Proceedings of Machine Learning Research vol 336:1–53, 2026 39th Annual Conference on Learning Theory

Near-Optimal Regret for Distributed Adversarial Bandits: A Black-Box Approach

Hao Qiu[*] QIUHAOSAI@GMAIL.COM
Università degli Studi di Milano

Mengxiao Zhang[*] MENGXIAO-ZHANG@UIOWA.EDU
University of Iowa

Nicolò Cesa-Bianchi NICOLO.CESA-BIANCHI@UNIMI.IT
Università degli Studi di Milano

Editors: Steve Hanneke and Tor Lattimore

Abstract

We study distributed adversarial bandits, where N agents cooperate to minimize the global average loss while observing only their own local losses. We show that the minimax regret for this problem is $\widetilde{\Theta}\Big(\sqrt{\big(\rho^{-1/2} + \frac{K}{N}\big)\,T}\Big)$, where T is the horizon, K is the number of actions, and ρ is the spectral gap of the communication matrix. Our algorithm, based on a novel black-box reduction to bandits with delayed feedback, requires agents to communicate only through gossip. It achieves an upper bound that significantly improves over the previous best bound $\widetilde{\mathcal{O}}\left(\rho^{-1/3}(KT)^{2/3}\right)$ of Yi and Vojnovic (2023). We complement this result with a matching lower bound, showing that the problem's difficulty decomposes into a communication cost $\rho^{-1/4}\sqrt{T}$ and a bandit cost $\sqrt{KT/N}$. We further demonstrate the versatility of our approach by deriving first-order and best-of-both-worlds bounds in the distributed adversarial setting. Finally, we extend our framework to distributed linear bandits in $\mathbb{R}^d$, obtaining a regret bound of $\widetilde{\mathcal{O}}\Big(\sqrt{\big(\rho^{-1/2} + \frac{1}{N}\big)\,dT}\Big)$, achieved with only $\mathcal{O}(d)$ communication cost per agent and per round via a volumetric spanner.

Keywords: distributed bandits, bandits with delayed feedback, linear bandits.

1. Introduction

Distributed online optimization (Yan et al., 2013; Hosseini et al., 2013; Wan et al., 2024) studies how multiple agents can cooperate to minimize regret in a decentralized fashion. The core challenge in this setting is the misalignment between the global objective and the local information: agents aim to minimize the average loss across the entire network, yet they only observe their own local loss functions. To bridge this gap, agents must exchange information over a communication network, typically interacting only with their immediate neighbors via gossip protocols (Xiao and Boyd, 2004; Boyd et al., 2006; Liu and Morse, 2011). This framework is naturally motivated by applications in federated learning with heterogeneous clients (Blaser et al., 2024) and decentralized collaborative learning (Zantedeschi et al., 2020). In this work, we investigate distributed online optimization under the challenging regime of bandit feedback: unlike full information settings, where agents observe the local losses for all actions, here agents only observe the local loss of their chosen action at each round, which they must propagate to optimize the global objective.

[*] Equal contribution, in alphabetical order.

In distributed online convex optimization (D-OCO), the regret scales with the connectivity of the network, formalized by the spectral gap ρ of the gossip matrix. A small ρ indicates network bottlenecks, which typically slow down the propagation of local losses. In the full information setting, the cost of this decentralization is well understood: the pioneering work of Yan et al. (2013); Hosseini et al. (2013) extends classical OCO algorithms to the distributed online setting, achieving an $\mathcal{O}(N^{1/4}\rho^{-1/2}\sqrt{T})$ regret, where N is the number of agents and T is the horizon. Building on this, Wan et al. (2024) utilized an accelerated variant of gossip to establish regret bounds of $\widetilde{\mathcal{O}}(\rho^{-1/4}\sqrt{T})$, which are optimal up to logarithmic factors.[1] However, for distributed K-armed bandits, the picture is less clear. The current state-of-the-art (Yi and Vojnovic, 2023) achieves $\widetilde{\mathcal{O}}(\rho^{-1/3}(KT)^{2/3})$ regret, which is significantly worse than the $\sqrt{T}$ rate achieved in the full-information case. Moreover, the bound of Yi and Vojnovic (2023) does not reveal how the regret should depend on the number of agents N. Seldin et al. (2014) show that K-armed bandits with N observations per round have minimax regret $\Theta(\sqrt{KT/N})$, suggesting that multiple bandit observations can in principle lead to improved rates. Although distributed K-armed bandits differ from this setting, they also generate N local observations per round. These contrasts motivate the following question:

What is the minimax regret rate for distributed K-armed bandits?

Contributions and technical challenges. In this work, we answer this question by providing a regret upper bound of $\widetilde{\mathcal{O}}\big(\sqrt{\big(\rho^{-1/2}+\frac{K}{N}\big)\,T}\big)$, thus significantly improving upon the previous $\widetilde{\mathcal{O}}\left(\rho^{-1/3}(KT)^{2/3}\right)$ bound of Yi and Vojnovic (2023). We complement this with a matching lower bound, characterizing the minimax rate as a combination of a *communication cost* of order $\rho^{-1/4}\sqrt{T}$ and a *bandit information cost* of order $\sqrt{KT/N}$. A key technical obstacle—leading to the $T^{2/3}$-type behavior in existing analyses—is the seemingly natural *round-by-round* update paradigm: each agent updates immediately from its local importance-weighted bandit estimate while gossiping. Due to the high variance of these estimates, the information exchanged through local gossip cannot be sufficiently mixed within a single round. Consequently, the local updates made by each agent are not close to those based on the global average loss estimate, causing a large error term in the regret. We resolve this obstacle by explicitly *decoupling* learning from communication via a block-based approach that intentionally introduces delayed feedback. Specifically, agents freeze their action distributions and buffer local bandit observations for the duration of a block. Within this block, they execute accelerated gossip iterations to ensure that, when the next block starts, every agent possesses a high-precision approximation of the global average loss from the previous block. Crucially, we incorporate a small $\Theta(1/T)$ uniform exploration term to ensure that the importance weights remain uniformly bounded. This allows the gossip protocol to achieve an accurate approximation of the global average loss estimate in only $\mathcal{O}\big(\rho^{-1/2}\log(KTN)\big)$ rounds. Ultimately, this construction yields a black-box reduction from distributed bandits with local gossip to standard adversarial bandits with delayed feedback (Cesa-Bianchi et al., 2019; Thune et al., 2019; Bistritz et al., 2019; Gyorgy and Joulani, 2021; van der Hoeven et al., 2023). Furthermore, because our reduction is algorithm-agnostic, instantiating it with adaptive delayed-bandit algorithms (Van Der Hoeven and Cesa-Bianchi, 2022; Masoudian et al., 2022) immediately confers adaptive guarantees to the distributed setting, including small-loss and best-of-both-worlds bounds. Finally, we extend our framework to distributed d-dimensional linear bandits with K actions, a setting that—to the best of our knowledge—has not been previously studied. Specifically, we provide a $\widetilde{\mathcal{O}}\big(\sqrt{\big(\rho^{-1/2}+1/N\big)dT}\big)$

1. We use $\widetilde{\mathcal{O}}(\cdot)$ and $\widetilde{\Theta}(\cdot)$ to suppress logarithmic factors in T, N, K.

upper bound and a $\Omega\big(\sqrt{(\rho^{-1/2} + d/N)T}\big)$ lower bound. Crucially, we achieve this using only gossip communication with $\mathcal{O}(d)$-sized messages per round, matching the communication efficiency of established D-OCO protocols (Yan et al., 2013; Hosseini et al., 2013; Wan et al., 2024; Wan, 2025).

1.1. Related works

Distributed optimization and gossiping. Our framework builds on the established line of work on gossip protocols (Xiao and Boyd, 2004; Boyd et al., 2006). Gossip is used to aggregate information over a communication network when each agent can only exchange information with its neighbors. Due to their simplicity, gossip algorithms have become a standard technique in distributed optimization. Early work on distributed convex optimization leveraged gossip (peer-to-peer) communication to achieve consensus and optimize a global objective. In classic setups, each agent holds a local convex cost and they seek to minimize the sum of costs without a central coordinator. For example, Nedic and Ozdaglar (2009) developed a distributed method where agents repeatedly take gradient steps on local functions and gossip their updates with neighbors, guaranteeing convergence to the global optimum. Building on such consensus ideas, Duchi et al. (2011) proposed a distributed dual averaging algorithm using gossip at each iteration to share gradient information. This method provided rigorous convergence rate bounds in terms of the network spectral properties (e.g., the second-largest eigenvalue of the gossip matrix). In distributed optimization Scaman et al. (2019) identified optimal algorithms and matching lower bounds for decentralized convex optimization, including the convex and strongly convex.

Distributed online convex optimization. Distributed online convex optimization (D-OCO) studies a network of agents that sequentially make decisions and incur convex losses, with the objective of collectively minimizing regret with respect to the global average loss function. In the full-information setting, each agent observes the entire loss function (or gradient) after making their decision, allowing the use of powerful online learning algorithms in a decentralized manner. The pioneering work of Yan et al. (2013); Hosseini et al. (2013) extend OGD and dual average into the D-OCO setting and achieve $\mathcal{O}(N^{1/4}\rho^{-1/2}\sqrt{T})$ regret bound. Nearly optimal regret bounds $\widetilde{\mathcal{O}}(\rho^{-1/4}\sqrt{T})$ for convex functions have been established by Wan et al. (2024) using an accelerated gossip strategy, see also Liu and Morse (2011); Ye et al. (2023b). They also provide $\Omega(\rho^{-1/4}\sqrt{T})$ lower regret bounds for convex loss functions. For comprehensive surveys on D-OCO, we refer the readers to Li et al. (2023) and Yuan et al. (2024).

Distributed K-armed bandits. The problem becomes significantly more challenging under bandit feedback, where each agent observes only the scalar loss of their chosen action rather than the full loss vector. Prior literature has investigated this setting in both the stochastic environment, where losses are drawn i.i.d. from fixed distributions, and the adversarial environment, where the loss sequences can be chosen arbitrarily. In a stochastic setting, previous works Zhu et al. (2021); Zhu and Liu (2023); Xu and Klabjan (2023); Zhang et al. (2025); Liu et al. (2025) study distributed K-armed bandits where agents run a gossip protocol *augmented with additional communication*, achieving regret guarantees of $\mathcal{O}(\log T)$. The closest work to ours is Yi and Vojnovic (2023), which studies distributed K-armed bandits using only a gossip protocol in the adversarial setting, and proves a suboptimal regret bound of $\widetilde{\mathcal{O}}\left(\rho^{-1/3}(KT)^{2/3}\right)$. They also prove a lower bound for a specific family of communication graphs and for an arbitrary gossip matrix supported on these graphs. Our lower bound construction in Theorem 8 is an adaptation of theirs.

