) = \max_{x\in\Delta^{K-1}}\{\langle x,\, y\rangle - F(x)\}.$$

We denote by $d_v = |\mathcal{N}(v)|$ the degree of node v, and by $d_{\max} = \max_{v\in[V]} d_v$ the maximum node degree in the graph. For a matrix B, we use $\sigma_i(B)$ to denote its i-th largest singular value. For a real symmetric matrix B, we use $\lambda_i(B)$ to denote its i-th largest eigenvalue. The dynamics of consensus averaging among agents is typically characterized by the Laplacian matrix M of the communication graph G, defined as:

$$M_{u,v} = \begin{cases} d_u & \text{if } u = v, \\ -1 & \text{if } u \neq v \text{ and } (u,v) \in E, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

4 Algorithm

In this section, we propose our FedFTRL method for the federated bandit problem. The details of the algorithm are presented in Algorithm 1. One challenge in federated bandits is that messages from other agents arrive with a delay that varies based on the network's connectivity. This scenario can be regard as a bandit problem with delayed feedback. Motivated by this idea, our FedFTRL algorithm adapts the FTRL framework using a hybrid regularizer similar to those in prior works on bandits with delay (Zimmert & Seldin, 2020; Masoudian et al., 2022). We present the regularizer used in FedFTRL as follows:

$$F_t(x) = -2\eta_t^{-1}\left(\sum_{k=1}^{K}\sqrt{x_k}\right) + \gamma_t^{-1}\left(\sum_{k=1}^{K} x_k(\log x_k - 1)\right).$$

We introduce a time-varying parameter C_t^P to quantify the delay caused by decentralized communication:

$$C_t^P = \frac{\min\{\log(Vt), \sqrt{V}\}}{1-\sigma_2(P)} + 2 + D, \tag{2}$$

Then we set the learning rates η and γ as

$$\eta_t^{-1} = 4\sqrt{Vt + 169V^2D} \quad \textit{and} \quad \gamma_t^{-1} = 8V\sqrt{C_t^P t/\log(K) + 36D^2(K-1)^{\frac{2}{3}} + 4(C_t^P)^2}.$$

At each time step t, each agent v computes a probability distribution for selecting arms as follows:

$$x_{v,t} = \nabla \bar{F}_t^*(-\hat{L}_{v,t}^{obs}) = \arg\min_{x\in\Delta^{K-1}}\{\langle x, \hat{L}_{v,t}^{obs} + F_t(x)\},\tag{3}$$

where P is the communication matrix, and $\hat{L}_{v,t}^{obs} \in \mathbb{R}^K$ is agent v's cumulative loss estimator up to time t. The agent then samples an arm $k_{v,t} \sim x_{v,t}$ and observes a local biased loss $\ell_{v,t}(k_{v,t})$. We construct unbiased and truncated loss estimators for this feedback as follows:

$$\hat{\ell}_{v,t}(k) = \frac{\ell_{v,t}(k)\mathbb{I}(k = k_{v,t})}{x_{v,t}(k)} \quad \text{and} \quad \tilde{\ell}_{v,t}(k) = \frac{\ell_{v,t}(k_{v,t})\mathbb{I}(k = k_{v,t})}{\max\{x_{v,t}(k), 12VC_t^P\gamma_t\}}.\tag{4}$$

Before communicating with neighbors at time t, agent v prepares a message consisting of (i) its current cumulative loss estimator $\hat{L}_{v,t}^{obs}$ and (ii) a deviation record set A_v. A new record $\langle v, t, k_{v,t}, w_{v,t}\rangle$ is appended to A_v if and only if $\hat{\ell}_{v,t} \neq \tilde{\ell}_{v,t}$, in which case we set the estimator's deviation $w_{v,t} = V\big(\hat{\ell}_{v,t}(k_{v,t}) - \tilde{\ell}_{v,t}(k_{v,t})\big)$. Next, agent v averages its cumulative loss estimates with those of its neighbors and merges incoming deviation records.

Algorithm 1 FEDFTRL (local routine for each agent v)

```
Input: a doubly stochastic matrix P ∈ [0,1]^{V×V}; the diameter D of graph G.
Initialize: a deviation record set A_v ← ∅; the loss estimate \hat{L}^{obs}_{v,1} = 0_K.
for t = 1, 2, 3, ... do
    Compute x_{v,t} = arg min_{x∈Δ^{K-1}} {⟨x, \hat{L}^{obs}_{v,t}⟩ + F_t(x)}.
    Sample k_{v,t} ~ x_{v,t} and observe ℓ_{v,t}(k_{v,t}).
    Construct \hat{ℓ}_{v,t} and \tilde{ℓ}_{v,t} by Eq. (4).
    if \hat{ℓ}_{v,t} ≠ \tilde{ℓ}_{v,t} then
        Set w_{v,t} = V(\hat{ℓ}_{v,t}(k_{v,t}) − \tilde{ℓ}_{v,t}(k_{v,t})) and append the record ⟨v, t, k_{v,t}, w_{v,t}⟩ to A_v.
    end if
    Send the message {\hat{L}^{obs}_{v,t}, A_v} to neighbors of agent v.
    Update cumulative loss estimate:
        \hat{L}^{obs}_{v,t+1} = Σ_{u: (u,v)∈E} P_{u,v} \hat{L}^{obs}_{u,t} + V \tilde{ℓ}_{v,t}.      (5)
    Update the deviation record set via A_v ← ∪_{(u,v)∈E} A_u.
    for each record ⟨u, s, k, w_{u,s}⟩ ∈ A_v do
        if t − s > D then
            Set \hat{L}^{obs}_{v,t+1}(k) ← \hat{L}^{obs}_{v,t+1}(k) + w_{u,s}, and remove the record ⟨u, s, k, w_{u,s}⟩ from A_v.
        end if
    end for
end for
```

4.1 INTUITION BEHIND THE TRUNCATED LOSS ESTIMATOR

One challenge in federated bandits is that the loss observed locally at agent v is biased relative to the average loss $\bar{\ell}_t$ of that arm. In FEDFTRL, we address this by updating $\hat{L}^{obs}_{v,t+1}$ using the *truncated* estimator $\tilde{\ell}_{v,t}$ instead of the unbiased $\hat{\ell}_{v,t}$. This choice keeps all agents' action probability distributions roughly aligned.

Specifically, when constructing $\tilde{\ell}_{v,t}$ we cap the denominator by $\max\{x_{v,t}(k), 12VC^P_t\gamma_t\}$, which prevents the loss estimate from exploding when $x_{v,t}(k)$ is extremely small. As a result, no single rare arm pull can trigger an excessively large update that would cause the agents' probability distributions to diverge. This stabilization ensures well-behaved and nearly aligned action distributions across agents. Indeed, we have $\mathbb{E}\big[\sum_{v\in[V]}\hat{\ell}_{v,t}\big] = V\,\bar{\ell}_t$, and $\sum_{v\in[V]}\tilde{\ell}_{v,t}$ closely tracks this same quantity except on rounds where truncation occurs, enabling more accurate estimation of the global loss.

4.2 INTUITION BEHIND THE COMMUNICATION

Since broadcast raw observations is not allowed in the federated learning, our method communicates the deviation record vector A_v in each round. Such communication is necessitated for reduce the deviation caused by the use of the truncated estimator. Specifically, while truncation keeps the probability distributions of agents nearly aligned, whenever $\hat{\ell}_{v,t} \neq \tilde{\ell}_{v,t}$, the local loss estimates deviate from the average loss. Consequently, whenever truncation occurs (i.e., $\hat{\ell}_{v,t} \neq \tilde{\ell}_{v,t}$), agent v appends the record $\langle v, t, k_{v,t}, w_{v,t}\rangle$ to A_v. Once a record $\langle u, s, k, w_{u,s}\rangle$ has been in the system for more than D rounds (i.e., $t - s > D$), every agent will have received it. At that point, adding the correction $w_{u,s}$ to $\hat{L}^{obs}_{v,t+1}(k)$ will no longer introduce any distribution mismatch among agents.

Finally, recall that we multiply cumulative loss by V in Equation 5. Communication averaging yields consensus on average losses, and this factor of V counteracts the averaging effect, ensuring that feedback information is not overly diluted.

Thus we can finally obtain the following regret guarantee for FEDFTRL, with the proof is provided in Appendix 11.

Theorem 1. *If* FEDFTRL *is run with a given doubly stochastic communication matrix P, then in the adversarial regime, the individual regret of each agent v is upper bounded as*

$$R_T(v) \leq 13\sqrt{KT/V} + 13\sqrt{C_T^P T \log(K)} + 156\sqrt{D} + 72D(K-1)^{\frac{1}{3}}\log(K) + 24C_T^P \log(K).$$

Furthermore, in the stochastic regime the individual regret of each agent v is bounded as

$$R_T(v) \leq \sum_{k \neq k^*} \frac{90\log(T)}{V\Delta_k} + \sum_{k \neq k^*} \frac{180C_T^P}{\Delta_k \log(K)} + 33\sqrt{D} + 15D(K-1)^{\frac{1}{3}}\log(K) + 11C_T^P \log(K).$$

For each agent v, the expected communication cost in each round is $O(K)$.

Remark 1. *If the doubly stochastic matrix P is constructed via the* max-degree *trick (Duchi et al., 2011), i.e.,*

$$P = I - \frac{W - A}{1 + d_{\max}},$$

where $W = diag(d_1, d_2, \ldots, d_V)$ is the degree matrix and A is the adjacency matrix of the communication graph G, then Corollary 1 of Duchi et al. (2011) implies the following result:

$$C_T^P = \Omega\left(\sqrt{\frac{1 + d_{\max}}{\lambda_{V-1}(M)}}\sqrt{\min\{\log(VT),\ \sqrt{V}\}}\right).$$

This result shows that only small polynomial gaps remains between our upper bound and lower bound $\Omega\big(\max\big\{\sqrt{\frac{KT}{V}}, \sqrt[4]{\frac{1+d_{\max}}{\lambda_{V-1}(M)}}\sqrt{\log(K)T}\big\}\big)$ in the adversarial setting.

5 A Sketch of the Proof of Theorem 1

In this section we provide a sketch of the proof of Theorem 1. We provide a proof sketch for the regret bound of adversarial and stochastic settings in Section 5.1 and Section 5.2, respectively. The detail proofs are provided in Appendix 11.

5.1 Adversarial Bound

We start by providing a key lemma (Lemma 1) that controls the ratio of the playing distribution between any two agents at the same time step, with the proof is provided in Section 10 in the appendix. This lemma also relates each agent's individual regret to the group regret, which represents that the sum of regrets over all agents.

Lemma 1. *For any two agents u and v, and for any action k at time t, it holds that*

$$x_{u,t}(k) \leq \frac{3}{2}x_{v,t}(k) \quad \text{and} \quad x_{v,t}(k) \leq \frac{3}{2}x_{u,t}(k).$$

Furthermore, for any agent v, its individual regret is bounded in terms of the group regret as

$$R_T(v) \leq \frac{3}{2V}\sum_{u=1}^{V} R_T(u).$$

To bound the group regret, we transform the federated bandit into a single-agent interaction with the environment over VT rounds. This reduction significantly simplifies the theoretical analysis. We introduce some additional definitions: define the instantaneous loss m_t and the drifted cumulative loss $\hat{L}_{v,t}$ as follows:

$$m_t = \frac{1}{V}\sum_{v=1}^{V}\hat{\ell}_{v,t} \quad \text{and} \quad \hat{L}_{v,t} = \sum_{s=1}^{t-1} Vm_s + (v-1)m_t.$$

Since $\mathbb{E}[m_t] = \bar{\ell}_t$, intuitively $\hat{L}_{v,t}$ staggers the cumulative loss by an offset proportional to $(v-1)$. This ensures that when we sum over all agents, the losses m_t line up as if they were incurred sequentially by a single agent over VT rounds. As a result, we can decompose the group regret into three terms:

$$\begin{aligned}\sum_{v=1}^{V} R_T(v) &= \sum_{v=1}^{V}\mathbb{E}\Big[\sum_{t=1}^{T}\langle \bar{\ell}_t, x_{v,t}\rangle - \bar{\ell}_t(k^*)\Big] \\ &\le \mathbb{E}\Bigg[\underbrace{\sum_{t=1}^{T}\sum_{v=1}^{V}\Big(\bar{F}_t^*(-\hat{L}_{v,t}^{obs}-m_t) - \bar{F}_t^*(-\hat{L}_{v,t}^{obs}) + \langle x_{v,t}, m_t\rangle\Big)}_{(A)} \\ &+ \underbrace{\sum_{t=1}^{T}\sum_{v=1}^{V}\Big(\bar{F}_t^*(-\hat{L}_{v,t}^{obs}) - \bar{F}_t^*(-\hat{L}_{v,t}^{obs}-m_t) - \bar{F}_t^*(-\hat{L}_{v,t}) + \bar{F}_t^*(-\hat{L}_{v+1,t})\Big)}_{(B)} \\ &+ \underbrace{\left(\sum_{v=1}^{V}\sum_{t=1}^{T}\bar{F}_t^*(-\hat{L}_{v,t}) - \bar{F}_t^*(-\hat{L}_{v+1,t})\right) - \hat{L}_{1,T+1}(k^*)\Bigg]}_{(C)}.\end{aligned}$$

Term (A) is a typical Bregman divergence term arising from the FTRL/OMD analysis, and it depends on the local norm of the regularizer. We can bound it as

$$\mathbb{E}[(A)] \le \frac{9}{32}\sum_{v=1}^{V}\sum_{t=1}^{T}\sqrt{\frac{K}{Vt}} \le \frac{9}{16}\sqrt{VKT}.$$

Term (B) is handled by the analysis in Zimmert & Seldin (2020), which yields

$$\mathbb{E}[(B)] \le \frac{9}{32}\sum_{v=1}^{V}\sum_{t=1}^{T}\sqrt{\frac{C_t^P\log(K)}{t}} \le \frac{9}{16}V\sqrt{C_T^P T\log(K)}.$$

Term (C) can be bounded using standard telescoping-sum techniques. Specifically, one obtains

$$\mathbb{E}[(C)] \le 8\sqrt{VKT} + 8V\sqrt{C_T^P T\log(K)} + 104V\sqrt{D} + 48VD(K-1)^{\frac{1}{3}}\log(K) + 24VC_T^P\log(K).$$

Combining the bounds for (A)–(C) above and simplifying, we complete the proof of the adversarial bound.