Cooperative adversarial bandits. Our work is also closely related to the literature on *cooperative adversarial bandits* (Cesa-Bianchi et al., 2019; Bar-On and Mansour, 2019; Ito et al., 2020). This setting can be viewed as a special case of distributed bandits where the local loss functions are identical for all agents at each time step. However, a key distinction lies in the communication protocol. While we rely on a gossip protocol, Cesa-Bianchi et al. (2019); Bar-On and Mansour (2019) employ a message-sending protocol, in which agents broadcast their local information rather than mixing information through gossip. Because the local loss functions are identical across agents, each agent can achieve $\mathcal{O}(\sqrt{T})$ regret even without communication, whereas $\Omega(T)$ regret would be unavoidable in the distributed bandits problem.

2. Preliminaries

For a positive integer K, we use the notation $[K] \triangleq \{1, 2, \dots, K\}$. For a finite set S, we denote with $\Delta(S)$ the probability simplex over S, defined as $\Delta(S) \triangleq \big\{\boldsymbol{p} \in \mathbb{R}^{|S|} : \sum_{k=1}^{|S|} p(k) = 1, p(k) \geqslant 0, \forall k \in [|S|]\big\}$. If $S = [K]$, then we write $\Delta(K)$ instead of $\Delta([K])$. We use boldface to denote vectors $\boldsymbol{u}, \boldsymbol{v}(i)$ and $u(j), v(i,j)$ to denote their components. We let $\mathbf{1}$ to be the all-one vector in an appropriate dimension and, given a set S, we let $\mathbf{1}_S$ to be the vector with entries $\mathbf{1}_S(i) = 1$ if $i \in S$ and $\mathbf{1}_S(i) = 0$ otherwise. Agents are the vertices $V = [N]$ of an undirected, connected communication graph $G = (V, E)$ with neighbors $\mathcal{N}(i) \triangleq \{j \in V : (i,j) \in E\} \cup \{i\}$ for $i \in V$.

Distributed K-armed bandits. We consider an oblivious adversary, specifying vectors $\boldsymbol{\ell}_t(i) \in [0,1]^K$ denoting the local loss of agent $i \in V$ at time t. At each round $t \in [T]$, each agent $i \in V$ maintains a distribution $\boldsymbol{p}_t(i) \in \Delta(K)$ and samples an action $A_t(i) \sim \boldsymbol{p}_t(i) \in [K]$. After playing $A_t(i)$, each agent i observes its realized local loss $\ell_t\big(i, A_t(i)\big)$. Define the global average loss at time t by $\bar{\boldsymbol{\ell}}_t \triangleq \frac{1}{N}\sum_{j=1}^N \boldsymbol{\ell}_t(j)$, so that $\bar{\ell}_t(k) = \frac{1}{N}\sum_{j=1}^N \ell_t(j,k)$. Then the expected global average loss of agent $i \in V$ at round t is $\mathbb{E}\left[\bar{\ell}_t\big(A_t(i)\big)\right] = \big\langle \boldsymbol{p}_t(i), \bar{\boldsymbol{\ell}}_t \big\rangle$.

After observing local feedback, each agent i may communicate with neighbors in $\mathcal{N}(i)$—according to the gossip protocol—and perform the update $\boldsymbol{p}_t(i) \to \boldsymbol{p}_{t+1}(i)$. The performance of agent $i \in V$ is measured by the (pseudo) regret

$$\mathrm{Reg}_T(i) \triangleq \mathbb{E}\left[\sum_{t=1}^T \bar{\ell}_t\big(A_t(i)\big)\right] - \min_{k\in[K]} \mathbb{E}\left[\sum_{t=1}^T \bar{\ell}_t(k)\right] = \max_{k\in[K]} \mathbb{E}\left[\sum_{t=1}^T \big\langle \boldsymbol{p}_t(i) - \boldsymbol{e}_k, \bar{\boldsymbol{\ell}}_t \big\rangle\right], \qquad (1)$$

where $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_K\}$ is the canonical basis of $\mathbb{R}^K$ and the expectation is with respect to the agents' internal randomization. We also define $\mathrm{Reg}_T \triangleq \max_{i\in V} \mathrm{Reg}_T(i)$. Without loss of generality, we assume $T \geqslant 3$ and $K \geqslant 2$.

Distributed linear bandits. In the linear bandits case, agents have a common set of K actions with feature vectors $\Omega = \{\boldsymbol{a}_1, \dots, \boldsymbol{a}_K\} \subset \mathbb{R}^d$. The oblivious adversary specifies local loss coefficients $\boldsymbol{\theta}_t(i) \in \mathbb{R}^d$ such that for any agent $i \in V$, action $\boldsymbol{a}_k \in \Omega$, and round $t \in [T]$, the linear local loss satisfies $\ell_t(i,k) = \langle \boldsymbol{\theta}_t(i), \boldsymbol{a}_k \rangle \in [-1, 1]$. At each round $t \in [T]$, each agent $i \in V$ maintains a distribution $\boldsymbol{p}_t(i) \in \Delta(K)$ and samples an action index $A_t(i) \sim \boldsymbol{p}_t(i) \in [K]$. After playing $A_t(i)$, agent i observes its realized local loss $\ell_t\big(i, A_t(i)\big)$. The regret is still defined by Eq. (1), with $\bar{\boldsymbol{\ell}}_t$ formed from the linear losses: $\bar{\ell}_t(k) = \frac{1}{N}\sum_{i=1}^N \langle \boldsymbol{\theta}_t(i), \boldsymbol{a}_k \rangle = \big\langle \frac{1}{N}\sum_{i=1}^N \boldsymbol{\theta}_t(i), \boldsymbol{a}_k \big\rangle$.

Algorithm 1 Reduction from distributed bandits to bandits with delay

Input: Agent index i, time horizon T, block length B, mixing coefficient κ, gossip matrix W, delayed bandit algorithm $\mathcal{A}$.
Initialization: Set $\boldsymbol{z}_1^{-1}(i) = \boldsymbol{z}_1^0(i) = \mathbf{0} \in \mathbb{R}^K$, exploration parameter $\alpha = \frac{1}{T}$.
for $\tau = 1, \ldots, T/B$ **do**
  Receive $\boldsymbol{p}'_\tau(i)$ from $\mathcal{A}$ and compute $\boldsymbol{p}_\tau(i) = (1-\alpha)\,\boldsymbol{p}'_\tau(i) + \frac{\alpha}{K}\mathbf{1}$
  Set $b = 0$
  for $t = (\tau-1)B+1, \ldots, \tau B$ **do**
    Play $A_t(i) \sim \boldsymbol{p}_\tau(i)$ and observe $\ell_t(i, A_t(i))$
    Compute the estimate $\widehat{\boldsymbol{\ell}}_t(i)$: for each $k \in [K]$, $\widehat{\ell}_t(i,k) = \dfrac{\ell_t(i,k)}{p_\tau(i,k)}\mathbb{1}\{A_t(i) = k\}$
    Update $\boldsymbol{z}_\tau^{b+1}(i) = (1+\kappa)\sum_{j\in\mathcal{N}(i)} W(i,j)\boldsymbol{z}_\tau^b(j) - \kappa \boldsymbol{z}_\tau^{b-1}(i)$ and set $b \leftarrow b+1$
  end
  if $\tau \geqslant 2$ **then** send $\big(\tau-1, \boldsymbol{z}_\tau^B(i)\big)$ to $\mathcal{A}$;
  Set $\boldsymbol{z}_{\tau+1}^{-1}(i) = \boldsymbol{z}_{\tau+1}^0(i) = \sum_{t=(\tau-1)B+1}^{\tau B} \widehat{\boldsymbol{\ell}}_t(i)$
end

Gossip protocol. We assume that agents exchange information *only* through a gossip protocol (Yan et al., 2013; Hosseini et al., 2013; Yi and Vojnovic, 2023; Wan et al., 2024). A *gossip matrix* is any symmetric and doubly stochastic matrix $W \in \mathbb{R}^{N\times N}$ supported on the graph G, i.e., $W(i,j) > 0$ only if $(i,j) \in E$ or $i = j$ and $\sum_i W(i,j) = \sum_j W(i,j) = 1$. We use $0 \leqslant \sigma_2(W) < 1$ to denote the second largest singular value of W (as W is doubly stochastic, the largest singular value is 1). The *spectral gap* is $\rho(W) \triangleq 1 - \sigma_2(W)$; when the dependence on W is clear, we write ρ. In each gossip step, every agent $i \in V$ sends a message to its neighbors $\mathcal{N}(i)$ on the communication network G and updates its local state by forming a weighted average using a common gossip matrix W supported on G. In the K-armed setting, messages are K-dimensional; in the linear bandit setting, agents exchange $\mathcal{O}(d)$-dimensional vectors. As in prior work, W is a parameter shared among agents and not individually learned. To describe gossip more formally, suppose each agent $i \in V$ holds a vector $\boldsymbol{x}(i) \in \mathbb{R}^K$ (or $\mathbb{R}^{\mathcal{O}(d)}$ in the linear case) and the goal of the agents is to compute an approximation of the global average $\bar{\boldsymbol{x}} \triangleq \frac{1}{N}\sum_{i\in V}\boldsymbol{x}(i)$. Depending on the learning problem, $\boldsymbol{x}(i)$ may represent, e.g., (local) loss gradients (Hosseini et al., 2013; Wan et al., 2024) or loss estimates (Yi and Vojnovic, 2023). To do so, agents run B rounds of (accelerated) gossip (Liu and Morse, 2011) where the update is defined by

$$\boldsymbol{x}^{b+1}(i) = (1+\kappa)\sum_{j\in\mathcal{N}(i)} W(i,j)\boldsymbol{x}^b(j) - \kappa\boldsymbol{x}^{b-1}(i) \qquad b \geqslant 0, \tag{2}$$

with initialization $\boldsymbol{x}^{-1}(i) = \boldsymbol{x}^0(i) = \boldsymbol{x}(i)$ for all $i \in V$, and mixing coefficient $\kappa \geqslant 0$. When $\kappa = 0$, this reduces to standard gossip (Xiao and Boyd, 2004): $\boldsymbol{x}^{b+1}(i) = \sum_{j\in\mathcal{N}(i)} W(i,j)\boldsymbol{x}^b(j)$.

3. Distributed K-armed Bandits: A Black-Box Reduction to Delayed Feedback

In this section, we present a black-box reduction that transforms a distributed K-armed bandit problem into a K-armed bandit problem with delayed feedback. The reduction, stated in Algorithm 1, can be combined with any bandit algorithm $\mathcal{A}$ that enjoys a regret guarantee under delayed feedback.

A central difficulty in distributed adversarial bandits is that agents observe only noisy estimates of their local losses. If agents attempt to (i) update their policies every round and simultaneously (ii) aggregate these noisy estimates via insufficient gossip steps, the resulting local approximations remain far from the true global average loss estimate, leading to large regret. Previous work by Yi and Vojnovic (2023) attempts to mitigate this by maintaining explicit $\Omega(T^{-1/3})$ uniform exploration to control variance, but this comes at the cost of a suboptimal $T^{2/3}$ regret rate. Algorithm 1 resolves this issue by *decoupling learning from communication*:

- **Blocking (allowing time for mixing).** Inspired by (Garber and Kretzu, 2020; Wan et al., 2024), we introduce a delay of one block to ensure sufficient information mixing across the network. By keeping the action distribution fixed for a block of length B (to be defined later) and executing B rounds of accelerated gossip within each block, we ensure that the local approximation of global average loss estimate fed to $\mathcal{A}$ at the end of the block is close enough to the true global average estimate.
- **Small uniform exploration (bounded variance).** We mix in a $\frac{1}{T}$-amount of uniform exploration to ensure $p_\tau(i,k) \geqslant 1/(KT)$. This guarantees that the importance-weighted estimates are uniformly bounded by KT, which allows the gossip protocol to control the consensus error uniformly across all blocks without degrading the regret rate.

More specifically, we partition the time horizon T into T/B blocks indexed by τ, where $\mathcal{T}_\tau \triangleq \{(\tau-1)B+1,\dots,\tau B\}$ denotes the time steps in block τ. The algorithm operates iteratively over these blocks as follows:

- First, at the beginning of block τ, each agent i queries its local instance of the delayed bandit algorithm $\mathcal{A}$ to obtain a base policy $\boldsymbol{p}'_\tau(i)$. It then mixes in a $\frac{1}{T}$-amount of uniform exploration (see Line 4) to form the strategy $\boldsymbol{p}_\tau(i)$, which is then played for all time steps $t \in \mathcal{T}_\tau$.
- Second, while playing these actions, the agent performs two parallel tasks: (i) it computes the sum $\sum_{t\in\mathcal{T}_\tau} \widehat{\boldsymbol{\ell}}_t(i)$ of its local loss estimates for the *current* block; and (ii) it runs B rounds of accelerated gossip (see Line 9) to mix the estimators collected in the *previous* block, $\tau-1$.
- Finally, at the end of block τ (for $\tau \geqslant 2$), the result of this gossip process, denoted $\boldsymbol{z}_\tau^B(i)$, is fed back to $\mathcal{A}$. This vector serves as a local approximation of the global average loss estimate $\bar{\boldsymbol{z}}_\tau$ for block $\tau-1$, where

$$\bar{\boldsymbol{z}}_\tau \triangleq \frac{1}{N}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{\tau-1}} \widehat{\boldsymbol{\ell}}_t(i), \qquad \tau \geqslant 2\,. \tag{3}$$

Since the information from block $\tau-1$ is processed during block τ and delivered only at its conclusion, the learner effectively operates with a constant delay of one block.

As our reduction is agnostic to the choice of the delayed bandit algorithm, we choose a general approach and instantiate $\mathcal{A}$ as Algorithm 2 using the technique of Joulani et al. (2013), which converts any non-delayed OLO algorithm $\mathcal{B}$ into an OLO algorithm robust to delay. While this framework typically requires maintaining $d_{\max}$ parallel instances to handle a maximum delay of $d_{\max}$, our reduction imposes a fixed delay of only one block. Consequently, Algorithm 2 needs to instantiate only two independent copies of $\mathcal{B}$ (alternating between even and odd blocks), so each instance effectively observes non-delayed feedback. Specifically, we employ Follow-the-Regularized-Leader (FTRL) as the base learner $\mathcal{B}$ (Algorithm 3). While in this section we instantiate ψ with the negative entropy regularizer, in the next section we derive adaptive regret bounds using different regularizers.

Algorithm 2 Reduction from delayed OLO to OLO

Input: Agent index $i \in V$, time horizon T, base OLO algorithm $\mathcal{B}$
Initialization: Instances $\mathcal{B}^{(0)}(i), \mathcal{B}^{(1)}(i)$ of $\mathcal{B}$ with initial distributions $\boldsymbol{q}_1^{(0)}(i), \boldsymbol{q}_1^{(1)}(i) \in \Delta(K)$.
for $\tau = 1, \cdots, T/B$ **do**
  Query $\mathcal{B}^{(\tau \bmod 2)}(i)$ and receive $\boldsymbol{q}_{\lfloor(\tau+1)/2\rfloor}^{(\tau \bmod 2)}(i)$
  Send $\boldsymbol{p}'_\tau(i) = \boldsymbol{q}_{\lfloor(\tau+1)/2\rfloor}^{(\tau \bmod 2)}(i)$ to Algorithm 1
  if $\tau \geqslant 2$ **then**
    Receive $\big(\tau - 1, \boldsymbol{z}_\tau^B(i)\big)$ from Algorithm 1 and send $\boldsymbol{z}_\tau^B(i)$ to $\mathcal{B}^{(\tau-1 \bmod 2)}(i)$
  end
end

Algorithm 3 Follow the Regularized Leader for MAB

Input: regularizer ψ_t for $t \in [T]$
Initialize: $\boldsymbol{q}_1 = \frac{1}{K}\mathbf{1}$
for $t = 1, \ldots, T$ **do**
  Output $\boldsymbol{q}_t$ and receive $\boldsymbol{z}_t$
  Update $\boldsymbol{q}_{t+1} = \arg\min_{\boldsymbol{q} \in \Delta(K)} \Big\{ \sum_{s \in [t]} \langle \boldsymbol{z}_s, \boldsymbol{q} \rangle + \psi_t(\boldsymbol{q}) \Big\}$
end

It remains to specify the block length B and the mixing coefficient κ. Our choice is driven by the convergence properties of the accelerated gossip protocol (Ye et al., 2023a, Proposition 11), included as Lemma 11 for completeness. That result shows that, for an appropriate choice of κ, the approximation error to the global average loss estimate shrinks exponentially fast with the number of gossip rounds B, scaled by the norm of the initial gap to the global average loss estimate. Note that the $\frac{1}{T}$ uniform exploration ensures that the importance-weighted loss estimates have an ℓ_2-norm bounded by KT. Consequently, we can drive an initial error of magnitude $\mathcal{O}(KT)$ down to $\operatorname{poly}\big((KT)^{-1}\big)$ with a number of rounds logarithmic in KT. B and κ are thus set as follows:

$$B = \left\lceil \frac{\ln\big((KT)^6\sqrt{14N}\big)}{(1-1/\sqrt{2})\sqrt{1-\sigma_2(W)}} \right\rceil \qquad \text{and} \qquad \kappa = \frac{1}{1+\sqrt{1-\sigma_2^2(W)}}\,. \tag{4}$$

With these choices, the following lemma shows that the gossip process leads to a local approximation of global average loss estimate with only a $\mathcal{O}((KT)^{-5})$ error. The full proof is deferred to Appendix B.1.

Lemma 1 *If all agents $i \in V$ run Algorithm 1 with gossip matrix W and parameters κ, B defined in Eq. (4), then*

$$\max_{i \in V} \max_{\tau \in [T/B]} \big\| \boldsymbol{z}_\tau^B(i) - \bar{\boldsymbol{z}}_\tau \big\|_2 \leqslant \frac{2B}{(KT)^5}, \tag{5}$$

where $\bar{\boldsymbol{z}}_\tau \triangleq \frac{1}{N}\sum_{i=1}^N \sum_{t \in \mathcal{T}_{\tau-1}} \widehat{\boldsymbol{\ell}}_t(i)$ is defined in Eq. (3) for all $\tau \in [T/B]$.

We now present our main result for distributed K-armed bandits. The following theorem shows that our algorithm achieves a regret bound of $\widetilde{\mathcal{O}}(\sqrt{(\rho^{-1/2} + K/N)T})$.

Theorem 2 *Let $\mathcal{A}$ be Algorithm 2 run with base OLO algorithm $\mathcal{B}$ set to Algorithm 3 with $\psi_t(\boldsymbol{q}) = \frac{1}{\eta}\sum_{k=1}^{K} q(k)\log q(k)$ for all $t \in [T]$ where $\eta = \sqrt{\frac{\log K}{2(B+\frac{3K}{N})T}}$. Assume all agents $i \in V$ run Algorithm 1 with bandit algorithm $\mathcal{A}$, gossip matrix W, and parameters κ and B defined in Eq. (4). Then the regret is bounded as*

$$\mathrm{Reg}_T = 2\sqrt{2(\log K)\left(B + \frac{3K}{N}\right)T} + 10 = \widetilde{\mathcal{O}}\left(\sqrt{\left(\rho^{-1/2} + \frac{K}{N}\right)T}\right) .$$

Theorem 2 significantly improves over the state-of-the-art by Yi and Vojnovic (2023), which obtained a suboptimal regret of $\widetilde{\mathcal{O}}\big(\rho^{-1/3}(KT)^{2/3}\big)$. Moreover, we show that our bound is minimax optimal up to logarithmic factors, as we provide a matching lower bound in Theorem 8. To gain a better understanding of this result, note that our regret guarantee decomposes into two distinct terms: a *communication cost* of $\widetilde{\mathcal{O}}\big(\sqrt{\rho^{-1/2}T}\big)$, which scales with the block length B (and thus inversely with the spectral gap), and a *bandit information cost* of $\widetilde{\mathcal{O}}\big(\sqrt{KT/N}\big)$, accounting for the acceleration provided by N cooperating agents. The full proof is deferred to Appendix B.

Proof Sketch of Theorem 2. The proof follows by combining two key ingredients. The first ingredient is the lemma showing that if, for all $i \in V$, the instance of $\mathcal{A}$ run by i enjoys a bounded regret with respect to the one-block-delayed estimates $\boldsymbol{z}_\tau^B(i)$, then the regret of each agent with respect to the true global average loss is also bounded, up to a constant additive overhead. Formally, this reduction is stated as follows:

Lemma 3 *Suppose that every agent $i \in V$ runs an instance of Algorithm 1 using a base algorithm $\mathcal{A}$ that guarantees*

$$\max_{i\in V}\max_{k\in[K]} \mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \boldsymbol{z}_{\tau+1}^B(i)\right\rangle\right] \leqslant \mathsf{R}^{\mathrm{del}} , \tag{6}$$

for some $\mathsf{R}^{\mathrm{del}} \geqslant 0$, then we have $\mathrm{Reg}_T \leqslant \mathsf{R}^{\mathrm{del}} + 4$.