5.2 Stochastic Bound

Inspired the analysis of stochastic bound for bandit with delay feedback in Masoudian et al. (2022), let $\tilde{x}_{v,t} = \nabla\bar{F}_t^*(-\hat{L}_{v,t})$, then we define the drifted pseudo-regret as

$$R_T^{drift}(v) = \mathbb{E}\Big[\sum_{t=1}^{T}\langle\tilde{x}_{v,t}, \bar{\ell}_t\rangle - \bar{\ell}_t(k^*)\Big].$$

We can use the drifted pseudo-regret to control the actual pseudo-regret as follows:

$$\begin{aligned}\sum_{t=1}^{T} R_T(v) &\le \frac{5}{3}\sum_{v=1}^{V} R_T^{drift}(v) + VD \\ &\le \underbrace{\frac{5}{3}\sum_{v=1}^{V}\sum_{k\neq k^*}2\sqrt{\frac{\tilde{x}_{v,t}^k}{Vt}}}_{(A)} + \underbrace{\frac{5}{3}\sum_{t=2}^{T}\sum_{v=1}^{V}\sum_{k=1}^{K}\frac{2C_t^P\gamma_{t-1}\tilde{x}_{v,t}^k\log(1/\tilde{x}_{v,t}^k)}{\log(K)}}_{(B)} \\ &+ \underbrace{13V\sqrt{D} + 6VD(K-1)^{\frac{1}{3}}\log(K) + 7VC_T^P\log(K)}_{(C)}.\end{aligned}$$

Self Bounding Analysis: We apply a self-bounding technique to combine terms (A) and (B) with the drifted regret. Specifically:

$$\sum_{t=1}^{T} R_T(v) \leq \frac{5}{3}\sum_{v=1}^{V}(3R_T^{drift}(v) - 2R_T^{drift}(v)) + VD$$
$$\leq \frac{5}{3}\left(3A - \sum_{v=1}^{V} R_T^{drift}(v) + 3B - \sum_{v=1}^{V} R_T^{drift}(v)\right) + (C).$$

Using the analysis in Masoudian et al. (2022), we have the following bound:

$$3(A) - \sum_{v=1}^{V} R_T^{drift}(v) \leq \sum_{k\neq k^*} \frac{36\log(T)}{\Delta_k}, \quad 3(B) - \sum_{v=1}^{V} R_T^{drift}(v) \leq \frac{72VC_T^P}{\Delta_k \log(K)}.$$

Combining these bounds with (C) and simplifying yields the stated stochastic regret bound.

6 Experiments

We conducted experiments on both synthetic and real-world datasets under various network topologies to evaluate the performance of our FedFTRL algorithm against several baseline methods. We consider the following baseline methods: FedExp3 (Yi & Vojnovic, 2023), Gossip_UCB (Zhu et al., 2021), DRBB-bandit (Zhang et al., 2025) and IND-FTRL, where IND-FTRL represents that each agent runs the Tsallis FTRL (Zimmert & Seldin, 2021) without any communication. Following the experimental design in Yi & Vojnovic (2023), we adopt the max-degree trick to construct the doubly stochastic matrix P, which is presented in Remark 1. We set the learning rate of our FedFTRL algorithm as $\eta_t^{-1} = 0.5\sqrt{Vt}$ and $\gamma_t^{-1} = 8V\sqrt{C_t^P t/\log(K) + 4}$. All experiments are repeated for 50 trials, with the average results plotted as lines.

Choice of the network graphs. We conduct experiments on several different network graphs including fully connected graph, $\sqrt{V}\times\sqrt{V}$ grid graph and random geometric graph (RGG). A random geometric graph RGG-g is constructed by uniformly placing each node in $[0,1]^2$ and connecting any two nodes whose distance is within g (Penrose, 2003). In our experiments, we choose $g = 0.5$.

6.1 Synthetic Datasets

For each agent v and each arm k, we independently sample a mean loss $\mu_{v,k}$ from the uniform distribution over $[0,1]$. When agent v pulls arm k at round t, its feedback $\ell_{v,t}(k)$ is then drawn from a Gaussian distribution with mean $\mu_{v,k}$ and variance 0.01. We set horizon $T = 3000$, number of agents $V = 16$, and number of arms $K = 20$.

The results in Figure 1 show that our FedFTRL algorithm outperforms all baselines for average regret. It is worth noting that IND-FTRL cannot achieve sublinear regret by only observing the local biased feedback, demonstrating the benefits brought by our communication mechanism.

6.2 MovieLens Dataset: Recommending Popular Movie Genres

We further evaluate our FedFTRL algorithm on a real-world dataset: the latest MovieLens dataset (Cantador et al., 2011). This dataset contains 87,585 movies classified into 20 genres, with 32,000,204 ratings (scores in $0.5, 1, \ldots, 5$) from over 280,000 users. Among these users, 3,963 have rated at least one movie in every genre; we select these users as our agents and treat each genre as an arm. Then we set $T = 3000$, $V = 3963$ and $K = 20$.

To simulate changes in user preferences over time, we sort each user's ratings in chronological order and construct the loss sequence as follows. Let $r_v^j(k)$ be the j-th rating of user v for genre k in this sorted order. The the loss for user v on arm k at time t defined as

$$\ell_{v,t}(k) = \frac{5.5 - r_v^j(k)}{5.0} \quad \text{for} \quad t \in \left[(j-1)\left\lfloor\frac{T}{n_v^k}\right\rfloor, j\left\lfloor\frac{T}{n_v^k}\right\rfloor\right),$$

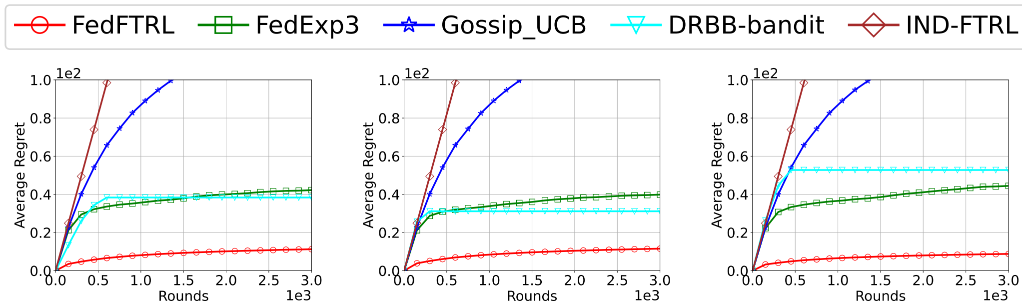

Figure 1: Average cumulative regret for FedFTRL, FEDEXP3, IND-FTRL, Gossip_UCB and DRBB-bandit in the synthetic dataest, under three different communication networks: (left) complete graph, (middle) grid graph, and (right) RGG-0.5.

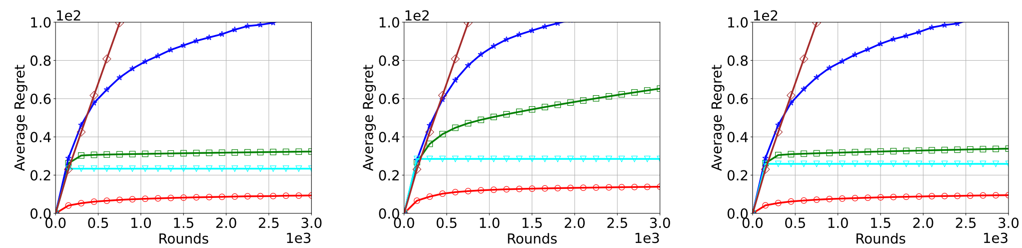

Figure 2: Average cumulative regret for FedFTRL, FEDEXP3, IND-FTRL, Gossip_UCB and DRBB-bandit in the MovieLens dataest, under three different communication networks: (left) complete graph, (middle) grid graph, and (right) RGG-0.5.

where n_v^k is the total number of ratings user v has for genre k. In words, we partition each user's interaction timeline into n_v^k segments of equal length (rounded down), and assign the j-th rating $r_v^j(k)$ as the loss (scaled to $[0, 1]$) for all time steps in the j-th segment for that user-genre pair.

As shown in Figure 2, FEDFTRL still significantly outperforms all baselines, which demonstrates the superiority of our FEDFTRL algorithm.

7 CONCLUSION

In this paper, we propose a novel federated bandit algorithm, called FEDFTRL, which, to the best of our knowledge, is the first to achieve a BOBW regret guarantee in both stochastic and adversarial settings. Our theoretical analysis shows that the regret upper bound matches the lower bound up to small polynomial factors. Furthermore, empirical results corroborate the theoretical analysis and demonstrate the superior performance of our algorithm. In addition, exploring how to close the gap between the upper and lower regret bounds in the adversarial setting is also worth investigating.

REPRODUCIBILITY STATEMENT

We provide the complete proofs for all theoretical claims in the appendices. We include the source code in the supplementary material for the reproducibility. We use the MovieLens[1] dataset in our experiments, which is publicly available online.

REFERENCES

Jacob D Abernethy, Chansoo Lee, and Ambuj Tewari. Fighting bandits with a new kind of smoothness. *Advances in Neural Information Processing Systems*, 28, 2015.

[1] https://grouplens.org/datasets/movielens/32m/

Peter Auer and Chao-Kai Chiang. An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits. In *Conference on Learning Theory*, pp. 116–120. PMLR, 2016.

Sébastien Bubeck and Aleksandrs Slivkins. The best of both worlds: Stochastic and adversarial bandits. In *Conference on Learning Theory*, pp. 42–1. JMLR Workshop and Conference Proceedings, 2012.

William Cai, Josh Grossman, Zhiyuan Jerry Lin, Hao Sheng, Johnny Tian-Zheng Wei, Joseph Jay Williams, and Sharad Goel. Bandit algorithms to personalize educational chatbots. *Machine Learning*, 110(9):2389–2418, 2021.

Iván Cantador, Peter Brusilovsky, and Tsvi Kuflik. Second workshop on information heterogeneity and fusion in recommender systems (hetrec2011). In *Proceedings of the fifth ACM conference on Recommender systems*, pp. 387–388, 2011.

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

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

John C Duchi, Alekh Agarwal, and Martin J Wainwright. Dual averaging for distributed optimization: Convergence analysis and network scaling. *IEEE Transactions on Automatic control*, 57(3): 592–606, 2011.

Saghar Hosseini, Airlie Chapman, and Mehran Mesbahi. Online distributed optimization via dual averaging. In *52nd IEEE Conference on Decision and Control*, pp. 1484–1489. IEEE, 2013.

Ruiquan Huang, Weiqiang Wu, Jing Yang, and Cong Shen. Federated linear contextual bandits. *Advances in neural information processing systems*, 34:27057–27068, 2021.

Shinji Ito. Hybrid regret bounds for combinatorial semi-bandits and adversarial linear bandits. *Advances in Neural Information Processing Systems*, 34:2654–2667, 2021a.

Shinji Ito. Parameter-free multi-armed bandit algorithms with hybrid data-dependent regret bounds. In *Conference on Learning Theory*, pp. 2552–2583. PMLR, 2021b.

Shinji Ito, Taira Tsuchiya, and Junya Honda. Nearly optimal best-of-both-worlds algorithms for online learning with feedback graphs. *Advances in Neural Information Processing Systems*, 35: 28631–28643, 2022.

Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *Foundations and trends® in machine learning*, 14(1–2):1–210, 2021.

Chung-Wei Lee, Haipeng Luo, Chen-Yu Wei, Mengxiao Zhang, and Xiaojin Zhang. Achieving near instance-optimality and minimax-optimality in stochastic and adversarial linear bandits simultaneously. In *International Conference on Machine Learning*, pp. 6142–6151. PMLR, 2021.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pp. 661–670, 2010.

Saeed Masoudian, Julian Zimmert, and Yevgeny Seldin. A best-of-both-worlds algorithm for bandits with delayed feedback. *Advances in Neural Information Processing Systems*, 35:11752–11762, 2022.

Mathew Penrose. *Random geometric graphs*, volume 5. OUP Oxford, 2003.

Clémence Réda, Sattar Vakili, and Emilie Kaufmann. Near-optimal collaborative learning in bandits. *Advances in Neural Information Processing Systems*, 35:14183–14195, 2022.

Chloé Rouyer, Dirk van der Hoeven, Nicolò Cesa-Bianchi, and Yevgeny Seldin. A near-optimal best-of-both-worlds algorithm for online learning with feedback graphs. *Advances in Neural Information Processing Systems*, 35:35035–35048, 2022.

Yevgeny Seldin and Gábor Lugosi. An improved parametrization and analysis of the exp3++ algorithm for stochastic and adversarial bandits. In *Conference on Learning Theory*, pp. 1743–1759. PMLR, 2017.

Chengshuai Shi, Cong Shen, and Jing Yang. Federated multi-armed bandits with personalization. In *International conference on artificial intelligence and statistics*, pp. 2917–2925. PMLR, 2021.

Taira Tsuchiya, Shinji Ito, and Junya Honda. Best-of-both-worlds algorithms for partial monitoring. In *International Conference on Algorithmic Learning Theory*, pp. 1484–1515. PMLR, 2023a.

Taira Tsuchiya, Shinji Ito, and Junya Honda. Further adaptive best-of-both-worlds algorithm for combinatorial semi-bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 8117–8144. PMLR, 2023b.

Chen-Yu Wei and Haipeng Luo. More adaptive algorithms for adversarial bandits. In *Conference On Learning Theory*, pp. 1263–1291. PMLR, 2018.

Jialin Yi and Milan Vojnovic. Doubly adversarial federated bandits. In *International Conference on Machine Learning*, pp. 39951–39967. PMLR, 2023.

Haoran Zhang, Xuchuang Wang, Hao-Xu Chen, Hao Qiu, Lin Yang, and Yang Gao. Near-optimal regret bounds for federated multi-armed bandits with fully distributed communication. In *The 41st Conference on Uncertainty in Artificial Intelligence*, 2025.

Zhaowei Zhu, Jingxuan Zhu, Ji Liu, and Yang Liu. Federated bandit: A gossiping approach. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 5(1):1–29, 2021.

Julian Zimmert and Yevgeny Seldin. An optimal algorithm for adversarial bandits with arbitrary delays. In *International Conference on Artificial Intelligence and Statistics*, pp. 3285–3294. PMLR, 2020.