The full proof of Lemma 3 is deferred to Appendix B.2, here we only sketch the main argument. A direct calculation shows that the overall regret is bounded by the regret of $\mathcal{A}$ (the left-hand side of Eq. (6)) plus two residual error terms: (i) $(1-\alpha)\mathbb{E}[\sum_{\tau=1}^{T/B}\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} - \boldsymbol{z}_{\tau+1}^B(i)\rangle]$, representing the gossip approximation error; and (ii) $\alpha\,\mathbb{E}[\sum_{\tau=1}^{T/B}\langle \boldsymbol{q} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\rangle]$, representing the cost of uniform exploration. To control the first term, we utilize our key concentration result Lemma 1, which guarantees that the gossip error $\|\bar{\boldsymbol{z}}_{\tau+1} - \boldsymbol{z}_{\tau+1}^B(i)\|_2$ is bounded by $\mathcal{O}(B/(TK)^5)$. Since $\|\boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k\|_2 \leqslant \sqrt{2}$, summing this error over all T/B blocks yields a total contribution of $\mathcal{O}(1)$. To control the second term, we recall that $\bar{\boldsymbol{z}}_{\tau+1}$ is an unbiased estimator of the cumulative global average loss over a block $\frac{1}{N}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_\tau}\boldsymbol{\ell}_t(i)$, which has magnitude $\mathcal{O}(B)$. Summing over T/B blocks, and recalling that $\alpha = \mathcal{O}(T^{-1})$, the total expected cost becomes $\mathcal{O}(\alpha(T/B)B) = \mathcal{O}(1)$. Combining these bounds proves Lemma 3.

The second ingredient is to show that our subroutine $\mathcal{A}$ defined in Theorem 2 satisfies Eq. (6) with $\mathsf{R}^{\mathrm{del}} = \widetilde{\mathcal{O}}\big(\sqrt{(\rho^{-1/2} + K/N)T}\big)$, formally stated in Lemma 4.

Lemma 4 *Under the same assumptions as Theorem 2,*

$$\max_{i\in V}\max_{k\in[K]}\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle \boldsymbol{p}'_\tau(i)-\boldsymbol{e}_k,\boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right]\leqslant 2\sqrt{2\left(B+\frac{3K}{N}\right)(\log K)T}+6\,. \tag{7}$$

The proof of Lemma 4 is deferred to Appendix B.3 and we provide a sketch here. Fix an agent $i \in V$. Recall that Algorithm 2 handles the one-block delay by alternating two FTRL instances across odd and even blocks. This mechanism ensures that each instance of $\mathcal{B}$ receives the feedback from its previous execution before it is queried again, so each instance effectively observes non-delayed feedback. Consequently, we can decompose the analysis by considering the sequences generated by the two instances separately and then summing their regrets.

Focusing on the odd blocks, let $\{\boldsymbol{p}'_{2\tau-1}(i)\}_{\tau=1}^{T/2B}$ denote the sequence of action distributions generated by the first instance, $\mathcal{B}^{(1)}(i)$, based on the feedback $\boldsymbol{z}^B_{2\tau}(i)$. For the analysis, we introduce a ghost sequence $\{\bar{\boldsymbol{p}}_{2\tau-1}\}_{\tau=1}^{T/2B}$, generated by a hypothetical FTRL instance receiving the true global average loss estimates $\bar{\boldsymbol{z}}_{2\tau}$. We then decompose the regret of $\mathcal{B}^{(1)}(i)$ into three parts: (i) $\mathbb{E}[\sum_\tau\langle \boldsymbol{p}'_{2\tau-1}(i)-\bar{\boldsymbol{p}}_{2\tau-1},\bar{\boldsymbol{z}}_{2\tau}\rangle]$, capturing the discrepancy between the actual and ghost policies; (ii) $\mathbb{E}[\sum_\tau\langle \boldsymbol{p}'_{2\tau-1}(i)-\boldsymbol{e}_k,\boldsymbol{z}^B_{2\tau}(i)-\bar{\boldsymbol{z}}_{2\tau}\rangle]$, which is the approximation error of the global average loss estimate; and (iii) $\mathbb{E}[\sum_\tau\langle \bar{\boldsymbol{p}}_{2\tau-1}-\boldsymbol{e}_k,\bar{\boldsymbol{z}}_{2\tau}\rangle]$, which is the regret of the ghost policy on the true global average losses.

We bound these terms as follows. For term (i), Lemma 1 together with the stability properties of FTRL imply that $\|\bar{\boldsymbol{p}}_{2\tau-1}-\boldsymbol{p}'_{2\tau-1}(i)\|_1$ is bounded by $\mathcal{O}(B/(T^4K^5))$. Since uniform exploration ensures $\|\widehat{\boldsymbol{\ell}}_t(i)\|_2\leqslant KT$ (and thus $\|\bar{\boldsymbol{z}}_{2\tau}\|_2\leqslant BKT$), summing these over the blocks yields an $\mathcal{O}(1)$ bound. Similarly, term (ii) is bounded by $\mathcal{O}(1)$ as the approximation error $\|\boldsymbol{z}^B_{2\tau}(i)-\bar{\boldsymbol{z}}_{2\tau}\|_2$ is negligible by Lemma 1. For term (iii), a standard FTRL analysis leads to a bound of $\mathcal{O}(\log K/\eta+\eta\sum_\tau\sum_k\bar{\boldsymbol{p}}_{2\tau-1}(k)\mathbb{E}[\bar{\boldsymbol{z}}_{2\tau}(k)^2])$. A key deviation from the classic analysis is that $\bar{\boldsymbol{z}}_{2\tau}$ is constructed using actions sampled from the actual policy $\boldsymbol{p}_{2\tau-1}(i)$, instead of the ghost policy $\bar{\boldsymbol{p}}_{2\tau-1}$. Fortunately, we show in Lemma 12 (using Lemma 1 again) that $\bar{p}_{2\tau-1}(k)\leqslant 3p_{2\tau-1}(i,k)$ for all $i\in[N]$ and $k\in[K]$. This constant-factor bound on the importance weight mismatch results in an $\mathcal{O}(\log K/\eta+\eta(B+K/N)T)$ upper bound on term (iii). Picking η as defined in Theorem 2 and summing the contributions from both subsequences yields the stated bound.

4. Adaptive Bounds for Distributed K-armed Bandits

To demonstrate the versatility of our black-box reduction beyond worst-case minimax guarantees, we show in this section that our framework naturally yields *adaptive* regret bounds for distributed bandits. The underlying principle is straightforward: if Algorithm 1 is instantiated with a delayed bandit algorithm $\mathcal{A}$ that achieves a specific adaptive regret guarantee under delayed feedback (implying that Lemma 3 holds with $\mathsf{R}^{\mathrm{del}}$ being a data-dependent quantity), then this same bound carries over to the distributed setting via our reduction. For illustration, we continue to use Algorithm 2 as our subroutine $\mathcal{A}$, but we modify its base algorithm $\mathcal{B}$ to target specific guarantees. We explore two regimes: *small-loss* guarantees in Section 4.1 and *best-of-both-worlds* guarantees in Section 4.2.

To obtain small-loss guarantees, we follow Van Der Hoeven and Cesa-Bianchi (2022) who derived small-loss bounds for delayed adversarial bandits using FTRL with a hybrid negative entropy and log-barrier regularizer. To obtain *best-of-both-worlds* guarantees, we use FTRL with a negative

Tsallis entropy and log-barrier regularizer, as proposed by Masoudian et al. (2022). These examples highlight a general principle: as long as an adaptive bound is achievable in the delayed-feedback model, our reduction automatically transfers it to the distributed bandit setting. The full proof is deferred to Appendix C.1.

4.1. Small-Loss Bound

We first formalize our small-loss guarantee. In this regime, our goal is to improve the regret bound by replacing the worst-case dependence on the horizon T with the cumulative global average loss of the optimal action, $L^\star$, defined as $L^\star \triangleq \min_{k\in[K]} \sum_{t=1}^T \bar{\ell}_t(k) = \min_{k\in[K]} \sum_{t=1}^T \frac{1}{N}\sum_{i=1}^N \ell_t(i,k)$. Since $\ell_t(i,k) \in [0,1]$, we have $L^\star \leqslant T$. Therefore, replacing T with $L^\star$ keeps the worst-case guarantee while offering better bounds in benign environments, where the optimal action incurs low cumulative loss. The following theorem establishes that instantiating Algorithm 2 with FTRL (Algorithm 3) using a hybrid negative entropy and log-barrier regularizer achieves an optimal small-loss guarantee for distributed bandits.

Theorem 5 *Let $\mathcal{A}$ be Algorithm 2 run with base OLO algorithm $\mathcal{B}$ set to Algorithm 3 with $\psi_t(\boldsymbol{q}) = \frac{1}{\eta}\sum_{i=1}^K q(i)\log q(i) - \frac{1}{\gamma}\sum_{k=1}^K \log q(k)$ for all $t \in [T]$ where $\eta = \min\left\{\frac{1}{4B}, \sqrt{\frac{\log K}{BL^\star}}\right\}$ and $\gamma = \min\left\{\frac{N}{12}, \sqrt{\frac{KN\log T}{L^\star}}\right\}$. Assume all agents $i \in V$ run Algorithm 1 with bandit algorithm $\mathcal{A}$, gossip matrix W, and κ and B defined in Eq. (4). Then*

$$\begin{aligned}\mathrm{Reg}_T &= \mathcal{O}\Big(\sqrt{BL^\star \log K} + \sqrt{\frac{KL^\star \log T}{N}} + B\log K + \frac{K\log T}{N}\Big)\\ &= \tilde{\mathcal{O}}\Big(\sqrt{\rho^{-1/2}L^\star} + \sqrt{\frac{KL^\star}{N}} + \rho^{-1/2} + \frac{K}{N}\Big)\,.\end{aligned}$$

To the best of our knowledge, Theorem 5 provides the first small-loss regret guarantee for distributed K-armed bandits. We remark that while Theorem 5 requires tuning the learning rates η and γ based on the unknown quantity $L^\star$, this restriction can be removed using a standard doubling trick, since the approximation error of the global average loss estimate is negligible—of order $\mathcal{O}(B/(TK)^5)$. We refer the reader to Wei and Luo (2018); Lee et al. (2020) for a detailed exposition on adapting to unknown $L^\star$.