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

Julian Zimmert, Haipeng Luo, and Chen-Yu Wei. Beating stochastic and adversarial semi-bandits optimally and simultaneously. In *International Conference on Machine Learning*, pp. 7683–7692. PMLR, 2019.

8 NOTATIONS

Defining the instantaneous loss $m_t = \frac{1}{V}\sum_{v=1}^{V}\hat{\ell}_{v,t}$, the average cumulative loss

$$\bar{L}_{t-1} = \frac{1}{V}\sum_{v=1}^{V}\hat{L}^{obs}_{v,t-1},$$

and the drifted cumulative loss $\hat{L}_{v,t} = \sum_{s=1}^{t-1} Vm_s + (v-1)m_t$.
For ease of writing, we sometimes use the index $\{V+1, t\}$ to represent the index $\{1, t+1\}$.

9 AUXILIARY LEMMAS

First, we analyze some properties of the regularizer

$$F_t(x) = -2\eta_t^{-1}\sum_{k=1}^{K} x_k^{\frac{1}{2}} + \gamma_t^{-1}\sum_{k=1}^{K} x_k(\log(x_k) - 1)$$

.
Given the function $f_t(x) = -2\eta_t^{-1}\sqrt{x} + \gamma_t^{-1}x(\log(x) - 1)$.

Fact 1. *$f_t'(x)$ is a concave function, $f_t''(x)$ is a monotonically decreasing function, $f_t''(x)^{-1}$ is a convex function, and $f_t^{*'}$ is a convex monotonically increasing function.*

Proof. By definition $f_t'(x) = -\eta_t^{-1}x^{-\frac{1}{2}} + \gamma_t^{-1}\log(x)$, whose second derivative is $-\frac{3}{4}\eta_t^{-1}x^{-\frac{5}{2}} - \gamma_t^{-1}x^{-2} < 0$, which conclude the first and the second statement. $f_t''(x)^{-1} = (\frac{1}{2}\eta_t^{-1}x^{-\frac{3}{2}} + \gamma_t^{-1}x^{-1})^{-1}$ so the second derivative is

$$\frac{\eta_t\gamma_t^2\left(2\eta_t x^{\frac{7}{2}} + 3\gamma_t x^3\right)}{2\sqrt{x}\left(2\eta_t x^{\frac{3}{2}} + \gamma_t x^{3/2}\right)^3} > 0,$$

which conclude the third claim. Since f_t are Legendre functions, we have $f_t^{*''}(y) = f_t''(f_t^{*'}(y))^{-1} > 0$. Therefore the function is monotonically increasing. Since both $f_t''(x)^{-1}$, as well as $f_t^{*'}(y)$ are increasing, the composition is as well and $f_t^{*'''} > 0$. □

Fact 2. *For any convex F, for $L \in \mathbb{R}^K$ and $c \in \mathbb{R}$:*

$$\bar{F}^*(L + c\mathbf{1}_K) = \bar{F}^*(L) + c.$$

Proof. By definition $\bar{F}^*(L + c1_k) = \max_{x\in\Delta^{K-1}}\langle x, L + c\mathbf{1}_K\rangle - F(x) = \max_{x\in\Delta^{K-1}}\langle x, L\rangle - F(x) + c = \bar{F}^*(L) + c$. □

Fact 3. *For ant $x_{v,t}$ there exists $c \in \mathbb{R}$, such that:*

$$x_{v,t} = \nabla\bar{F}_t^*(-\hat{L}^{obs}_{v,t}) = \nabla F_t^*(-\hat{L}^{obs}_{v,t} + c\mathbf{1}_k) = \nabla F_t^*(\nabla F_t(x_{v,t})).$$

Proof. By the KKT conditions, there exists $c \in \mathbb{R}$, such that $x_{v,t} = \arg\max_{x\in\Delta^{K-1}}\langle x, -\hat{L}^{obs}_{v,t}\rangle + F_t(x)$ satisfies $\nabla F_t(x_{v,t}) = -\hat{L}^{obs}_{v,t} + c\mathbf{1}_k$. The rest follows by the standard property $\nabla F = (\nabla F^*)^{-1}$ of Legendre F. □

Fact 4. *For any Legendre function F and $L \in \mathbb{R}^K$ it holds that*

$$\bar{F}^*(L) \le F^*(L),$$

with equality if and only if there exists $x \in \Delta^{K-1}$ such that $L = \nabla F(x)$.

Proof. The first statement follows from the definition, since for any $A \subset B$:$\max_{x\in A} f(x) \le \max_{x\in B} f(x)$. The second part follows because equality means that $\arg\max_x\big(\langle x, L\rangle - F(x)\big) = \nabla F^*(L) \in \Delta^{K-1}$, which is equivalent to the statement. □

Fact 5. *For any $x \in \Delta^{K-1}$, $L \geq 0$ and $k \in [K]$, we have*

$$\big(\nabla \bar{F}_t^*\big(\nabla F_t(x) - L\big)\big)_k \geq \big(\nabla F_t^*\big(\nabla F_t(x) - L\big)\big)_k.$$

Proof. By Fact 3, there exists some $c \in \mathbb{R}$ such that $\nabla \bar{F}_t^*\big(\nabla F_t(x) - L\big) = \nabla F_t^*\big(\nabla F_t(x) - L + c\mathbf{1}_K\big)$. The statement is equivalent to c being non-negative, since $F_t^{*'}$ are monotonically increasing. If $c < 0$, then

$$\begin{aligned} 1 &= \sum_{k=1}^{K}\Big(\nabla \bar{F}_t^*\big(\nabla F_t(x) - L\big)\Big)_k = \sum_{k=1}^{K}\Big(\nabla F_t^*\big(\nabla F_t(x) - L + c\mathbf{1}_K\big)\Big)_k \\ &= \sum_{k=1}^{K} F_t^{*'}\big(F_t'(x_k) - L_k + c\big) < \sum_{k=1}^{K} F_t^{*'}\big(F_t'(x_k)\big) = 1, \end{aligned}$$

which is a contradiction. Hence c must be non-negative, and the proof is complete. □

Fact 6. *Let $D_F(x, y) = F(x) - F(y) - \langle x - y, \nabla F(y)\rangle$ be the Bregman divergence of a function F. For any Legendre function f with monotonically decreasing second derivative,$x \in \operatorname{dom}(f)$, and $\ell \geq 0$, such that $f'(x) - \ell \in \operatorname{dom}\big(f^*\big)$, we have*

$$D_{f^*}\big(f'(x) - \ell, f'(x)\big) \leq \frac{\ell^2}{2f''(x)}.$$

Proof. By Taylor's theorem, there exists some $\tilde{x} \in \big[f^{*'}(f'(x) - \ell), x\big]$ such that

$$D_{f^*}\big(f'(x) - \ell, f'(x)\big) = \frac{\ell^2}{2f''\big(\tilde{x}\big)}.$$

Note that $\tilde{x}$ is smaller than x, since $f^{*'}$ is monotonically increasing. Finally, using the fact that the second derivative is decreasing allows us to bound

$$f''\big(\tilde{x}\big)^{-1} \leq f''(x)^{-1}.$$

Hence the stated inequality follows. □

Fact 7. *For any convex function F, and $L_2 \geq L_1$ (coordinate wise), we have*

$$\bar{F}^*(-L_1) \geq \bar{F}^*(-L_2).$$

Proof.

$$\begin{aligned} \bar{F}^*(-L_2) &= \big\langle \nabla \bar{F}^*(-L_2), -L_2\big\rangle + F\big(\nabla \bar{F}^*(-L_2)\big) \\ &\leq \big\langle \nabla \bar{F}^*(-L_2), -L_1\big\rangle + F\big(\nabla \bar{F}^*(-L_2)\big) \\ &\leq \max_{x \in \Delta^{K-1}}\Big(\langle x, -L_1\rangle + F(x)\Big) \\ &= \bar{F}^*(-L_1). \end{aligned}$$

□

Lemma 2. *For any fixed k, t, given $x_1 = \nabla \bar{F}_t^*\,(-L_1)$ and $x_2 = \nabla \bar{F}_t^*\,(-L_2)$, if we have*

$$\frac{\sum_{k=1}^{K} f_t^{''}(x_1^k)^{-1}(L_2(k) - L_1(k))}{\sum_{k=1}^{K} f_t^{''}(x_1^k)^{-1}} \leq \frac{\alpha_1}{\gamma_t} \quad \text{and} \quad L_1(k) - L_2(k) \leq \frac{\alpha_2}{\gamma_t},$$

where $\alpha_1, \alpha_2 \in [0, \frac{1}{2})$. Then we can obtain

$$x_2^k \leq \frac{1}{1 - \alpha_1 - \alpha_2} x_1^k.$$

Proof. By the KKT conditions $\exists\, \mu_1, \mu_2$ s.t. $\forall\, k$:

$$f_t'(x_1^k) = -L_1(k) + \mu_1, \quad f_t'(x_2^k) = -L_2(k) + \mu_2.$$

From the concavity of $f'(x_1)$, derived from Fact 1, we have

$$\left(x_1^k - x_2^k\right) f_t''(x_1^k) \le f_t'(x_1^k) - f_t'(x_2^k) \le \left(x_1^k - x_2^k\right) f_t''(x_2^k). \tag{6}$$

Using left side of equation 6 and the fact $f_t''(x_1^k) \ge 0$ gives us

$$\begin{aligned}
&x_1^k - x_2^k \le f_t^{''}(x_1^k)^{-1}\left(\mu_1 - \mu_2 + L_2(k) - L_1(k)\right) \Rightarrow \\
&\sum_{k=1}^{K} x_1^k - x_2^k = 0 \le \sum_{k=1}^{K} f_t^{''}(x_1^k)^{-1}(\mu_1 - \mu_2 + L_2(k) - L_1(k)) \Rightarrow \\
&\mu_2 - \mu_1 \le \frac{\sum_{k=1}^{K} f_t^{''}(x_1^k)^{-1}(L_2(k) - L_1(k))}{\sum_{k=1}^{K} f_t^{''}(x_1^k)^{-1}}.
\end{aligned}$$

Using the upper bound for $f_t^{'}(x_1^k) - f_t^{'}(x_2^k)$ in equation 6 along with the upper bound for $\mu_2 - \mu_1$ and the fact that $f_t^{'}(x_1^k) - f_t^{'}(x_2^k) = \mu_1 - \mu_2 + L_2(k) - L_1(k)$ result in

$$\begin{aligned}
&(x_2^k - x_1^k) f_t^{''}(x_2^k) \le \mu_2 - \mu_1 + L_1(k) - L_2(k) \\
&\qquad\qquad \le \frac{\sum_{k=1}^{K} f_t^{''}(x_1^k)^{-1}(L_2(k) - L_1(k))}{\sum_{k=1}^{K} f_t^{''}(x_1^k)^{-1}} - (L_2(k) - L_1(k)) \Rightarrow \\
&x_2^k \le x_1^k + f_t^{''}(x_2^k)^{-1} \times \frac{\sum_{k=1}^{K} f_t^{''}(x_1^k)^{-1}(L_2(k) - L_1(k))}{\sum_{k=1}^{K} f_t^{''}(x_1^k)^{-1}} - f_t^{''}(x_2^k)^{-1}(L_2(k) - L_1(k)) \\
&x_2^k \overset{(a)}{\le} x_1^k + \gamma_t x_2^k \times \frac{\sum_{k=1}^{K} f_t^{''}(x_1^k)^{-1}(L_2(k) - L_1(k))}{\sum_{k=1}^{K} f_t^{''}(x_1^k)^{-1}} - \gamma_t x_2^k (L_2(k) - L_1(k)) \\
&x_2^k \le x_1^k + \alpha_1 x_2^k + \alpha_2 x_2^k \Rightarrow x_2^k \le \frac{1}{1 - \alpha_1 - \alpha_2} x_1^k,
\end{aligned}$$

where (a) holds because $f_t^{''}(x_2^k)^{-1} = \left(\frac{1}{2}\eta_t^{-1}(x_2^k)^{-3/2} + \gamma_t^{-1}(x_2^k)^{-1}\right)^{-1}$. □

Lemma 3. *For any time step t and agent $v \in [V]$, we have*

$$\|\bar{L}_t - \hat{L}_{v,t}^{obs}\|_\infty \le \frac{1}{12\gamma_t} \quad \textit{and} \quad \|\hat{L}_{v,t}^{obs} - \bar{L}_t\|_\infty \le \frac{1}{12\gamma_t}.$$

Proof. As mentioned before, using the deviation record to update will not affect $\|\bar{L}_t - \hat{L}_{v,t}^{obs}\|_\infty$, because all agents will perform this operation. So we only consider the impact of equation 5.

From equation 4, for any $v,\ t$, we have the following inequalities:

$$\|\tilde{\ell}_{v,t}\|_\infty = \frac{\ell_{v,t}(k_{v,t})}{\max\{x_{v,t}(k_{v,t}), 12VC_t^P\gamma_t\}} \le \frac{1}{12VC_t^P\gamma_t}.$$

Since $\{\gamma_t\}$ is non-increasing, let $L = \frac{1}{12C_t^P\gamma_t}$ in Lemma 6 in Hosseini et al. (2013), we can get

$$\|\bar{L}_t - \hat{L}_{v,t}^{obs}\|_\infty \le V\left(\frac{1}{12VC_t^P\gamma_t}\left(\frac{\sqrt{V}}{1 - \sigma_2(P)} + 2\right)\right) = \frac{1}{12C_t^P\gamma_t}\left(\frac{\sqrt{V}}{1 - \sigma_2(P)} + 2\right). \tag{7}$$

Follow the definition $\bar{L}_t$, we have

$$\begin{aligned}
\left\|\bar{L}_t - \tilde{L}_{v,t}^{obs}\right\|_\infty &= V\left\|\sum_{s=1}^{t-1}\sum_{u=1}^{V}\left(\mathbf{1}_K/V - P_{u,v}^{t-s+1}\right)\tilde{\ell}_{u,s} + \left(\frac{1}{V}\sum_{u=1}^{V}\tilde{\ell}_{u,t} - \tilde{\ell}_{v,t}\right)\right\|_\infty \\
&\le \sum_{s=1}^{t-1}\sum_{u=1}^{V} V\|\tilde{\ell}_{v,t}\|_\infty \left|\mathbf{1}_K/V - P_{u,v}^{t-s+1}\right| + \sum_{u=1}^{V}\left\|\tilde{\ell}_{u,t} - \tilde{\ell}_{v,t}\right\|_\infty
\end{aligned}$$

$$\leq \sum_{s=1}^{t-1} \frac{1}{12C_t^P\gamma_t}\|P_{u,v}^{t-s+1} - \mathbf{1}_K/V\|_1 + \frac{1}{12C_t^P\gamma_t}. \tag{8}$$

From (23) in Duchi et al. (2011), $\|P_{u,v}^{t-s+1} - \mathbf{1}_K/V\|_1 \leq \sqrt{V}\sigma_2(P)^{t-s+1}$. Hence, if

$$t - s \geq \frac{\log \epsilon^{-1}}{\log \sigma_2(P)^{-1}} - 1 \quad \text{we immediately have} \quad \left\|[P_{u,v}^{t-s+1} - \mathbf{1}_K/V\right\|_1 \leq \sqrt{V}\epsilon.$$

Thus, by setting $\epsilon^{-1} = Vt$, for $t - s + 1 \geq \frac{\log(Vt)}{\log \sigma_2(P)^{-1}}$, we have

$$\left\|[P_{u,v}^{t-s+1} - \mathbf{1}_K/V\right\|_1 \leq 1/t. \tag{9}$$

For large s, we simply have $\left\|[P_{u,v}^{t-s+1} - \mathbf{1}_K/V\right\|_1 \leq 1$. The above The above suggests that we split the sum at $\hat{t} = \frac{\log(Vt)}{\log \sigma_2(P)^{-1}}$. We break apart the sum in equation 8 and use equation 9 to see that since $t - 1 - (t - \hat{t}) = \hat{t}$ and there are at most t steps in the summation,

$$\begin{aligned}\|\bar{L}_t - \hat{L}_{v,t}^{obs}\|_\infty &\leq \frac{1}{12C_t^P\gamma_t}\left(\sum_{s=t-\hat{t}}^{t-1}\|P_{u,v}^{t-s+1} - \mathbf{1}_K/V\|_1 + \sum_{s=1}^{t-1-\hat{t}}\|P_{u,v}^{t-s+1} - \mathbf{1}_K/V\|_1 + 1\right) \\ &\leq \frac{1}{12C_t^P\gamma_t}\left(\frac{\log(Vt)}{\log \sigma_2(P)^{-1}} + 2\right) \leq \frac{1}{12\gamma_t},\end{aligned} \tag{10}$$

Where the last inequality follows from the concavity of $\log(\cdot)$, since $\log \sigma_2(P)^{-1} \geq 1 - \sigma_2(P)$. Combine equation 7 and equation 10 completes the left statement. Similarly, we can use the above analysis to completes the right statement. □

10 Proof of Lemma 1

Proof. By Lemma 3, for any two agent u, v, and any fixed k, t, we can get

$$\|\hat{L}_{v,t}^{obs} - \hat{L}_{u,t}^{obs}\|_\infty \leq \|\hat{L}_{v,t}^{obs} - \bar{L}_t\|_\infty + \|\bar{L}_t - \hat{L}_{u,t}^{obs}\|_\infty \leq \frac{1}{6\gamma_t},$$

and

$$\|\hat{L}_{u,t}^{obs} - \hat{L}_{v,t}^{obs}\|_\infty \leq \|\hat{L}_{u,t}^{obs} - \bar{L}_t\|_\infty + \|\bar{L}_t - \hat{L}_{v,t}^{obs}\|_\infty \leq \frac{1}{6\gamma_t}.$$

Since

$$\begin{aligned}\frac{\sum_{k=1}^K f_t''(x_{u,t}^k)^{-1}(L_{v,t}(k) - L_{u,t}(k))}{\sum_{k=1}^K f_t''(x_{u,t}^k)^{-1}} &\leq \frac{\sum_{k=1}^K f_t''(x_{u,t}^k)^{-1}\|L_{v,t} - L_{u,t}\|_\infty}{\sum_{k=1}^K f_t''(x_{u,t}^k)^{-1}} \\ &= \|L_{v,t} - L_{u,t}\|_\infty \leq \frac{1}{6\gamma_t},\end{aligned}$$

and

$$L_{u,t}(k) - L_{v,t}(k) \leq \|L_{u,t} - L_{v,t}\|_\infty \leq \frac{1}{6\gamma_t}.$$

Using Lemma 2 gives us that

$$x_{v,t}^k \leq \frac{1}{1 - 1/6 - 1/6}x_{u,t}^k = \frac{3}{2}x_{u,t}^k.$$

Similarly, we can get

$$x_{u,t}^k \leq \frac{1}{1 - 1/6 - 1/6}x_{v,t}^k = \frac{3}{2}x_{v,t}^k.$$

□

11 PROOF OF THEOREM 1

11.1 LEMMAS

Lemma 4. *For any two agents u, v, assume that for t and s there exists α such that $x_{v,t}^k \leq \alpha x_{u,s}^k$ for all $k \in [K]$ and let $f(x) = -2\eta_t^{-1}\sum_{k=1}^K x_k^{\frac{1}{2}} + \gamma_t^{-1}\sum_{k=1}^K x_k(\log(x_k) - 1)$, then*

$$\frac{\sum_{k=1}^K f''(x_{v,t}^k)^{-1}\hat{\ell}_{u,s}^k}{\sum_{k=1}^K f''(x_{v,t}^k)^{-1}} \leq 2\alpha(K-1)^{\frac{1}{3}}.$$

Proof. Now we aim to bound for any $s \in A$.

$$\begin{aligned}
\frac{\sum_{k=1}^K f''(x_{v,t}^k)^{-1}\hat{\ell}_{u,s}^k}{\sum_{k=1}^K f''(x_{v,t}^k)^{-1}} &= \frac{f''(x_{v,t}(k_{u,s}))^{-1}x_{u,s}(k_{u,s})^{-1}\ell_{u,s}(k_{u,s})}{\sum_{k=1}^K f''(x_{v,t}^k)^{-1}} \\
&\leq \frac{f''(x_{v,t}(k_{u,s}))^{-1}x_{v,t}(k_{u,s})^{-1}(x_{v,t}(k_{u,s})/x_{u,s}(k_{u,s}))}{\sum_{k=1}^K f''(x_{v,t}^k)^{-1}} \\
&\leq \frac{f''(x_{v,t}(k_{u,s}))^{-1}\alpha x_{v,t}(k_{u,s})^{-1}}{\sum_{k=1}^K f''(x_{v,t}^k)^{-1}} \\
&\leq \frac{\alpha f''(x_{v,t}(k_{u,s}))^{-1}x_{v,t}(k_{u,s})^{-1}}{(K-1)f''(\frac{1-x_{v,t}(k_{u,s})}{K-1})^{-1}) + f''(x_{v,t}(k_{u,s}))^{-1}} \quad \text{Define } z := x_{v,t}(k_{u,s}) \\
&= \frac{\alpha(\eta_t z^{-3/2} + 2\gamma_t z^{-1})^{-1}z^{-1}}{(K-1)(\eta_t(\frac{1-z}{K-1})^{-3/2} + 2\gamma_t(\frac{1-z}{K-1})^{-1})^{-1} + (\eta_t z^{-3/2} + 2\gamma_t z^{-1})^{-1}} \\
&= \alpha\left((1-z)\frac{\eta_t z^{-1/2} + 2\gamma_t}{\eta_t\sqrt{K-1}(1-z)^{-1/2} + 2\gamma_t} + z\right)^{-1} \qquad (11)
\end{aligned}$$

where the first inequality follows by $\ell_{u,s}(k_{u,s}) \leq 1$, the second one holds because of induction assumption that tells us for $s \leq t : t - s \leq D \Rightarrow x_{v,t}^k \leq \alpha x_{u,s}^k$, and the third inequality is due to convexity of $f''(x)^{-1}$ from Fact 1. Now for z we have two cases, $z < \frac{1}{K}$ and $z \geq \frac{1}{K}$.

a) $z \leq \frac{1}{K}$: This case implies

$$\begin{aligned}
\frac{1-z}{z} = \frac{1}{z} - 1 \geq K - 1 &\Rightarrow (1-z)^{-1/2}\sqrt{K-1} \leq z^{-1/2} \\
&\Rightarrow 1 \leq \frac{\eta_t z^{-1/2} + 2\gamma_t}{\eta_t\sqrt{K-1}(1-z)^{-1/2} + 2\gamma_t} \qquad (12)
\end{aligned}$$

Plugging equation 12 into equation 11 gives us

$$\frac{\sum_{k=1}^K f''(x_{v,t}^k)^{-1}\hat{\ell}_{u,s}^k}{\sum_{k=1}^K f''(x_{v,t}^k)^{-1}} \leq \alpha(1 - z + z)^{-1} = \alpha$$

b) $z \geq \frac{1}{K}$: Similar to previous case $z \geq \frac{1}{K}$ implies $\eta_t z^{-1/2} \leq \eta_t\sqrt{K-1}(1-z)^{-1/2}$ so the minimum of $\frac{\eta_t z^{-1/2}+2\gamma_t}{\eta_t\sqrt{K-1}(1-z)^{-1/2}+2\gamma_t}$ occurs when $2\gamma_t = 0$. So substituting $2\gamma_t = 0$ in equation 11 leads us to have

$$\frac{\sum_{k=1}^K f''(x_{v,t}^k)^{-1}\hat{\ell}_{u,s}^k}{\sum_{k=1}^K f''(x_{v,t}^k)^{-1}} \leq \alpha((1-z)^{3/2}z^{-1/2}(K-1)^{-1/2} + z)^{-1} \qquad (13)$$

In this case again we have two following cases

b1) $z \geq \frac{1}{(K-1)^{1/3}+1}$: With this we have

$$\alpha((1-z)^{3/2}z^{-1/2}(K-1)^{-1/2}+z)^{-1} \leq \alpha z^{-1} \leq \alpha V\left((K-1)^{1/3}+1\right) \leq 2\alpha(K-1)^{1/3}$$

b2) $z \leq \frac{1}{(K-1)^{1/3}+1}$: This tells us that $(1-z) \geq \frac{(K-1)^{1/3}}{(K-1)^{1/3}+1} \geq \frac{1}{2}$ where we can use it in equation 13 as the following

$$\begin{aligned}\alpha\left((1-z)^{3/2}z^{-1/2}(K-1)^{-1/2}+z\right)^{-1} &\leq \alpha(\frac{z^{-1/2}(K-1)^{-1/2}}{\sqrt{8}}+z)^{-1} \\ &= \alpha(\frac{z^{-1/2}(K-1)^{-1/2}}{2\sqrt{8}}+\frac{z^{-1/2}(K-1)^{-1/2}}{2\sqrt{8}}+z)^{-1} \\ &\leq \frac{\alpha}{3}(\frac{(K-1)^{-1}}{32})^{-1/3} \leq 2\alpha(K-1)^{1/3}\end{aligned}$$

where the second inequality uses AM-GM inequality.

So at the end combining results of all cases to complete the proof. □

Lemma 5. *For any fixed s, t, given $x_1 = \nabla \bar{F}_s^*(-L)$ and $x_2 = \nabla \bar{F}_t^*(-L)$, if we have $s \leq t$ and $t - s \leq D$, then*

$$\forall\, k \in [K]: \quad x_2^k \leq \frac{5}{4}x_1^k.$$

Proof. Since $x_1 = \nabla \bar{F}_s^*(-\hat{L})$ and $x_2 = \nabla \bar{F}_t^*(-\hat{L})$, by the KKT conditions $\exists \mu_1, \mu_2$ s.t. $\forall k$:

$$f_s^{'}(x_1^k) = -L(k) + \mu_1, \quad f_t^{'}(x_2^k) = -L(k) + \mu_2.$$

We also know that $\exists\, k : x_1^k \geq x_2^k$ which leads to have

$$-L(k) + \mu_2 = f_t^{'}(x_2^k) \leq f_s^{'}(x_2^k) \leq f_s^{'}(x_1^k) = -L_(k) + \mu_1,$$

where the first inequality holds because the learning rates are decreasing and the second inequality is due to the fact that $f_s^{'}(x)$ is increasing. This implies that $\mu_2 \leq \mu_1$ which gives us the following inequality for all k:

$$f_t^{'}(x_2^k) = -\frac{1}{\eta_t\sqrt{x_2^k}} + \gamma_t^{-1}\log(x_2^k) \leq -\frac{1}{\eta_s\sqrt{x_1^k}} + \gamma_s^{-1}\log(x_1^k) = f_s^{'}(x_1^k).$$

Define $\beta = x_2^k/x_1^k$. So using above inequality we have

$$\begin{aligned}\frac{1}{\eta_s\sqrt{x_1^k}} - \gamma_s^{-1}\log(x_1^k) &\leq \frac{1}{\eta_t\sqrt{\beta x_1^k}} - \gamma_t^{-1}\log(x_1^k) - \gamma_t^{-1}\log(\beta) \\ \Rightarrow \frac{1}{\sqrt{\beta}} &\geq \frac{\eta_t}{\eta_s} + 2\sqrt{x_1^k}\log(\sqrt{x_1^k})\left(\frac{\eta_t}{\gamma_t} - \frac{\eta_t}{\gamma_s}\right) + \log(\beta)\frac{\eta_t}{\gamma_t}\sqrt{x_1^k} \\ &\geq \frac{\eta_t}{\eta_s} + \min_{0<z\leq 1}\left\{2z\log(z)\left(\frac{\eta_t}{\gamma_t} - \frac{\eta_t}{\gamma_s}\right) + \log(\beta)\frac{\eta_t}{\gamma_t}z\right\} \\ &\overset{(a)}{=} \frac{\eta_t}{\eta_s} - \frac{2}{e}\left(\frac{\eta_t}{\gamma_t} - \frac{\eta_t}{\gamma_s}\right)\left(\frac{1}{\sqrt{\beta}}\right)^{\frac{\gamma_t^{-1}}{\gamma_t^{-1}-\gamma_s^{-1}}} \\ &\overset{(b)}{\geq} \frac{\eta_t}{\eta_s} - \frac{2}{e}\left(\frac{\eta_t}{\gamma_t} - \frac{\eta_t}{\gamma_s}\right)\frac{1}{\sqrt{\beta}}.\end{aligned}$$

□

where (a) holds because the subject function of the minimization problem is convex and equating the first derivative to zero gives $z = \beta^{\frac{\gamma_t^{-1}}{\gamma_t^{-1}-\gamma_s^{-1}}}$, and (b) follows by $\frac{\gamma_t^{-1}}{\gamma_t^{-1}-\gamma_s^{-1}} \geq 1$. So rearranging the above result gives

$$\beta \leq \left(\frac{\eta_s}{\eta_t} + \frac{2}{e}\left(\frac{\eta_t}{\gamma_t} - \frac{\eta_t}{\gamma_s}\right)\right)^2. \tag{14}$$