4.2. Best of Both Worlds

In this section, we extend our results to provide simultaneous guarantees for both adversarial and stochastic environments. Specifically, for the stochastic setting, we assume that for each agent $i \in V$ and arm $k \in [K]$, the loss $\ell_t(i,k) \in [0,1]$ at each round $t \in [T]$ is drawn i.i.d. from a distribution with mean $\mu(i,k) = \mathbb{E}[\ell_t(i,k)]$. We define the global average mean loss as $\boldsymbol{\mu} \in [0,1]^K$, where $\mu(k) \triangleq \frac{1}{N}\sum_{i=1}^N \mu(i,k)$. Following previous works (Zimmert and Seldin, 2020), we assume there exists a unique optimal arm $k^\star \triangleq \arg\min_{k\in[K]} \mu(k)$ and define the sub-optimality gaps $\boldsymbol{\delta} \in [0,1]^K$ as $\delta(k) \triangleq \mu(k) - \mu(k^\star)$. In this environment, the regret defined in Eq. (1) can be equivalently written as $\mathrm{Reg}_T = \max_{i\in V} \mathbb{E}\big[\sum_{t=1}^T \langle \boldsymbol{p}_t(i), \boldsymbol{\delta}\rangle\big]$.

The following theorem shows that by instantiating the base algorithm $\mathcal{B}$ with FTRL using a hybrid regularizer, specifically a combination of Shannon entropy and Tsallis entropy, we obtain bounds that are optimal up to logarithmic factors in the adversarial setting, while also enjoying logarithmic instance-dependent regret in the stochastic setting.

Theorem 6 *Let $\mathcal{A}$ be Algorithm 2 run with base OLO algorithm $\mathcal{B}$ set to Algorithm 3 with $\psi_t(\boldsymbol{q}) = \frac{1}{\eta_t}\sum_{i=1}^{K} q(i)\log q(i) - \frac{2}{\gamma_t}\sum_{k=1}^{K}\sqrt{q(k)}$ where $\eta_t = \min\left\{\frac{1}{B}, \sqrt{\frac{\log K}{tB^2}}\right\}$ and $\gamma_t = \sqrt{\frac{N}{tB}}$ and. Assume all agents $i \in V$ run Algorithm 1 with bandit algorithm $\mathcal{A}$, gossip matrix W, and parameters κ and B defined in Eq. (4), Then, in the adversarial environment, the agents' regret is bounded as*

$$\mathrm{Reg}_T = \mathcal{O}\Big(\sqrt{BT\log K} + \sqrt{\frac{KT}{N}} + B\log K\Big) = \widetilde{\mathcal{O}}\Big(\sqrt{\rho^{-1/2}T} + \sqrt{\frac{KT}{N}} + \rho^{-1/2}\Big)\,.$$

In the stochastic environment, the regret of the agents is bounded as

$$\mathrm{Reg}_T = \mathcal{O}\Big(\frac{B}{N}\sum_{k\neq k^\star}\frac{\log(T/B)}{\delta(k)} + \sum_{k\neq k^\star}\frac{B^2}{\delta(k)\log K} + B^2\log K\Big)\,.$$

Since $B = \Theta(\rho^{-1/2}\log(KNT))$, the bound for the stochastic setting is dominated by the second term, which scales as $\mathcal{O}(\sum_{k\neq k^\star}\frac{\log^2(KTN)}{\rho\cdot\delta(k)\log K})$. To the best of our knowledge, no prior work establishes minimax rates for best-of-both-worlds guarantees in distributed bandits, not even for the pure stochastic setting with heterogeneous losses and gossip-only communication. Specifically, in the pure stochastic setting, Zhang et al. (2025) achieve $\mathcal{O}(\sum_{k\neq k^\star}\frac{\log T}{N\delta(k)} + \frac{K\sqrt{N}}{\rho})$; however, their protocol requires additional communication overhead beyond the standard gossip permitted in our framework. Similarly, Martínez-Rubio et al. (2019) obtain $\mathcal{O}(\sum_{k\neq k^\star}\frac{\log T}{N\delta(k)} + \rho^{-1/2}K\log N)$, but their result is restricted to the homogeneous setting where $\mu(i,k) = \mu(k)$ for all agents. Determining the tight minimax rate for distributed stochastic bandits with heterogeneous losses remains an interesting open direction.

5. Distributed Adversarial Linear Bandits

In this section, we extend our black-box reduction to the setting of adversarial linear bandits with K actions in $\mathbb{R}^d$. To adapt the importance-weighting estimator (Line 8 in Algorithm 1) from K-armed to linear bandits, a standard approach is to construct a local loss coefficient estimator $\widehat{\boldsymbol{\theta}}_t(i) = M_\tau^{-1}(i)\boldsymbol{a}_{A_t(i)}\ell_t(i, A_t(i))$ at each round $t \in \mathcal{T}_\tau$, where $M_\tau(i) = \sum_{j\in[K]} p_\tau(i,j)\boldsymbol{a}_j\boldsymbol{a}_j^\top$ denotes the correlation matrix of the sampling distribution.

A naive strategy to estimate the global average loss would be to directly gossip $\widehat{\boldsymbol{\theta}}_t(i)$. However, unlike the MAB setting where the estimator's norm is bounded by KT due to the uniform exploration, $\widehat{\boldsymbol{\theta}}_t(i)$ may have an arbitrarily large ℓ_2-norm, preventing the direct application of Lemma 1 to control the consensus error. Alternatively, one could explicitly control the scale of the loss estimator by forcing exploration on a certain basis, and then gossiping the reconstructed loss estimate $\widehat{\ell}_t(i,k) = \langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i)\rangle$ for all $k \in [K]$. While this ensures boundedness of $\widehat{\boldsymbol{\ell}}_t(i)$, it will incur an $\mathcal{O}(K)$ communication cost per agent per round, which is prohibitive in linear bandits, where K is typically large. Therefore, in order to achieve $\mathcal{O}(d)$ communication cost, matching the communication efficiency of previous D-OCO protocols, we exploit the linear structure of the losses. Specifically, we restrict the gossip process to loss estimates defined only on a *volumetric spanner* $\mathcal{S} \subseteq \Omega$. As formalized below, this spanner forms a good basis for Ω, ensuring that any action can be represented as a linear combination with bounded coefficients.

Definition 1 (Volumetric Spanner (Hazan and Karnin, 2016)) *A subset $\mathcal{S} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_{|\mathcal{S}|}\} \subset \mathbb{R}^d$ is a volumetric spanner of $\Omega = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_K\} \subset \mathbb{R}^d$ if $\mathcal{S} \subseteq \Omega$ and for every $\boldsymbol{a}_k \in \Omega$, $\boldsymbol{a}_k$ can be expressed as $\boldsymbol{a}_k = \sum_{j=1}^{|\mathcal{S}|}\lambda^{(k)}(j)\,\boldsymbol{b}_j$ for some coefficients $\boldsymbol{\lambda}^{(k)} \in \mathbb{R}^{|\mathcal{S}|}$ satisfying $\|\boldsymbol{\lambda}^{(k)}\|_2 \leqslant 1$.*

Algorithm 4 Black-box reduction for linear bandits

Input: Agent index i, time horizon T, block length B, mixing coefficients κ, gossip matrix W, delayed bandit algorithm $\mathcal{A}$, action set $\Omega = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_K\}$, volumetric spanner $\mathcal{S} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_{|\mathcal{S}|}\} \subseteq \Omega$, exploration parameters β.

Initialization: Set $\boldsymbol{z}_1^{-1}(i) = \boldsymbol{z}_1^0(i) = \mathbf{0} \in \mathbb{R}^{|\mathcal{S}|}$, exploration parameters $\alpha = \frac{1}{T}$.

for $\tau = 1, \cdots, T/B$ **do**

  Receive $\boldsymbol{p}'_\tau(i)$ from $\mathcal{A}$ and compute $\boldsymbol{p}_\tau(i) = \big(1 - \alpha - \beta\big)\, \boldsymbol{p}'_\tau(i) + \frac{\alpha}{K}\mathbf{1} + \frac{\beta}{|\mathcal{S}|}\mathbf{1}_{\mathcal{S}}$

  Compute $M_\tau(i) = \sum_{j \in [K]} p_\tau(i,j)\boldsymbol{a}_j\boldsymbol{a}_j^\top$

  Set $b = 0$

  for $t \in \{(\tau-1)B + 1, \ldots, \tau B\}$ **do**

    Play $A_t(i) \sim \boldsymbol{p}_\tau(i)$ and observe $\ell_t\big(i, A_t(i)\big) = \langle \boldsymbol{a}_{A_t(i)}, \boldsymbol{\theta}_t(i)\rangle$

    Compute $\widehat{\boldsymbol{\theta}}_t(i) = M_\tau(i)^{-1}\boldsymbol{a}_{A_t(i)}\ell_t\big(i, A_t(i)\big)$

    Compute the estimator $\widehat{\boldsymbol{\ell}}_t(i) \in \mathbb{R}^{|\mathcal{S}|}$: For each $k \in \big[|\mathcal{S}|\big]$, $\widehat{\ell}_t(i,k) = \langle \boldsymbol{b}_k, \widehat{\boldsymbol{\theta}}_t(i)\rangle$

    Update $\boldsymbol{z}_\tau^{b+1}(i) = (1+\kappa)\sum_{j \in \mathcal{N}(i)} W(i,j)\boldsymbol{z}_\tau^b(j) - \kappa\boldsymbol{z}_\tau^{b-1}(i)$ and set $b \leftarrow b+1$

  end

  if $\tau \geqslant 2$ **then** send $\big(\tau - 1, \boldsymbol{z}_\tau^B(i)\big)$ to $\mathcal{A}$;

  Set $\boldsymbol{z}_{\tau+1}^{-1}(i) = \boldsymbol{z}_{\tau+1}^0(i) = \sum_{t=(\tau-1)B+1}^{\tau B} \widehat{\boldsymbol{\ell}}_t(i)$

end

This property allows us to reduce the communication complexity from K to $\mathcal{O}(d)$. Specifically, agents need only gossip the estimated losses for the basis vectors in $\mathcal{S}$. Using the linear relation in Definition 1, each agent can then locally reconstruct an approximation of the global average loss estimate for any action $\boldsymbol{a}_k \in \Omega$. Crucially, the constraint $\|\boldsymbol{\lambda}^{(k)}\|_2 \leqslant 1$ ensures that the consensus error does not grow during this reconstruction. The following proposition guarantees that such a spanner set has size at most $3d$ and can be found efficiently. In light of this, we assume that all agents pre-compute a common spanner $\mathcal{S}$ during initialization.