Therefore, we have

$$\frac{\eta_t}{\eta_s} = \frac{4\sqrt{Vt + 169V^2 D\log(K)}}{4\sqrt{Vs + 169V^2 D}} = \sqrt{1 + \frac{V(t-s)}{169V^2 D}} \le \sqrt{1 + \frac{1}{169}}.$$

where $V \ge 1$, $D \ge 1$ and $t - s \le D$. First, we give an inequality

$$\sqrt{x+a} - \sqrt{y+a} \le \sqrt{x-y}, \quad x \ge y \ge 0, \quad a \ge 0.$$

We square the left side:

$$(\sqrt{x+a} - \sqrt{y+a})^2 = x+a-2\sqrt{x+a}\sqrt{y+a}+y+a \le x+a-2\sqrt{y+a}\sqrt{y+a}+y+a = x-y.$$

By this inequality, we have

$$\begin{aligned}
\frac{2}{e}\left(\frac{\eta_t}{\gamma_t} - \frac{\eta_t}{\gamma_s}\right) &= \frac{2}{e}\left(\frac{8V\sqrt{t/\log(K) + 36D^2(K-1)^{\frac{2}{3}} + 4(C_t^P)^2}}{4\sqrt{Vt + 169V^2 D}}\right. \\
&\quad \left. - \frac{8V\sqrt{s/\log(K) + 36D^2(K-1)^{\frac{2}{3}} + 4(C_t^P)^2}}{4\sqrt{Vt + 169V^2 D}}\right) \\
&\le \frac{2}{e}\left(\frac{2V(\sqrt{t-s})}{\sqrt{Vt + 169V^2 D}}\right) \\
&\le \frac{4}{e}\left(\frac{V\sqrt{t-s}}{13V\sqrt{D}}\right) \le \frac{4}{13e}.
\end{aligned}$$

Plugging the above inequalities gives us the following bound:

$$\beta \le \left(\sqrt{1 + \frac{1}{169}} + \frac{4}{13e}\right)^2 < \frac{5}{4}.$$

Lemma 6. *For any time step $t \ge s > D$, $t - s \le D$ and any fixed arm k, then we have*

$$x_{v,t}^k \le 2x_{u,s}^k.$$

Proof. First, we decompose $\hat{L}_{v,t}^{obs}$ into the following two parts:

$$\hat{L}_{v,t}^{obs} = \hat{L}_{v,1\to s}^{obs} + \hat{L}_{v,s+1\to t}^{obs},$$

where the former represents the cumulative loss estimate observed by agent v from time step 1 to s, and the latter is the cumulative loss estimate from time step $s+1$ to t.

Using the same analytical method in Lemma 3, we can obtain:

$$\begin{aligned}
\|\hat{L}_{u,s}^{obs} - \hat{L}_{v,t}^{obs}\|_\infty &\le \|\hat{L}_{u,s}^{obs} - \hat{L}_{v,1\to s}^{obs}\|_\infty \\
&\le \frac{1}{12C_t^P\gamma_t}\left(\frac{\min\{\sqrt{V}, \log(Vt)\}}{1 - \sigma_2(P)} + 2\right) \le \frac{1}{12\gamma_t}.
\end{aligned} \tag{15}$$

Where the first inequality because that $\hat{L}_{v,s+1\to t}^{obs}(k) \ge 0$. As mentioned before, for any fixed k we have

$$\begin{aligned}
\hat{L}_{v,1\to s}^{obs}(k) - \hat{L}_{u,s}^{obs}(k) &\le V\sum_{\hat{t}=s-D}^{t-D} m_{\hat{t}}(k) + \|\hat{L}_{v,1\to s}^{obs} - L_{u,s}^{obs}(k)\|_\infty \\
&\le V\sum_{\hat{t}=s-D}^{t-D} m_{\hat{t}}(k) + \frac{1}{12C_t^P\gamma_t}\left(\frac{\min\{\sqrt{V}, \log(Vt)\}}{1 - \sigma_2(P)} + 2\right).
\end{aligned} \tag{16}$$

Where the first inequality because only the records generated in time period $[s-D, t-D]$ will be used in $\hat{L}^{obs}_{v,1\to s}(k)$. By Lemma 4 and mathematical induction, we have

$$V\sum_{\hat{t}=s-D}^{t-D}\frac{\sum_{k=1}^{K}f''(x_{u,s}^{k})^{-1}m_{\hat{t}}(k)}{\sum_{k=1}^{K}f''(x_{u,s}^{k})^{-1}}=\sum_{u=1}^{V}\frac{\sum_{k=1}^{K}f''(x_{u,s}^{k})^{-1}\hat{\ell}_{u,\hat{t}}(k)}{\sum_{k=1}^{K}f''(x_{u,s}^{k})^{-1}}\le 8V(K-1)^{\frac{1}{3}}. \tag{17}$$

According to our update rules for deviation records, no records generated in time period $[t-D+1, t]$ will be used, so we have

$$\begin{aligned}\|\hat{L}^{obs}_{v,s+1\to t}\|_\infty &= V\left\|\sum_{\hat{t}=s+1}^{t-1}\sum_{u=1}^{V}P_{u,v}^{t-\hat{t}-1}\tilde{\ell}_{u,t}+\tilde{\ell}_{v,t}\right\|_\infty\\ &\le \sum_{\hat{t}=s+1}^{t-1}\sum_{u=1}^{V}V\|\tilde{\ell}_{u,t}\|_\infty P_{u,v}^{t-\hat{t}-1}+V\|\tilde{\ell}_{v,t}\|_\infty\le\frac{D}{12C_t^P\gamma_t}. \end{aligned}\tag{18}$$

Combine equation 15, equation 16, equation 17 and equation 18, we can complete the right statement

$$\begin{aligned}&\frac{\sum_{k=1}^{K}f''(x_{u,s}^{k})^{-1}(\hat{L}_{u,s}(k)-\hat{L}^{obs}_{v,t}(k))}{\sum_{k=1}^{K}f''(x_{u,s}^{k})^{-1}}\\ &\qquad\le\frac{1}{12C_t^P\gamma_t}\left(\frac{\min\{\sqrt{V},\log(Vt)\}}{1-\sigma_2(P)}+2+D\right)+8V(K-1)^{\frac{1}{3}}\\ &\qquad\le\frac{1}{4\gamma_t}.\end{aligned}$$

Where the last inequality uses the fact

$$\gamma_t^{-1}=8V\sqrt{C_t^Pt/\log(K)+36D^2(K-1)^{2/3}+4(C_t^P)^2}\ge 48V(K-1)^{\frac{1}{3}}.$$

Using Lemma 2 and Lemma 5, we can complete the proof:

$$x_{v,t}^k\le\frac{1}{1-1/12-1/4}\times\frac{5}{4}x_{u,s}^k\le 2x_{u,s}^k.$$

□

Lemma 7. *For any time step $t>D$ and fixed arm k, for any two agents u, v, then*

$$\|\hat{L}_{u,t-D}-\hat{L}^{obs}_{v,t}\|_\infty\le\frac{1}{12\gamma_t}\quad\textit{and}\quad\frac{\sum_{k=1}^{K}f''(x_{u,t-D}^{k})^{-1}(\hat{L}_{u,t-D}(k)-\hat{L}^{obs}_{v,t}(k))}{\sum_{k=1}^{K}f''(x_{u,t-D}^{k})^{-1}}\le\frac{1}{6\gamma_t}.$$

Proof. First, we decompose $\hat{L}^{obs}_{v,t}$ into the following two parts:

$$\hat{L}^{obs}_{v,t}=\hat{L}^{obs}_{v,1\to t-D}+\hat{L}^{obs}_{v,t-D+1\to t},$$

where the former represents the cumulative loss estimate observed by agent v from time step 1 to $t-D$, and the latter is the cumulative loss estimate from time step $t-D+1$ to t.

Using the same analytical method in Lemma 3, we can obtain:

$$\begin{aligned}\|\hat{L}_{u,t-D}-\hat{L}^{obs}_{v,1\to t-D}\|_\infty &< \|\hat{L}_{1,t-D+1}-\hat{L}^{obs}_{v,1\to t-D}\|_\infty\\ &\le\frac{1}{12C_t^P\gamma_t}\left(\frac{\min\{\sqrt{V},\log(Vt)\}}{1-\sigma_2(P)}+2\right)\le\frac{1}{12\gamma_t}.\end{aligned}\tag{19}$$

Since for any k we have $\hat{L}^{obs}_{v,t-D+1\to t}(k)\ge 0$, the left statement is complete. As mentioned before, for any fixed k we have

$$\hat{L}^{obs}_{v,1\to t-D}(k)-\hat{L}_{u,t-D}(k)\le(V-u)m_{t-D}(k)+\|\hat{L}^{obs}_{v,1\to t-D}-\hat{L}_{1,t-D+1}\|_\infty$$

$$\leq Vm_t(k) + \frac{1}{12C_t^P\gamma_t}\left(\frac{\min\{\sqrt{V},\log(Vt)\}}{1-\sigma_2(P)}+2\right). \tag{20}$$

By Lemma 4, we have

$$\frac{\sum_{k=1}^K f''(x_{u,t-D}^k)^{-1}Vm_{t-D}(k)}{\sum_{k=1}^K f''(x_{u,t-D}^k)^{-1}} = \sum_{u=1}^V \frac{\sum_{k=1}^K f''(x_{u,t-D}^k)^{-1}\hat{\ell}_{t-D}(k)}{\sum_{k=1}^K f''(x_{u,t-D}^k)^{-1}} \leq 4V(K-1)^{\frac{1}{3}}. \tag{21}$$

According to our update rules for deviation records, no records generated in time period $[t-D+1,t]$ will be used, so we have

$$\begin{aligned}\|\hat{L}^{obs}_{v,t-D+1\to t}\|_\infty &= V\left\|\sum_{s=t-D+1}^{t-1}\sum_{u=1}^V P_{u,v}^{t-s-1}\tilde{\ell}_{u,t}+\tilde{\ell}_{v,t}\right\|_\infty \\ &\leq \sum_{s=t-D+1}^{t-1}\sum_{u=1}^V V\|\tilde{\ell}_{u,t}\|_\infty P_{u,v}^{t-s-1} + V\|\tilde{\ell}_{v,t}\|_\infty \leq \frac{D}{12C_t^P\gamma_t}. \end{aligned}\tag{22}$$

Combine equation 19, equation 20, equation 21 and equation 22, we can complete the right statement

$$\begin{aligned}&\frac{\sum_{k=1}^K f''(x_{u,t-D}^k)^{-1}(\hat{L}_{u,t-D}(k)-\hat{L}^{obs}_{v,t}(k))}{\sum_{k=1}^K f''(x_{u,t-D}^k)^{-1}} \\ &\qquad\leq \frac{1}{12C_t^P\gamma_t}\left(\frac{\min\{\sqrt{V},\log(Vt)\}}{1-\sigma_2(P)}+2+D\right)+4V(K-1)^{\frac{1}{3}} \\ &\qquad\leq \frac{1}{6\gamma_t}.\end{aligned}$$

Where the last inequality uses the fact

$$\gamma_t^{-1} = 8V\sqrt{C_t^P t/\log(K) + 36D^2(K-1)^{2/3} + 4(C_t^P)^2} \geq 48V(K-1)^{\frac{1}{3}}.$$

□

Lemma 8. *For any two agents u, v, and any time step $t > D$, defining $\tilde{x}_{u,t} = \nabla\bar{F}_t^*(-\hat{L}_{u,t})$. By Lemma 2, Lemma 5 and Lemma 7, for any fixed k we can get*

$$x_{v,t}^k \leq \frac{1}{1-1/12-1/6}\times\frac{5}{4}\tilde{x}_{u,t-D}^k = \frac{5}{3}\tilde{x}_{u,t-D}^k.$$

Lemma 9. *For any two time steps t, any fixed arm k, and any two agents u, v, defining $\tilde{x}_{u,t}^k = \nabla\bar{F}_t^*(-\hat{L}_{v,t})$, then*

$$x_{u,t}^k \leq 2\tilde{x}_{v,t}^k.$$

Proof. Using the same analytical method in Lemma 3, we can obtain:

$$\begin{aligned}\|\hat{L}^{obs}_{v,t}-\hat{L}_{u,t}\|_\infty &\leq \|\hat{L}^{obs}_{v,t-1}-\bar{L}_{t-1}\|_\infty + \|\tilde{\ell}_{v,t}\|_\infty \\ &\leq \frac{1}{12C_t^P\gamma_t}\left(\frac{\min\{\sqrt{V},\log(Vt)\}}{1-\sigma_2(P)}+2\right)+\frac{1}{12C_t^P\gamma_t} \\ &\leq \frac{1}{12C_t^P\gamma_t}\left(\frac{\min\{\sqrt{V},\log(Vt)\}}{1-\sigma_2(P)}+2+D\right) \leq \frac{1}{12\gamma_t}.\end{aligned}\tag{23}$$

As mentioned before, for any fixed k we have

$$\hat{L}_{u,t}(k) - \hat{L}^{obs}_{v,t}(k) \leq V\sum_{\hat{t}=t-D}^{t-1} m_{\hat{t}}(k) + \|\hat{L}^{obs}_{v,t}-\bar{L}_t\|_\infty$$

$$\leq V \sum_{\hat{t}=t-D}^{t-1} m_{\hat{t}}(k) + \frac{1}{12C_t^P \gamma_t} \left(\frac{\min\{\sqrt{V}, \log(Vt)\}}{1-\sigma_2(P)} + 2 \right). \quad (24)$$

Using mathematical analysis, we assume that from 1 to t the lemma holds. Then using the same analytical method in Lemma 6, for any time step $t-D \leq \hat{t} \leq t-1$, we can obtain:

$$\tilde{x}^k_{v,t} \leq 2\tilde{x}^k_{u,\hat{t}} \leq 4x^k_{u,\hat{t}}.$$

By Lemma 4, we have

$$\sum_{\hat{t}=t-D}^{t-1} \frac{\sum_{k=1}^K f''(\tilde{x}^k_{u,t})^{-1} V m_{\hat{t}}(k)}{\sum_{k=1}^K f''(\tilde{x}^k_{v,t})^{-1}} = \sum_{\hat{t}=t-D}^{t} \sum_{u=1}^{V} \frac{\sum_{k=1}^K f''(\tilde{x}^k_{v,t})^{-1} \hat{\ell}_{\hat{t}}(k)}{\sum_{k=1}^K f''(\tilde{x}^k_{v,t})^{-1}} \leq 8VD(K-1)^{\frac{1}{3}}. \quad (25)$$

According to our update rules for deviation records, no records generated in time period $[t-D+1, t]$ will be used, so we have

$$\begin{aligned}
\|\hat{L}^{obs}_{v,t-D+1\to t}\|_\infty &= V \left\| \sum_{s=t-D+1}^{t-1} \sum_{u=1}^{V} P^{t-s-1}_{u,v} \tilde{\ell}_{u,t} + \tilde{\ell}_{v,t} \right\|_\infty \\
&\leq \sum_{s=t-D+1}^{t-1} \sum_{u=1}^{V} V\|\tilde{\ell}_{u,t}\|_\infty P^{t-s-1}_{u,v} + V\|\tilde{\ell}_{v,t}\|_\infty \leq \frac{D}{12C_t^P\gamma_t}. \quad (26)
\end{aligned}$$

Combine equation 19, equation 20, equation 21 and equation 22, we can complete the right statement

$$\begin{aligned}
&\frac{\sum_{k=1}^K f''(x^k_{v,t})^{-1}(\hat{L}_{u,t-D}(k) - \hat{L}^{obs}_{v,t}(k))}{\sum_{k=1}^K f''(x^k_{v,t})^{-1}} \\
&\qquad \leq \frac{1}{12C_t^P\gamma_t}\left(\frac{\min\{\sqrt{V}, \log(Vt)\}}{1-\sigma_2(P)} + 2 + D\right) + 8VD(K-1)^{\frac{1}{3}} \\
&\qquad \leq \frac{1}{4\gamma_t}.
\end{aligned}$$

Where the last inequality uses the fact

$$\gamma_t^{-1} = 8V\sqrt{C_t^P t/\log(K) + 36D^2(K-1)^{2/3} + 4(C_t^P)^2} \geq 48V(K-1)^{\frac{1}{3}}.$$

By By Lemma 2, we can get

$$\tilde{x}^k_{v,t} \leq \frac{1}{1-1/12-1/4} x^k_{v,t} \leq 2x^k_{v,t}.$$

□

11.2 Adversarial Bounds

As a consequence, the group regret bound as follows:

$$\begin{aligned}
\sum_{v=1}^{V} R_T(v) &= \sum_{v=1}^{V} \mathbb{E}\left[\sum_{t=1}^{T} \langle \bar{\ell}_t, x_{v,t}\rangle - \bar{\ell}_t(k^*)\right] \\
&\overset{(a)}{=} \sum_{v=1}^{V} \mathbb{E}\left[\sum_{t=1}^{T} (\langle \mathbb{E}[m_t], x_{v,t}\rangle - \mathbb{E}[m_t(k^*)])\right] \\
&\overset{(b)}{\leq} \mathbb{E}\left[\sum_{v=1}^{V}\sum_{t=1}^{T} \langle m_t, x_{v,t}\rangle - \hat{L}_{1,T+1}(k^*)\right] \\
&= \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{v=1}^{V} \left(\bar{F}_t^*(-\hat{L}^{obs}_{v,t} - m_t) - \bar{F}_t^*(-\hat{L}^{obs}_{v,t}) + \langle x_{v,t}, m_t\rangle\right)\right.}_{(A)}
\end{aligned}$$

$$+\underbrace{\sum_{t=1}^{T}\sum_{v=1}^{V}\Big(\bar{F}_t^*(-\hat{L}_{v,t}^{obs}) - \bar{F}_t^*(-\hat{L}_{v,t}^{obs} - m_t) - \bar{F}_t^*(-\hat{L}_{v,t}) + \bar{F}_t^*(-\hat{L}_{v+1,t})\Big)}_{(B)}$$

$$+\underbrace{\left(\sum_{v=1}^{V}\sum_{t=1}^{T}\bar{F}_t^*(-\hat{L}_{v,t}) - \bar{F}_t^*(-\hat{L}_{v+1,t})\right) - \hat{L}_{1,T+1}(k^*)\Bigg]}_{(C)}.$$

Where (a) holds because the following facts for all arms k:

$$\bar{\ell}_t(k) = \frac{1}{V}\sum_{v=1}^{V}\ell_{v,t}(k) = \frac{1}{V}\sum_{v=1}^{V}x_{v,t}(k)\frac{\ell_{v,t}(k)}{x_{v,t}(k)} = \mathbb{E}[m_t(k)],$$

(b) holds because the following definition:

$$\hat{L}_{1,T+1}(k^*) = \sum_{t=1}^{T}Vm_t(k^*) = \sum_{v=1}^{V}\sum_{t=1}^{T}m_t(k^*).$$

11.2.1 Bounding (A)

We set an indicator variable $Y_{k,t} = \sum_{v=1}^{V}\mathbb{I}(k_{v,t} = k)$, which represents how many agents have selected arm k at round t. Through this definition, for each agent v we have:

$$m_{k,t} = \frac{1}{V}\sum_{v:k_{v,t}=k}\frac{\ell_{v,t}}{x_{v,t}(k)} \leq \frac{1}{V}\sum_{k_{v,t}=k}\frac{1}{x_{v,t}(k)} \leq \frac{3Y_{k,t}}{2Vx_{v,t}(k)}. \tag{27}$$

where the first inequality from all $\ell_{v,t} \leq 1$, and the second inequality follows that Lemma 1.

$$\begin{aligned}
&\sum_{v=1}^{V}\bar{F}_t^*(-\hat{L}_{v,t}^{obs} - m_t) - \bar{F}_t^*(-\hat{L}_{v,t}^{obs}) + \langle x_{v,t}, m_t\rangle \\
&\overset{(a)}{=} \sum_{v=1}^{V}\bar{F}_t^*(-\nabla\bar{F}_t(x_{v,t}) - m_t) - \bar{F}_t^*(-\nabla\bar{F}_t(x_{v,t})) + \langle x_{v,t}, m_t\rangle \\
&\overset{(b)}{\leq} \sum_{v=1}^{V}F_t^*(-\nabla\bar{F}_t(x_{v,t}) - m_t) - F_t^*(-\nabla\bar{F}_t(x_{v,t})) + \langle x_{v,t}, m_t\rangle \\
&= \sum_{v=1}^{V}\sum_{k=1}^{K}D_{f_t^*}\left(f'(x_{v,t}) - m_{k,t}, f'(x_{v,t})\right) \\
&\overset{(c)}{=} \sum_{v=1}^{V}\frac{V}{Y_{k_{v,t},t}}D_{f_t^*}\left(f'(x_{v,t}(k_{v,t})) - m_{k_{v,t},t}), f'(x_{v,t}(k_{v,t}))\right) \\
&\overset{(d)}{\leq} \sum_{v=1}^{V}\frac{V}{Y_{k_{v,t},t}}D_{f_t^*}\left(f'(x_{v,t}(k_{v,t})) - \frac{3Y_{k_{v,t},t}}{2Vx_{v,t}(k_{v,t})}, f'(x_{v,t}(k_{v,t}))\right) \\
&\overset{(e)}{\leq} \sum_{v=1}^{V}\frac{9Y_{k_{v,t},t}}{8V(x_{v,t}(k_{v,t}))^2 f_t''(x_{v,t}(k_{v,t}))} \\
&\leq \frac{9}{8}\sum_{v=1}^{V}\frac{1}{(x_{v,t}(k_{v,t}))^2 f_t''(x_{v,t}(k_{v,t}))} \\
&\leq \frac{9}{32}\sum_{v=1}^{V}\frac{x_{v,t}(k_{v,t})^{3/2}}{(x_{v,t}(k_{v,t}))^2\sqrt{Vt + 169V^2D}}
\end{aligned}$$

$$= \frac{9}{32} \sum_{v=1}^{V} \frac{1}{(x_{v,t}(k_{v,t}))^{\frac{1}{2}} \sqrt{Vt}}.$$

Where (a) applies Facts 2 and 3, the (b) follows from both parts of Fact 4, (d) holds because (27), (e) uses Fact 6, and (c) uses the following equality for any arm k:

$$\sum_{v=1}^{V} D_{f_t^*}\left(f'(x_{v,t}) - m_{k,t}, f'(x_{v,t})\right) = \frac{V}{Y_{k,t}} \sum_{k_{v,t}=k} D_{f_t^*}\left(f'(x_{v,t}) - m_{k,t}, f'(x_{v,t})\right).$$

In expectation we get

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{v=1}^{V}\left(\bar{F}_t^*(-L_{v,t}^{obs} - m_t) - \bar{F}_t^*(-L_{v,t}^{obs}) + \langle x_{v,t}, \tilde{\ell}_t\rangle\right)\right] \leq \frac{9}{32}\sum_{t=1}^{T}\sum_{v=1}^{V}\sum_{k=1}^{K}\frac{x_{v,t}(k)^{\frac{1}{2}}}{\sqrt{Vt}} \leq \frac{9}{16}\sqrt{VKT}. \tag{28}$$

11.2.2 BOUNDING (B)

We define $\hat{L}_{v,t}^{miss} = \hat{L}_{v,t} - \hat{L}_{v,t}^{obs}$. Then we have for any $v \in [V]$ and $t \in [T]$:

$$\begin{aligned}
-\bar{F}_t^*(-\hat{L}_{v,t}) + \bar{F}_t^*(-\hat{L}_{v+1,t}) &= -\bar{F}_t^*(-\hat{L}_{v,t}) + \bar{F}_t^*(-\hat{L}_{v,t} - m_t) \\
&= -\int_0^1 \langle m_t, \nabla \bar{F}_t^*(-\hat{L}_{v,t} - x m_t)\rangle dx \\
&= -\int_0^1 \langle m_t, \nabla \bar{F}_t^*(-\hat{L}_{v,t}^{obs} - \hat{L}_{v,t}^{miss} - x m_t)\rangle dx.
\end{aligned}$$

Where the second equality holds by the fundamental theorem of calculus. Therefore, we have for any $v \in [V]$ and $t \in [T]$:

$$\begin{aligned}
&\sum_{v=1}^{V} \bar{F}_t^*(-\hat{L}_{v,t}^{obs}) - \bar{F}_t^*(-\hat{L}_{v,t}^{obs} - m_t) - \bar{F}_t^*(-\hat{L}_{v,t}) + \bar{F}_t^*(-\hat{L}_{v+1,t}) \\
&\overset{(a)}{\leq} \sum_{v=1}^{V} \int_0^1 \langle m_t, \nabla \bar{F}_t^*(-\hat{L}_{v,t}^{obs} - x m_t)\rangle dx - \int_0^1 \langle m_t, \nabla \bar{F}_t^*(-\hat{L}_{v,t}^{obs} - \hat{L}_{v,t}^{miss} - x m_t\rangle dx \\
&\overset{(b)}{=} \sum_{v=1}^{V}\sum_{k=1}^{K} \int_0^1 \langle m_{k,t}, \tilde{z}(x) - \nabla \bar{F}_t^*(\nabla F_t(\tilde{z}(x)) - \hat{L}_{v,t}^{miss})\rangle dx \\
&\overset{(c)}{\leq} \sum_{v=1}^{V}\sum_{k=1}^{K} \int_0^1 \langle m_{k,t}, \tilde{z}(x) - \nabla \bar{F}_t^*(\nabla F_t(\tilde{z}(x)) - \hat{L}_{v,t}^{miss}(k)\rangle dx \\
&\overset{(d)}{\leq} \sum_{v=1}^{V}\sum_{k=1}^{K} \int_0^1 \langle m_{k,t}, \tilde{z}(x) - \nabla F_t^*(\nabla F_t(\tilde{z}(x)) - \hat{L}_{v,t}^{miss}(k))\rangle dx \\
&= \sum_{v=1}^{V}\sum_{k=1}^{K} \int_0^1 m_{k,t}(\tilde{z}_k(x) - f_t^{*'}(f_t^{'}(\tilde{z}_k(x)) - \hat{L}_{v,t}^{miss}(k))dx \\
&\overset{(e)}{\leq} \sum_{v=1}^{V}\sum_{k=1}^{K} \int_0^1 m_{k,t} f_t^{*''}(f_t^{'}(\tilde{z}_k(x))\hat{L}_{v,t}^{miss}(k)dx \\
&= \sum_{v=1}^{V}\sum_{k=1}^{K} \int_0^1 m_{k,t} f_t^{''}(\tilde{z}_k(x))^{-1}\hat{L}_{v,t}^{miss}(k)dx \\
&\overset{(f)}{\leq} \sum_{v=1}^{V}\sum_{k=1}^{K} \int_0^1 m_{k,t} f_t^{''}\left(\frac{3}{2}x_{v,t}(k)\right)^{-1}\hat{L}_{v,t}^{miss}(k)dx \\
&\leq \frac{3\gamma_t}{2}\sum_{v=1}^{V}\sum_{k=1}^{K} m_{k,t} x_{v,t}(k)\hat{L}_{v,t}^{miss}(k)dx
\end{aligned}$$

$$\overset{(g)}{\leq} \frac{9\gamma_t}{4V} \sum_{v=1}^{V} \sum_{u=1}^{V} \hat{L}_{v,t}^{miss}(k_{u,t}).$$