Proposition 2 (Bhaskara et al. (2023)) *Given a set $\Omega \subset \mathbb{R}^d$ of size K, there exists an efficient algorithm to compute a volumetric spanner $\mathcal{S}$ of Ω of size $|\mathcal{S}| \leqslant 3d$ in time $\mathcal{O}(Kd^3 \log d)$.*

With the volumetric spanner $\mathcal{S}$ in hand, we present our distributed linear bandit algorithm in Algorithm 4. While the overall structure mirrors the MAB reduction (Algorithm 1), we introduce three specific modifications to leverage the linear geometry and control communication costs. First, in each block τ, the sampling distribution $\boldsymbol{p}_\tau(i)$ mixes the base policy not only with $\frac{1}{T}$-uniform exploration over the action set Ω but also with an explicit exploration term over the spanner $\mathcal{S}$ (controlled by parameter β). This ensures the second moment matrix $M_\tau(i)$ is well-conditioned, making the estimated loss $\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i)\rangle$ for $k \in [K]$ bounded by $|\mathcal{S}|/\beta$ (as shown in Lemma 17 in Appendix D). Second, after playing an action $A_t(i)$ and observing the feedback, the agent constructs the standard unbiased local loss coefficient estimate $\widehat{\boldsymbol{\theta}}_t(i)$. Instead of communicating this potentially unbounded $\widehat{\boldsymbol{\theta}}_t(i)$ directly, the agent projects the loss onto the spanner basis and computes $\widehat{\boldsymbol{\ell}}_t(i) \in \mathbb{R}^{|\mathcal{S}|}$ where $\widehat{\ell}_t(i,k) = \langle \boldsymbol{b}_k, \widehat{\boldsymbol{\theta}}_t(i)\rangle$. Finally, the agents run B rounds of accelerated gossip on these $|\mathcal{S}|$-dimensional vectors. When the delayed feedback $\boldsymbol{z}_\tau^B(i)$ (an approximation of the global average loss estimate over the spanner) is finally passed to the base learner $\mathcal{B}$, we em-

ploy a modified base FTRL algorithm (Algorithm 5 deferred to Appendix D.1). Specifically, the learner first *reconstructs* the estimated loss for every original action $\boldsymbol{a}_k \in \Omega$ using the coefficients $\boldsymbol{\lambda}^{(k)}$ from Definition 1, and then performs a standard FTRL update with entropy regularizer over Ω (Dani et al., 2007; Cesa-Bianchi and Lugosi, 2012; Bubeck et al., 2012). The following theorem shows the guarantee of Algorithm 4 for the linear bandits setting.

Theorem 7 *Let $\mathcal{A}$ be Algorithm 2 run with base OLO algorithm $\mathcal{B}$ set to Algorithm 5 with $\psi(\boldsymbol{q}) = \frac{1}{\eta}\sum_{k=1}^{K} q(k)\log q(k)$ and $\eta = \min\left\{\frac{1}{6Bd}, \sqrt{\log K/(dTB + \frac{dT}{N})}\right\}$. Assume all agents $i \in V$ run Algorithm 4 with linear bandit algorithm $\mathcal{A}$, action set Ω, volumetric spanner $\mathcal{S} \subseteq \Omega$, gossip matrix W, $\beta = 3Bd\eta$, and exploration parameters κ and B defined in Eq. (4). Then, the regret of each agent is bounded as*

$$\mathrm{Reg}_T = \mathcal{O}\left(\sqrt{(\log K)\left(B + \tfrac{1}{N}\right)dT} + dB\log K\right) = \tilde{\mathcal{O}}\left(\sqrt{\left(\rho^{-1/2} + \tfrac{1}{N}\right)dT} + d\rho^{-1/2}\right) .$$

The proof is deferred to Appendix D. To the best of our knowledge, Theorem 7 provides the first regret guarantee for adversarial distributed linear bandits using only gossip communication. Crucially, our algorithm achieves this rate with $\mathcal{O}(d)$ communication per round and a dependence on the action set size K that is merely logarithmic. In Section 6, we complement this result with a lower bound of $\Omega(\rho^{-1/4}\sqrt{T} + \sqrt{dT/N})$. Comparing the two results reveals a gap of $\sqrt{d\log KNT}$ in the communication-dependent term. This is likely an artifact of our block-based approach, and we leave closing this gap as an open problem.

6. Lower Bound

In this section, we provide lower bounds for both distributed K-armed and distributed linear bandits. Recall that these bounds feature the sum of a communication cost term and a bandit cost term. For each setting, the proof shows that there exist problem instances forcing either term in the bound. Hence, the overall lower bound is asymptotic to the sum of these two terms. We start with the lower bound for K-armed bandits, which is a combination of the lower bound construction for the K-armed bandits with NT observed losses (Seldin et al., 2014) and the one in Yi and Vojnovic (2023). The full proof is deferred to Appendix E.

Theorem 8 *For any distributed K-armed bandit algorithm with $K \geqslant 2$, and for any N, T large enough, there exists a communication graph G with N nodes, a gossip matrix W, and a sequence $\ell_1, \ldots, \ell_T$ of loss matrices $\ell_t(\cdot,\cdot) \in [0,1]^{N\times K}$ such that $\mathrm{Reg}_T = \Omega\left(\sqrt{\rho^{-1/2}T\log K + \frac{KT}{N}}\right)$.*

Next, we state the lower bound for the linear setting. The proof is a direct combination of the K-armed lower bound and a lower bound for cooperative linear bandits (Ito et al., 2020). The full proof is deferred to Appendix E.

Theorem 9 *For any distributed linear bandit algorithm, and for any N, T large enough, there exists an action set $\Omega \subset \mathbb{R}^d$ with $|\Omega| = K \geqslant 2$, a communication graph G with N nodes, a gossip matrix W supported on G, and local loss coefficients $\boldsymbol{\theta}_t(i) \in \mathbb{R}^d$ for $t \in T$ and $i \in V$ such that $\mathrm{Reg}_T = \Omega\left(\sqrt{(\log K)\left(\rho^{-1/2} + \frac{d}{N}\right)T}\right)$.*

Acknowledgments

NCB acknowledges the financial support from the EU Horizon CL4-2021-HUMAN-01 research and innovation action under grant agreement 101070617, project ELSA (European Lighthouse on Secure and Safe AI) and the EU Horizon CL4-2022-HUMAN-02 research and innovation action under grant agreement 101120237, project ELIAS (European Lighthouse of AI for Sustainability).

References

William N Anderson Jr and Thomas D Morley. Eigenvalues of the Laplacian of a graph. *Linear and multilinear algebra*, 18(2):141–145, 1985.

Yogev Bar-On and Yishay Mansour. Individual regret in cooperative nonstochastic multi-armed bandits. *Advances in Neural Information Processing Systems*, 32, 2019.

Aditya Bhaskara, Sepideh Mahabadi, and Ali Vakilian. Tight bounds for volumetric spanners and applications. *Advances in Neural Information Processing Systems*, 36:916–930, 2023.

Ilai Bistritz, Zhengyuan Zhou, Xi Chen, Nicholas Bambos, and Jose Blanchet. Online exp3 learning in adversarial bandits with delayed feedback. *Advances in neural information processing systems*, 32, 2019.

Ethan Blaser, Chuanhao Li, and Hongning Wang. Federated linear contextual bandits with heterogeneous clients. In *International Conference on Artificial Intelligence and Statistics*, pages 631–639. PMLR, 2024.

Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE transactions on information theory*, 52(6):2508–2530, 2006.

Sébastien Bubeck, Nicolò Cesa-Bianchi, and Sham M Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Conference on Learning Theory*, pages 41–1. JMLR Workshop and Conference Proceedings, 2012.

Nicolò Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

Nicolò Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. Delay and cooperation in nonstochastic bandits. *Journal of Machine Learning Research*, 20(17):1–38, 2019.

Varsha Dani, Sham M Kakade, and Thomas Hayes. The price of bandit information for online optimization. *Advances in Neural Information Processing Systems*, 20, 2007.

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.

Dan Garber and Ben Kretzu. Improved regret bounds for projection-free bandit convex optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 2196–2206. PMLR, 2020.

Andras Gyorgy and Pooria Joulani. Adapting to delays and data in adversarial multi-armed bandits. In *International Conference on Machine Learning*, pages 3988–3997. PMLR, 2021.

Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016.

Elad Hazan and Zohar Karnin. Volumetric spanners: an efficient exploration basis for learning. *The Journal of Machine Learning Research*, 17(1):4062–4095, 2016.

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

Shinji Ito, Daisuke Hatano, Hanna Sumita, Kei Takemura, Takuro Fukunaga, Naonori Kakimura, and Ken-Ichi Kawarabayashi. Delay and cooperation in nonstochastic linear bandits. *Advances in Neural Information Processing Systems*, 33:4872–4883, 2020.

Pooria Joulani, Andras Gyorgy, and Csaba Szepesvári. Online learning under delayed feedback. In *International Conference on Machine Learning*, pages 1453–1461. PMLR, 2013.

Chung-Wei Lee, Haipeng Luo, and Mengxiao Zhang. A closer look at small-loss bounds for bandits with graph feedback. In *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 2516–2564. PMLR, 09–12 Jul 2020.

Xiuxian Li, Lihua Xie, and Na Li. A survey on distributed online optimization and online games. *Annual Reviews in Control*, 56:100904, 2023.

Ji Liu and A Stephen Morse. Accelerated linear iterations for distributed averaging. *Annual Reviews in Control*, 35(2):160–165, 2011.

Jingyuan Liu, Hao Qiu, Lin Yang, and Mengfan Xu. Distributed multi-agent bandits erdos-rényi random networks. In *Advances in Neural Information Processing Systems*, volume 38, pages 159769–159798. Curran Associates, Inc., 2025.

David Martínez-Rubio, Varun Kanade, and Patrick Rebeschini. Decentralized cooperative stochastic bandits. *Advances in Neural Information Processing Systems*, 32, 2019.

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.

Angelia Nedic and Asuman Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009.

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213v8*, 2025.

Kevin Scaman, Francis Bach, Sébastien Bubeck, Yin Tat Lee, and Laurent Massoulié. Optimal convergence rates for convex distributed optimization in networks. *Journal of Machine Learning Research*, 20(159):1–31, 2019.