Where (a) uses the Fundamental theorem of calculus together with the inequality above, (b) substitutes $\tilde{z}(x) = \nabla \bar{F}_t(-\hat{L}_{v,t}^{obs} - xm_t)$ and applies Fact 3, (c) follows from the fact that $\nabla \bar{F}_t^*(-L)_k$ decreases if the loss in coordinates other than k is reduced, (d) applies Fact 5, (e) $f_t^{*'}$ is convex, so $-f_t^{*'}(f_t'(\tilde{z}_k(x)) - \hat{L}_{v,t}^{miss}(k)) \leq -\tilde{z}_k(x) + f_t^{*''}(f_t'(\tilde{z}_k(x))\hat{L}_{v,t}^{miss}(k)$, (f) follows because $\tilde{z}_k \leq \frac{3}{2}x_{v,t}(k)$ and $F_t''(x)^{-1}$ is monotonically increasing, and (g) holds because the following inequality:

$$\begin{aligned}
\sum_{v=1}^{V}\sum_{k=1}^{K} m_{k,t}x_{v,t}(k)\hat{L}_{v,t}^{miss}(k) &= \frac{1}{V}\sum_{v=1}^{V}\sum_{k=1}^{K}\sum_{u=1}^{V} \mathbb{I}(k = k_{u,t})\hat{\ell}_{u,t}(k_{u,t})x_{v,t}(k)\hat{L}_{v,t}^{miss}(k) \\
&= \frac{1}{V}\sum_{v=1}^{V}\sum_{u=1}^{V} \hat{\ell}_{u,t}(k_{u,t})x_{v,t}(k_{u,t})\hat{L}_{v,t}^{miss}(k_{u,t}) \\
&= \frac{1}{V}\sum_{v=1}^{V}\sum_{u=1}^{V} V\ell_{u,t}(k_{u,t})\hat{L}_{v,t}^{miss}(k_{u,t})\frac{x_{v,t}(k_{u,t})}{x_{u,t}(k_{u,t})} \\
&\leq \frac{3}{2V}\sum_{v=1}^{V}\sum_{u=1}^{V} \hat{L}_{v,t}^{miss}(k_{u,t}).
\end{aligned}$$

For any fixed k, v, t, we have

$$\begin{aligned}
\mathbb{E}\left[\|\tilde{\ell}_{v,t}(k)\|_\infty\right] &= \mathbb{E}\left[\mathbb{I}(k_{v,t} = k)\left\|\frac{\ell_{v,t}(k)}{\max\{x_{v,t}(k), 12C_t^P\gamma_t\}}\right\|_\infty\right] \\
&= \left\|\frac{\ell_{v,t}(k)x_{v,t}(k)}{\max\{x_{v,t}(k), 12C_t^P\gamma_t\}}\right\|_\infty \leq 1,
\end{aligned}$$

and

$$\mathbb{E}\left[\|\hat{\ell}_{v,t}(k)\|_\infty\right] = \mathbb{E}\left[\mathbb{I}(k_{v,t} = k)\left\|\frac{\ell_{v,t}(k)}{x_{v,t}(k)}\right\|_\infty\right] = \left\|\frac{\ell_{v,t}(k)x_{v,t}(k)}{x_{v,t}(k)}\right\|_\infty \leq 1,$$

Using the same analytical method in Lemma 6, we can obtain

$$\begin{aligned}
\mathbb{E}\left[\hat{L}_{v,t}^{miss}(k_{u,t})\right] &\leq \mathbb{E}\left[\left\|\hat{L}_{v,1\to t-D} - \hat{L}_{v,1\to t-D}^{obs}\right\|_\infty + \left\|\hat{L}_{v,t-D+1\to t} - \hat{L}_{v,t-D+1\to t}^{obs}\right\|_\infty\right] \\
&\leq V\left(\frac{\min\{\sqrt{V}, \log(Vt)\}}{1-\sigma_2(P)} + 2\right) + VD = VC_t^P.
\end{aligned}$$

Finally, we have in expectation

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T}\sum_{v=1}^{V}\left(\bar{F}_t^*(-\hat{L}_{v,t}^{obs}) - \bar{F}_t^*(-\hat{L}_{v,t}^{obs} - m_t) - \bar{F}_t^*(-\hat{L}_{v,t}) + \bar{F}_t^*(-\hat{L}_{v+1,t})\right)\right] \\
&\leq \sum_{t=1}^{T}\frac{9\gamma_t}{4V}\mathbb{E}\left[\sum_{v=1}^{V}\sum_{u=1}^{V}\hat{L}_{v,t}^{miss}(k_{u,t})\right] \leq \sum_{t=1}^{T}\frac{9\gamma_t}{4V}\cdot V^3C_t^P = \frac{9V^2}{4}\sum_{t=1}^{T}C_t^P\gamma_t \\
&= \frac{9V^2}{4}\sum_{t=1}^{T}\frac{C_t^P}{8V\sqrt{C_t^P t/\log(K) + 36D^2(K-1)^{\frac{2}{3}}}} \leq \frac{9}{16}V\sqrt{C_T^P T\log(K)}.
\end{aligned} \tag{29}$$

11.2.3 BOUNDING (C)

Let $\tilde{x}_{1,t} = \underset{x\in\Delta^{K-1}}{\arg\max}\langle x, -\hat{L}_{1,t}\rangle - F_t(x)$, then

$$\bar{F}_t^*(-\hat{L}_{1,t}) = \langle \tilde{x}_{1,t}, -\hat{L}_{1,t}\rangle - F_t(\tilde{x}_{1,t}).$$

Furthermore, since $\bar{F}^*(-\hat{L}_{v,t}) = \max_{x\in\Delta^{K-1}} \langle x, -\hat{L}_{v,t}\rangle - F(x)$, we have

$$-\bar{F}^*_{t-1}(-\hat{L}_{1,t}) \le -\langle \tilde{x}_{1,t}, -\hat{L}_{1,t}\rangle + F_{t-1}(\tilde{x}_{1,t})$$
$$-\bar{F}^*_T(-\hat{L}_{1,T+1}) \le -\langle \mathbf{e}_{k^*}, -\hat{L}_{1,T+1}\rangle + F_T(\mathbf{e}_{k^*}) \le \hat{L}_{1,T+1}(k^*).$$

Plugging these inequalities into the LHS leads to

$$\begin{aligned}
&\sum_{t=1}^{T}\left(\sum_{v=1}^{V}\left(\bar{F}^*_t(-\hat{L}_{v,t}) - \bar{F}^*_t(-\hat{L}_{v+1,t})\right)\right) - \hat{L}_{1,T+1}(k^*)\\
&= \sum_{t=1}^{T}\left(\bar{F}^*_t(-\hat{L}_{1,t}) - \bar{F}^*_t(-\hat{L}_{1,t+1})\right) - \hat{L}_{1,T+1}(k^*)\\
&\le -F_1(\tilde{x}_{1,1}) + \sum_{t=2}^{T}\left(F_{t-1}(\tilde{x}_{1,t}) - F_t(\tilde{x}_{1,t})\right)\\
&\le \max_{x\in\Delta^{K-1}} -F_1(x) + \sum_{t=2}^{T}\max_{x\in\Delta^{K-1}}\left(F_{t-1}(x) - F_t(x)\right)\\
&= -F_T(\mathbf{1}_K/K)\\
&= 8\sqrt{VKT + 169V^2D} + 8V\log(K)\sqrt{C_T^P T/\log(K) + 36D^2(K-1)^{\frac{2}{3}} + 4(C_t^P)^2}\\
&\le 8\sqrt{VKT} + 8V\sqrt{C_T^P T\log(K)} + 104V\sqrt{D} + 48VD(K-1)^{\frac{1}{3}}\log(K) + 16VC_T^P\log(K).
\end{aligned} \tag{30}$$

Combine equation 28, equation 29, and equation 30, we can get

$$\sum_{v=1}^{V} R_T(v) \le \frac{137}{16}\sqrt{VKT} + \frac{137}{16}V\sqrt{C_T^P T\log(K)} + 104V\sqrt{D} + 48VD(K-1)^{\frac{1}{3}}\log(K) + 16VC_T^P\log(K).$$

For any k, v, t, by Lemma 1, we can get the individual regret for each agent v:

$$\begin{aligned}
R_T(v) &\le \frac{3}{2V}\sum_{v=1}^{V} R_T(v)\\
&< 13\sqrt{KT/V} + 13\sqrt{C_T^P T\log(K)} + 156\sqrt{D} + 72D(K-1)^{\frac{1}{3}}\log(K) + 24C_T^P\log(K).
\end{aligned}$$

11.3 Stochastic Bounds

Inspired the analysis of stochastic bound for bandit with delay feedback in Masoudian et al. (2022), let $\tilde{x}_{v,t} = \nabla\bar{F}^*_t(-\hat{L}_{v,t})$, then we define the drifted pseudo-regret as

$$R_T^{drift}(v) = \mathbb{E}\left[\sum_{t=1}^{T}\left(\langle\tilde{x}_{v,t}, \bar{\ell}_t\rangle - \bar{\ell}_t(k^*)\right)\right].$$

We rewrite the drifted regret as

$$\begin{aligned}
R_T^{drift}(v) = \mathbb{E}\left[\sum_{t=1}^{T}\left(\langle\tilde{x}_{v,t}, \bar{\ell}_t\rangle - \bar{\ell}_t(k^*)\right)\right] &= \sum_{t=1}^{T}\sum_{k=1}^{K}\mathbb{E}\left[\left(\tilde{x}^k_{v,t}, \bar{\ell}_{k,t} - \bar{\ell}_t(k^*)\right)\right]\\
&= \sum_{t=1}^{T}\sum_{k=1}^{K}\mathbb{E}[\tilde{x}^k_{v,t}]\Delta_k.
\end{aligned}$$

Using the Lemma 8, for any agent v we have

$$\frac{5}{3}R_T^{drift}(v) = \frac{5}{3}\sum_{t=1}^{T}\sum_{k=1}^{K}\mathbb{E}[\tilde{x}^k_{v,t}]\Delta_k \ge \sum_{t=1}^{T-D}\sum_{k=1}^{K}\mathbb{E}[x^k_{v,t+D}]\Delta_k$$

$$= \sum_{t=D+1}^{T} \sum_{k=1}^{K} \mathbb{E}[x_{v,t}^k]\Delta_k$$
$$\geq \sum_{t=1}^{T} \sum_{k=1}^{K} \mathbb{E}[x_{v,t}^k]\Delta_k - D = R_T(v) - D.$$

Where the second inequality uses $\sum_{t=1}^{D} \sum_{k=1}^{K} \mathbb{E}[x_{v,t}^k]\Delta_k \leq D$. As a result, we have $R_T(v) \leq \frac{5}{3} R_T^{drift}(v) + D$ and it suffices to upper bound $R_T^{drift}(v)$. As a consequence, the drifted pseudo-regret bound as follows:

$$\begin{aligned}
\sum_{v=1}^{V} R_T^{drift}(v) &= \sum_{v=1}^{V} \mathbb{E}\left[\sum_{t=1}^{T} \langle \bar{\ell}_t, \tilde{x}_{v,t}\rangle - \bar{\ell}_t(k^*)\right] \\
&= \mathbb{E}\left[\sum_{v=1}^{V}\sum_{t=1}^{T} \langle \bar{\ell}_t, \tilde{x}_{v,t}\rangle - \bar{\ell}_t(k^*)\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\sum_{v=1}^{V}\sum_{t=1}^{T} \left(\langle \mathbb{E}[m_t], \tilde{x}_{v,t}\rangle - \mathbb{E}[m_t(k^*)]\right)\right] \\
&\overset{(b)}{\leq} \mathbb{E}\left[\sum_{v=1}^{V}\sum_{t=1}^{T} \langle m_t, \tilde{x}_{v,t}\rangle - \hat{L}_{1,T+1}(k^*)\right] \\
&= \mathbb{E}\Bigg[\underbrace{\sum_{t=1}^{T}\sum_{v=1}^{V} \left(\bar{F}_t^*(-\hat{L}_{v+1,t}) - \bar{F}_t^*(-\hat{L}_{v,t}) + \langle \tilde{x}_{v,t}, m_t\rangle\right)}_{(A)} \\
&\quad + \underbrace{\left(\sum_{t=1}^{T}\sum_{v=1}^{V} \bar{F}_t^*(-\hat{L}_{v,t}) - \bar{F}_t^*(-\hat{L}_{v+1,t})\right) - \hat{L}_{1,T+1}(k^*)\Bigg]}_{(B)}.
\end{aligned}$$

Where (a) holds because the following facts for all arms k:

$$\bar{\ell}_t(k) = \frac{1}{V}\sum_{v=1}^{V} \ell_{v,t}(k) = \frac{1}{V}\sum_{v=1}^{V} x_{v,t}(k)\frac{\ell_{v,t}(k)}{x_{v,t}(k)} = \mathbb{E}[m_{v,t}(k)],$$

(b) holds because the following definition:

$$\hat{L}_{1,T+1}(k^*) = \sum_{t=1}^{T} V m_t(k^*) = \sum_{v=1}^{V}\sum_{t=1}^{T} m_t(k^*).$$

11.3.1 BOUNDING (A)

$$\begin{aligned}
&\sum_{v=1}^{V} \bar{F}_t^*(-\hat{L}_{v+1,t}) - \bar{F}_t^*(-\hat{L}_{v,t}) + \langle \tilde{x}_{v,t}, m_t\rangle \\
&= \sum_{v=1}^{V} \bar{F}_t^*(-\hat{L}_{v,t} - m_t) - \bar{F}_t^*(-\hat{L}_{v,t}) + \langle \tilde{x}_{v,t}, m_t\rangle \\
&= \sum_{v=1}^{V} \bar{F}_t^*(-\hat{L}_{v,t} - (m_t - \tilde{x}_{v,t}\odot m_t)) - \bar{F}_t^*(-\hat{L}_{v,t}) + \langle \tilde{x}_{v,t}, m_t - \tilde{x}_{v,t}\odot m_t\rangle \\
&\overset{(a)}{=} \sum_{v=1}^{V} \bar{F}_t^*(-\nabla\bar{F}_t(\tilde{x}_{v,t}) - (m_t - \tilde{x}_{v,t}\odot m_t)) - \bar{F}_t^*(-\nabla\bar{F}_t(x_{v,t})) + \langle \tilde{x}_{v,t}, m_t - \tilde{x}_{v,t}\odot m_t)\rangle
\end{aligned}$$

$$\overset{(b)}{\leq} \sum_{v=1}^{V} F_t^*(-\nabla \bar{F}_t(\tilde{x}_{v,t}) - (m_t - \tilde{x}_{v,t} \odot m_t)) - F_t^*(-\nabla \bar{F}_t(\tilde{x}_{v,t})) + \langle \tilde{x}_{v,t}, m_t - \tilde{x}_{v,t} \odot m_t \rangle$$

$$= \sum_{v=1}^{V} \sum_{k=1}^{K} D_{f_t^*} \left(f'(\tilde{x}_{v,t}) - (m_t(k) - \tilde{x}_{v,t}^k m_t(k)), f'(\tilde{x}_{v,t}) \right)$$

$$= \sum_{v=1}^{V} \sum_{k=1}^{K} D_{f_t^*} \left(f'(\tilde{x}_{v,t}) - \frac{1}{V} \sum_{u=1}^{V} \frac{\ell_{u,t}(1 - \tilde{x}_{v,t}^k)}{x_{u,t}^k}, f'(\tilde{x}_{v,t}) \right)$$

$$\leq \sum_{v=1}^{V} \sum_{k=1}^{K} D_{f_t^*} \left(f'(\tilde{x}_{v,t}) - \frac{1}{V} \sum_{u=1}^{V} \frac{1 - \tilde{x}_{v,t}^k}{x_{u,t}^k}, f'(\tilde{x}_{v,t}) \right)$$

$$\overset{(c)}{\leq} \sum_{v=1}^{V} \sum_{k=1}^{K} \frac{f_t''(\tilde{x}_{v,t}(k))^{-1}}{2V^2} \left(\sum_{u=1}^{V} \frac{1 - \tilde{x}_{v,t}^k}{x_{u,t}^k} \right)^2$$

$$\overset{(d)}{\leq} \sum_{v=1}^{V} \sum_{k=1}^{K} \frac{\tilde{x}_{v,t}(k)^{\frac{3}{2}}}{8V^2\sqrt{Vt}} \left(2V \frac{1 - \tilde{x}_{v,t}^k}{\tilde{x}_{v,t}^k} \right)^2$$

$$= \sum_{v=1}^{V} \sum_{k=1}^{K} \frac{(1 - \tilde{x}_{v,t}^k)^2}{2(\tilde{x}_{v,t}^k)^{\frac{1}{2}}\sqrt{Vt}}.$$

Where (a) applies Facts 2 and 3, the (b) follows from both parts of Fact 4, (c) uses Fact 6, and (d) uses the Lemma 9. In expectation we get

$$\mathbb{E}\left[\sum_{v=1}^{V} \left(\bar{F}_t^*(-\hat{L}_{v+1,t}) - \bar{F}_t^*(-\hat{L}_{v,t}) + \langle \tilde{x}_{v,t}, m_t \rangle \right) \right] \leq \sum_{v=1}^{V} \sum_{k=1}^{K} \frac{(1 - \tilde{x}_{v,t}^k)^2 (\tilde{x}_{v,t}^k)^{\frac{1}{2}}}{2\sqrt{Vt}}.$$

$$\leq \sum_{v=1}^{V} \sum_{k \neq k^*} \frac{(\tilde{x}_{v,t}^k)^{\frac{1}{2}}}{2\sqrt{Vt}} + \sum_{v=1}^{V} \frac{(1 - \tilde{x}_{v,t}(k^*))^2 (\tilde{x}_{v,t}(k^*))^{\frac{1}{2}}}{2\sqrt{Vt}}$$

$$\leq \sum_{v=1}^{V} \sum_{k \neq k^*} \frac{(\tilde{x}_{v,t}^k)^{\frac{1}{2}}}{\sqrt{Vt}}. \tag{31}$$

11.3.2 Bounding (B)

We have the following bound by Abernethy et al. (2015)

$$\left(\sum_{t=1}^{T} \sum_{v=1}^{V} \bar{F}_t^*(-\hat{L}_{v,t}) - \bar{F}_t^*(-\hat{L}_{v+1,t}) \right) - \hat{L}_{1,T+1}(k^*)$$
$$\leq \sum_{t=2}^{T} \sum_{v=1}^{V} ((F_{t-1}(\tilde{x}_{v,t})) - (F_{t-1}(\tilde{x}_{v+1,t}))) + F_T(x^*) - F_1(x_{1,1}).$$

By replacing the closed form of the regularizer in this bound and using the facts that $\eta_t^{-1} - \eta_{t-1}^{-1} \leq 2\eta_{t-1}$ and $\gamma_t^{-1} - \gamma_{t-1}^{-1} \leq 4C_t^P \gamma_{t-1} / \log(K)$ we obtain

$$\left(\sum_{t=1}^{T} \sum_{v=1}^{V} \bar{F}_t^*(-\hat{L}_{v,t}) - \bar{F}_t^*(-\hat{L}_{v+1,t}) \right) - \hat{L}_{1,T+1}(k^*)$$
$$\leq \sum_{t=2}^{T} \sum_{v=1}^{V} \sum_{k \neq k^*} \frac{(\tilde{x}_{v,t}^k)^{\frac{1}{2}}}{\sqrt{Vt}} + \sum_{t=2}^{T} \sum_{v=1}^{V} \sum_{k=1}^{K} \frac{2C_t^P \gamma_{t-1} \tilde{x}_{v,t}^k \log(1/\tilde{x}_{v,t}^k)}{\log(K)}$$
$$+ 13V\sqrt{D} + 6VD(K-1)^{\frac{1}{3}} \log(K) + 2VC_T^P \log(K). \tag{32}$$

Combine equation 31 and equation 32, we can get

$$\sum_{v=1}^{V} R_T^{drift}(v) \leq 2 \sum_{t=2}^{T} \sum_{v=1}^{V} \sum_{k \neq k^*} \frac{(\tilde{x}_{v,t}^k)^{\frac{1}{2}}}{\sqrt{Vt}} + \sum_{t=2}^{T} \sum_{v=1}^{V} \sum_{k=1}^{K} \frac{2C_t^P \gamma_{t-1} \tilde{x}_{v,t}^k \log(1/\tilde{x}_{v,t}^k)}{\log(K)}$$

$$+ 13V\sqrt{D} + 6VD(K-1)^{\frac{1}{3}}\log(K) + 2VC_T^P\log(K). \tag{33}$$

11.3.3 Self Bounding Analysis

We use the self-bounding technique to write $\sum_{v=1}^{V} R_T^{drift}(v) = 3\sum_{v=1}^{V} R_T^{drift}(v) - 2\sum_{v=1}^{V} R_T^{drift}(v)$, and then based on equation 33 we have

$$\begin{aligned}\sum_{v=1}^{V} R_T^{drift}(v) \le{}& 6\sum_{v=1}^{V}\sum_{k\neq k^*}\frac{(\tilde{x}_{v,t}^k)^{\frac{1}{2}}}{\sqrt{Vt}} - \sum_{v=1}^{V} R_T^{drift}(v)\\ &+ \sum_{t=2}^{T}\sum_{v=1}^{V}\sum_{k=1}^{K}\frac{6C_t^P\gamma_{t-1}\tilde{x}_{v,t}^k\log(1/\tilde{x}_{v,t}^k)}{\log(K)} - \sum_{v=1}^{V} R_T^{drift}(v)\\ &+ 13V\sqrt{D} + 6VD(K-1)^{\frac{1}{3}}\log(K).\end{aligned}$$

Here we give bound for the first term:

$$\begin{aligned}6\sum_{v=1}^{V}\sum_{k\neq k^*}\frac{(\tilde{x}_{v,t}^k)^{\frac{1}{2}}}{\sqrt{Vt}} - \sum_{v=1}^{V} R_T^{drift}(v) &= \sum_{t=1}^{T}\sum_{v=1}^{V}\sum_{k\neq k^*}\left(\frac{6(\tilde{x}_{v,t}^k)^{\frac{1}{2}}}{\sqrt{Vt}} - \tilde{x}_{v,t}^k\Delta_k\right)\\ &\le \sum_{t=1}^{T}\sum_{v=1}^{V}\sum_{k\neq k^*}\frac{36}{Vt\Delta_k} \le \sum_{k\neq k^*}\frac{36\log(T)}{\Delta_k}.\end{aligned}$$

where the first inequality uses $\forall x, y \ge 0 : x + y \ge 2\sqrt{xy} \Rightarrow 2\sqrt{xy} - y \le x$ so called AM-GM. According the proof of Lemma 8 in Masoudian et al. (2022), we can get bound for the second term:

$$\sum_{t=2}^{T}\sum_{v=1}^{V}\sum_{k=1}^{K}\frac{6C_t^P\gamma_{t-1}\tilde{x}_{v,t}^k\log(1/\tilde{x}_{v,t}^k)}{\log(K)} - \sum_{v=1}^{V} R_T^{drift}(v) \le \sum_{k\neq k^*}\frac{72VC_T^P}{\Delta_k\log(K)}.$$

In summary, we have

$$\begin{aligned}&\sum_{v=1}^{V} R_T(v) \le \frac{5}{3}\sum_{v=1}^{V} R_T^{drift}(v) + VD\\ &\le \sum_{k\neq k^*}\frac{60\log(T)}{\Delta_k} + \sum_{k\neq k^*}\frac{120VC_T^P}{\Delta_k\log(K)} + 22V\sqrt{D} + 10VD(K-1)^{\frac{1}{3}}\log(K) + 7VC_T^P\log(K).\end{aligned}$$

For any k, v, t, by Lemma 1, we can get the individual regret for each agent v:

$$\begin{aligned}&R_T(v) \le \frac{3}{2V}\sum_{v=1}^{V} R_T(v)\\ &\le \sum_{k\neq k^*}\frac{90\log(T)}{V\Delta_k} + \sum_{k\neq k^*}\frac{180C_T^P}{\Delta_k\log(K)} + 33\sqrt{D} + 15D(K-1)^{\frac{1}{3}}\log(K) + 11C_T^P\log(K).\end{aligned}$$

11.4 Proof for Communication Cost

For each agent v, let $trunc\_round(v)$ denote the number of rounds in which a new deviation record is generated (i.e., loss truncation is triggered), and let $comm\_cost(v,t)$ denote the communication size at round t. Recalling the Algorithm 1, at round t, the probability that an agent v generates a new deviation record is $x_{v,t}(k_{v,t}) \le 12VC_t^P\gamma_t$. So we can get

$$\mathbb{E}[trunc\_round(v)] \le \sum_{t=1}^{T} 12VC_t^P \le \sum_{t=1}^{T}\frac{12VC_t^P\gamma_t}{8V\sqrt{C_t^P t/\log(K)}} \le 3\sqrt{C_T^P T\log(K)}.$$

So we can guarantee that $Truncated\_rounds(v) = O(\sqrt{T})$ holds for any agent v. At any round t, the message sent by an agent v consists of two parts, $\hat{L}_{v,t}^{obs}$ and A_v. The size of $\hat{L}_{v,t}^{obs}$ is $O(K)$, and the size of A_v, is at most the number of deviation records generated by all agents during the interval $t - D < s \leq t$. Thus, combining this with the probability of generating deviation records, we obtain:

$$\begin{aligned}\mathbb{E}[Comm\_size(v,t)] &= O(K) + O\left(\sum_{s=t-D+1}^{t}\sum_{i=1}^{V} i(12VC_t^P\gamma_t)^i\right)\\ &= O(K) + O\left(\sum_{i=1}^{VD} i(12VC_t^P\gamma_t)^i\right)\\ &= O(K) + O\left(\sum_{i=1}^{\infty} i(12VC_t^P\gamma_t)^i\right)\\ &= O(K + \frac{12VC_t^P\gamma_t}{(1-12VC_t^P\gamma_t)^2}).\end{aligned}$$

Where the last equality comes form the inequality:

$$\sum_{i=1}^{\infty} ia^i = \frac{a}{(1-a)^2}.$$

Since we have

$$12VC_t^P\gamma_t = \frac{12VC_t^P}{8V\sqrt{C_t^P t/\log(K) + 144D^2(K-1)^{\frac{2}{3}} + 4(C_t^P)^2}} \leq 0.75$$

Then

$$\mathbb{E}[comm\_cost(v,t)] = O(K+12) = O(K).$$

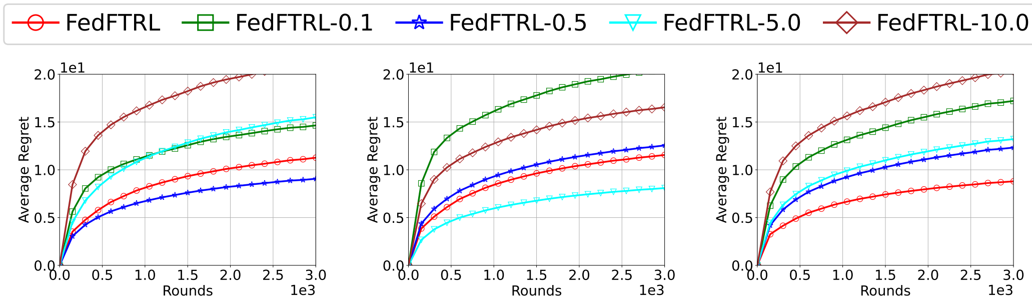

Figure 3: Average cumulative regret for FedFTRL, FedFTRL-0.1, FedFTRL-0.5, FedFTRL-5.0 and FedFTRL-10.0 in the synthetic dataest, under three different communication networks: (left) complete graph, (middle) grid graph, and (right) RGG-0.5.

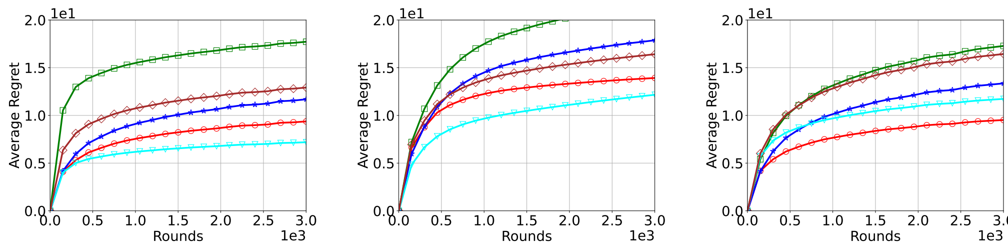

Figure 4: Average cumulative regret for FedFTRL, FedFTRL-0.1, FedFTRL-0.5, FedFTRL-5.0 and FedFTRL-10.0 in the MovieLens dataest, under three different communication networks: (left) complete graph, (middle) grid graph, and (right) RGG-0.5.

12 Supplementary Experiments

12.1 Sensitivity of C_t^P

In this section, we conduct experiments to investigate the sensitivity of the topology parameter C_t^P. We keep the experimental setup identical to that in Section 6 and only rescale C_t^P by factors $0.1, 0.5, 1.0$ (default), 5, and 10. We denote the corresponding variants by FedFTRL-ε, where ε is the scaling factor. When $\varepsilon = 1.0$, C_t^P is unchanged and we simply write FedFTRL. All experiments are repeated for 50 trials, and we report the averaged performance as plotted curves.

The results in Figure 3 and Figure 4 show that our FedFTRL algorithm is robust to the choice of the topology parameter C_t^P. Even with a misspecified C_T^P, our algorithms still achieve sublinear regret.