Yevgeny Seldin, Peter Bartlett, Koby Crammer, and Yasin Abbasi-Yadkori. Prediction with limited advice and multiarmed bandits with paid observations. In *International Conference on Machine Learning*, pages 280–287. PMLR, 2014.

Tobias Sommer Thune, Nicolò Cesa-Bianchi, and Yevgeny Seldin. Nonstochastic multiarmed bandits with unrestricted delays. *Advances in Neural Information Processing Systems*, 32, 2019.

Dirk Van Der Hoeven and Nicolò Cesa-Bianchi. Nonstochastic bandits and experts with arm-dependent delays. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 2022.

Dirk van der Hoeven, Lukas Zierahn, Tal Lancewicki, Aviv Rosenberg, and Nicolò Cesa-Bianchi. A unified analysis of nonstochastic delayed feedback for combinatorial semi-bandits, linear bandits, and mdps. In *Conference on Learning Theory*, pages 1285–1321. PMLR, 2023.

Yuanyu Wan. Black-box reductions for decentralized online convex optimization in changing environments. In *Conference on Learning Theory*, pages 5605–5631. PMLR, 2025.

Yuanyu Wan, Tong Wei, Mingli Song, and Lijun Zhang. Nearly optimal regret for decentralized online convex optimization. In *Conference on Learning Theory*, pages 4862–4888. PMLR, 2024.

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

Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53(1):65–78, 2004.

Mengfan Xu and Diego Klabjan. Decentralized randomly distributed multi-agent multi-armed bandit with heterogeneous rewards. *Advances in Neural Information Processing Systems*, 36:74799–74855, 2023.

Feng Yan, Shreyas Sundaram, SVN Vishwanathan, and Yuan Qi. Distributed autonomous online learning: Regrets and intrinsic privacy-preserving properties. *IEEE Transactions on Knowledge and Data Engineering*, 25(11):2483–2493, 2013.

Haishan Ye, Luo Luo, Ziang Zhou, and Tong Zhang. Multi-consensus decentralized accelerated gradient descent. *Journal of machine learning research*, 24(306):1–50, 2023a.

Mang Ye, Xiuwen Fang, Bo Du, Pong C Yuen, and Dacheng Tao. Heterogeneous federated learning: State-of-the-art and research challenges. *ACM Computing Surveys*, 56(3):1–44, 2023b.

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

Deming Yuan, Alexandre Proutiere, Guodong Shi, et al. Multi-agent online optimization. *Foundations and Trends® in Optimization*, 7(2-3):81–263, 2024.

Valentina Zantedeschi, Aurélien Bellet, and Marc Tommasi. Fully decentralized joint learning of personalized models and collaboration graphs. In *International Conference on Artificial Intelligence and Statistics*, pages 864–874. PMLR, 2020.

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.

Jingxuan Zhu and Ji Liu. Distributed multiarmed bandits. *IEEE Transactions on Automatic Control*, 68(5):3025–3040, 2023.

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*, pages 3285–3294. PMLR, 2020.

Appendix A. Auxiliary Results

Lemma 10 *Let $\psi : \Delta(K) \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex function that is μ-strongly convex with respect to $\|\cdot\|_2$ on $\Delta(K)$, i.e., for all $\boldsymbol{x}, \boldsymbol{y} \in \Delta(K)$ and all $\boldsymbol{g} \in \partial\psi(\boldsymbol{x})$,*

$$\psi(\boldsymbol{y}) \geqslant \psi(\boldsymbol{x}) + \langle \boldsymbol{g}, \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{\mu}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 \,.$$

Let $\boldsymbol{\xi} : \mathbb{R}^K \to \Delta(K)$ be the argmin map

$$\boldsymbol{\xi}(\boldsymbol{a}) = \operatorname*{arg\,min}_{\boldsymbol{u} \in \Delta(K)} \langle \boldsymbol{a}, \boldsymbol{u} \rangle + \psi(\boldsymbol{u}) \,.$$

Then the argmin map is $1/\mu$-Lipschitz w.r.t. the ℓ_2-norm,

$$\|\boldsymbol{\xi}(\boldsymbol{a}) - \boldsymbol{\xi}(\boldsymbol{b})\|_2 \leqslant \frac{1}{\mu}\,\|\boldsymbol{a} - \boldsymbol{b}\|_2 \,.$$

Proof Pick $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^K$ and let $\boldsymbol{x} = \boldsymbol{\xi}(\boldsymbol{a})$ and $\boldsymbol{y} = \boldsymbol{\xi}(\boldsymbol{b})$. By first-order optimality of $\boldsymbol{x}, \boldsymbol{y}$, there exist $\boldsymbol{g}_x \in \partial\psi(\boldsymbol{x})$ and $\boldsymbol{g}_y \in \partial\psi(\boldsymbol{y})$ such that

$$\langle \boldsymbol{a} + \boldsymbol{g}_x, \boldsymbol{u} - \boldsymbol{x} \rangle \geqslant 0 \qquad \text{and} \qquad \langle \boldsymbol{b} + \boldsymbol{g}_y, \boldsymbol{u} - \boldsymbol{y} \rangle \geqslant 0 \qquad \forall \boldsymbol{u} \in \Delta(K) \,.$$

Taking $\boldsymbol{u} = \boldsymbol{y}$ in the first inequality and $\boldsymbol{u} = \boldsymbol{x}$ in the second inequality, and adding them together gives

$$\langle \boldsymbol{a} - \boldsymbol{b}, \boldsymbol{y} - \boldsymbol{x} \rangle \geqslant \langle \boldsymbol{g}_x - \boldsymbol{g}_y, \boldsymbol{x} - \boldsymbol{y} \rangle \,.$$

Direct calculation shows that strong convexity implies strong monotonicity of the subgradient mapping:

$$\langle \boldsymbol{g}_x - \boldsymbol{g}_y, \boldsymbol{x} - \boldsymbol{y} \rangle \geqslant \mu\|\boldsymbol{x} - \boldsymbol{y}\|_2^2.$$

Hence $\langle \boldsymbol{a} - \boldsymbol{b}, \boldsymbol{y} - \boldsymbol{x} \rangle \geqslant \mu\|\boldsymbol{x} - \boldsymbol{y}\|_2^2$. By Cauchy-Schwarz inequality, we know that $\langle \boldsymbol{a} - \boldsymbol{b}, \boldsymbol{x} - \boldsymbol{y} \rangle \leqslant \|\boldsymbol{a} - \boldsymbol{b}\|_2\,\|\boldsymbol{x} - \boldsymbol{y}\|_2$. Dividing $\|\boldsymbol{x} - \boldsymbol{y}\|_2$ on both sides yields the claim. ■

A.1. Convergence properties for accelerated gossip

Given a set of vectors denoted as $\boldsymbol{x}(1), \boldsymbol{x}(2), \ldots, \boldsymbol{z}(N) \in \mathbb{R}^K$ and let $\boldsymbol{x}^0(i) = \boldsymbol{x}^{-1}(i) = \boldsymbol{x}(i)$ for all $i \in V$. To approximate the average $\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i\in V}\boldsymbol{x}(i)$, Liu and Morse (2011) considers the following accelerated gossip process:

$$\boldsymbol{x}^{b+1}(i) = (1+\kappa)\sum_{j\in\mathcal{N}(i)} W(i,j)\boldsymbol{x}^b(j) - \kappa\boldsymbol{x}^{b-1}(i)\,, \tag{8}$$

for all $b \geqslant 0$, where κ is the mixing coefficient. Let $X^b \triangleq \left[\boldsymbol{x}^b(1)^\top, \boldsymbol{x}^b(2)^\top, \ldots, \boldsymbol{x}^b(N)^\top\right] \in \mathbb{R}^{N\times K}$ and $\bar{X} \triangleq \left[\bar{\boldsymbol{x}}^\top, \bar{\boldsymbol{x}}^\top, \ldots, \bar{\boldsymbol{x}}^\top\right] \in \mathbb{R}^{N\times K}$. According to iteration in Eq. (8), it is not hard to verify that

$$X^{b+1} = (1+\kappa)WX^b - \kappa X^{b-1}\,,$$

for all $b \geqslant 0$. Ye et al. (2023a) shows the following convergence property.

Lemma 11 (Proposition 1 in Ye et al. (2023a)) *For $B \geqslant 1$, the iterations of Eq. (8) with gossip matrix W and parameters κ defined in Eq. (4) ensure*

$$\left\|X^B - \bar{X}\right\|_F \leqslant \sqrt{14}\left(1 - \left(1 - \frac{1}{\sqrt{2}}\right)\sqrt{1-\sigma_2(W)}\right)^B \left\|X^0 - \bar{X}\right\|_F.$$

Appendix B. Omitted Proof Details for Distributed K-armed Bandits

B.1. Omitted proof details for Lemma 1

For completeness, we restate Lemma 1 and provide its proof as follows.

Lemma 1 *If all agents $i \in V$ run Algorithm 1 with gossip matrix W and parameters κ, B defined in Eq. (4), then*

$$\max_{i\in V}\max_{\tau\in[T/B]}\left\|\boldsymbol{z}_\tau^B(i) - \bar{\boldsymbol{z}}_\tau\right\|_2 \leqslant \frac{2B}{(KT)^5}, \tag{5}$$

where $\bar{\boldsymbol{z}}_\tau \triangleq \frac{1}{N}\sum_{i=1}^N\sum_{t\in\mathcal{T}_{\tau-1}}\widehat{\boldsymbol{\ell}}_t(i)$ is defined in Eq. (3) for all $\tau \in [T/B]$.

Proof Fix an epoch index τ and define $X_\tau^b \triangleq \left[\boldsymbol{z}_\tau^b(1)^\top, \boldsymbol{z}_\tau^b(2)^\top, \ldots, \boldsymbol{z}_\tau^b(N)^\top\right] \in \mathbb{R}^{N\times K}$ for any integer $b \geqslant -1$. Let $\mathcal{T}_\tau \triangleq \{(\tau-1)B+1, (\tau-1)B+2, \ldots, \tau B\}$ be the round index in epoch τ. Recall that for each $i \in V$,

$$\boldsymbol{z}_\tau^{-1}(i) = \boldsymbol{z}_\tau^0(i) = \sum_{t\in\mathcal{T}_{\tau-1}}\widehat{\boldsymbol{\ell}}_t(i)\,,$$

which means that

$$X_\tau^{-1} = X_\tau^0 = \left[\sum_{t\in\mathcal{T}_{\tau-1}}\widehat{\boldsymbol{\ell}}_t(1)^\top, \ldots, \sum_{t\in\mathcal{T}_{\tau-1}}\widehat{\boldsymbol{\ell}}_t(N)^\top\right]. \tag{9}$$

According to the dynamic of $\boldsymbol{z}_\tau^b(i)$ defined in Algorithm 1, it is straightforward to verify that for all $b \in \{0, 1, \ldots, B-1\}$,

$$X_\tau^{b+1} = (1+\kappa)WX_\tau^b - \kappa X_\tau^{b-1}. \tag{10}$$

In order to bound $\|\bar{\boldsymbol{z}}_\tau - \boldsymbol{z}_\tau^B(i)\|_2$, we define $\bar{X}_\tau \triangleq \left[\bar{\boldsymbol{z}}_\tau^\top, \ldots, \bar{\boldsymbol{z}}_\tau^\top\right]$. When $\tau = 1$, since $\bar{\boldsymbol{z}}_1 = \boldsymbol{z}_1^B(i) = \mathbf{0}$, the conclusion holds trivially. When $\tau \geqslant 2$, by definition of X_τ^B and $\bar{X}_\tau$, we know that

$\left\|\boldsymbol{z}_\tau^B(i) - \bar{\boldsymbol{z}}_\tau\right\|_2 \leqslant \left\|X_\tau^B - \bar{X}_\tau\right\|_F$. Then, according to Ye et al. (2023a, Proposition 1) (included in Appendix A as Lemma 11), we have

$$\left\|X_\tau^B - \bar{X}_\tau\right\|_F \leqslant \sqrt{14}\left(1 - c\sqrt{1-\sigma_2(W)}\right)^B \left\|X_\tau^0 - \bar{X}_\tau\right\|_F \ , \tag{11}$$

where $c = 1 - 1/\sqrt{2}$. Further plugging in the definition of B in Eq. (4) leads to the following

$$\begin{aligned}
&\sqrt{14}\left(1 - c\sqrt{1-\sigma_2(W)}\right)^B \left\|X_\tau^0 - \bar{X}_\tau\right\|_F \\
&\leqslant \sqrt{14}\left(1 - c\sqrt{1-\sigma_2(W)}\right)^{\frac{\ln(K^6T^6\sqrt{14N})}{c\sqrt{1-\sigma_2(W)}}} \left\|X_\tau^0 - \bar{X}_\tau\right\|_F \\
&\leqslant \sqrt{14}\left(1 - c\sqrt{1-\sigma_2(W)}\right)^{\frac{\ln(K^6T^6\sqrt{14N})}{-\ln(1-c\sqrt{1-\sigma_2(W)})}} \left\|X_\tau^0 - \bar{X}_\tau\right\|_F \\
&= \frac{\left\|X_\tau^0 - \bar{X}_\tau\right\|_F}{T^6K^6\sqrt{N}} \ ,
\end{aligned} \tag{12}$$

where the first inequality is because of the definition of B and the second inequality is due to $-\ln x \geqslant 1 - x$ for any $x > 0$.

Combining Eq. (11) and Eq. (12), for any $i \in V$ and $\tau \geqslant 2$, we have

$$\begin{aligned}
\left\|\boldsymbol{z}_\tau^B(i) - \bar{\boldsymbol{z}}_\tau\right\|_2 &\leqslant \frac{\left\|X_\tau^0 - \bar{X}_\tau\right\|_F}{T^6K^6\sqrt{N}} \\
&\leqslant \frac{\left\|X_\tau^0\right\|_F + \left\|\bar{X}_\tau\right\|_F}{T^6K^6\sqrt{N}} && \text{(triangle inequality)} \\
&\leqslant \frac{2\sqrt{\sum_{i=1}^N \sum_{k=1}^K \left(\sum_{t\in\mathcal{T}_{\tau-1}} \widehat{\ell}_t(i,k)\right)^2}}{T^6K^6\sqrt{N}} \ ,
\end{aligned} \tag{13}$$

where the last inequality uses Eq. (9) and the fact that $\|\bar{X}_\tau\|_F \leqslant \|X_\tau^0\|_F$ due to Jensen's inequality.

To further bound the right-hand side of Eq. (13), since $\ell_t(i,k) \in [0,1]$ and $p_\tau(i,k) \geqslant 1/(KT)$ for all $t \in [T]$, $i \in V$, $k \in [K]$, and $\tau \geqslant 1$, and $\widehat{\boldsymbol{\ell}}_t(i)$ has at most one nonzero coordinate, we know that that $\|\boldsymbol{\ell}_t(i)\|_2 \leqslant KT$. Therefore, by triangular inequality,

$$\left\|\sum_{t\in\mathcal{T}_\tau} \widehat{\boldsymbol{\ell}}_t(i)\right\|_2 \leqslant \sum_{t\in\mathcal{T}_\tau} \left\|\widehat{\boldsymbol{\ell}}_t(i)\right\|_2 \leqslant |\mathcal{T}_\tau| KT \leqslant BKT, \tag{14}$$

where B is an upper bound on the block length (i.e., $|\mathcal{T}_\tau| \leqslant B$). Substituting Eq. (14) into Eq. (13) yields the final conclusion:

$$\left\|\boldsymbol{z}_\tau^B(i) - \bar{\boldsymbol{z}}_\tau\right\|_2 \leqslant \frac{2B}{T^5K^5}.$$

■

B.2. Omitted Proof Details for Lemma 3

In this subsection, we present the proof for Lemma 3, which formalizes the connection between the performance of the subroutine $\mathcal{A}$ used in Algorithm 1 and our global regret objectives. The result shows that a regret guarantee for $\mathcal{A}$ on the given loss vectors implies a corresponding bound on the target regret, subject only to a small additive constant.

Lemma 3 *Suppose that every agent $i \in V$ runs an instance of Algorithm 1 using a base algorithm $\mathcal{A}$ that guarantees*

$$\max_{i\in V}\max_{k\in[K]} \mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right] \leqslant \mathsf{R}^{\mathrm{del}}\,, \tag{6}$$

for some $\mathsf{R}^{\mathrm{del}} \geqslant 0$, then we have $\mathrm{Reg}_T \leqslant \mathsf{R}^{\mathrm{del}} + 4$.

Proof Recall that Algorithm 1 uses $\boldsymbol{p}_\tau(i) = (1-\alpha)\boldsymbol{p}'_\tau(i) + \alpha\boldsymbol{q}$ where $\boldsymbol{q} = \frac{1}{K}\mathbf{1}$, $\alpha = \frac{1}{T}$, and $\boldsymbol{p}'_\tau(i)$ is the distribution over $\Delta(K)$ received by $\mathcal{A}$. Let

$$k \triangleq \operatorname*{arg\,max}_{k'\in[K]} \mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}_\tau(i) - \boldsymbol{e}_{k'}, \sum_{t\in\mathcal{T}_\tau} \bar{\boldsymbol{\ell}}_t \right\rangle\right]\,.$$

Then, we decompose the regret as follows:

$$\begin{aligned}
\mathrm{Reg}_T(i) &= \mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}_\tau(i) - \boldsymbol{e}_k, \sum_{t\in\mathcal{T}_\tau} \bar{\boldsymbol{\ell}}_t \right\rangle\right] \\
&= \mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right] && \text{(by definition of } \bar{\boldsymbol{z}}_{\tau+1} \text{ in Eq. (3))} \\
&= \mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle (1-\alpha)\,\boldsymbol{p}'_\tau(i) + \alpha\boldsymbol{q} - \boldsymbol{e}_k, \boldsymbol{z}^B_{\tau+1}(i) + \bar{\boldsymbol{z}}_{\tau+1} - \boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right] \\
& && \text{(by definition of } \boldsymbol{p}_\tau(i)) \\
&= (1-\alpha)\mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right] \\
&\quad + (1-\alpha)\mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} - \boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right] \\
&\quad + \alpha\,\mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{q} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right] \\
&\leqslant \mathsf{R}^{\mathrm{del}} && \text{(using Eq. (6))}
\end{aligned}$$

$$+ (1-\alpha)\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} - \boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right] \tag{15}$$

$$+ \alpha\,\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle \boldsymbol{q} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right]. \tag{16}$$

We now bound Eq. (15) and Eq. (16) respectively.

Bounding Eq. (15). By Cauchy-Schwarz inequality, for each block $\tau \geqslant 1$,

$$\left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} - \boldsymbol{z}^B_{\tau+1}(i)\right\rangle \leqslant \left\|\boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k\right\|_2 \cdot \left\|\bar{\boldsymbol{z}}_{\tau+1} - \boldsymbol{z}^B_{\tau+1}(i)\right\|_2 .$$

Since $\boldsymbol{p}'_\tau(i) \in \Delta(K)$, $\|\boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k\|_2 \leqslant \sqrt{2}$. Moreover, by Lemma 1, we know that the second term is bounded as follows:

$$\left\|\bar{\boldsymbol{z}}_{\tau+1} - \boldsymbol{z}^B_{\tau+1}(i)\right\|_2 \leqslant \frac{2B}{T^5K^5}.$$

Therefore,

$$\left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} - \boldsymbol{z}^B_{\tau+1}(i)\right\rangle \leqslant \frac{2\sqrt{2}B}{T^5K^5} .$$

Taking summation over $\tau = 1, \ldots, T/B$ yields

$$(1-\alpha)\,\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} - \boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right] \leqslant \frac{T}{B}\,\frac{2\sqrt{2}B}{T^5K^5} = \frac{2\sqrt{2}}{T^4K^5} \leqslant 3, \tag{17}$$

where the last inequality uses the fact that $T \geqslant 1$ and $K \geqslant 1$.

Bounding Eq. (16). Since $\ell_t(i,k) \in [0,1]$ for all $i \in [N]$, $t \in [T]$, and $k \in [K]$, we know that each coordinate of $\bar{\boldsymbol{z}}_{\tau+1}$ is non-negative. Therefore, $\langle \boldsymbol{q} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\rangle \leqslant \langle \boldsymbol{q}, \bar{\boldsymbol{z}}_{\tau+1}\rangle$, which leads to

$$\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle \boldsymbol{q} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right] \leqslant \sum_{\tau=1}^{T/B}\mathbb{E}\big[\langle \boldsymbol{q}, \bar{\boldsymbol{z}}_{\tau+1}\rangle\big] = \sum_{\tau=1}^{T/B}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\big[\bar{z}_{\tau+1}(k)\big].$$

Since $\mathbb{E}\left[\widehat{\ell}_t(i,k)\right] = \ell_t(i,k)$ via a direct calculation, we know that

$$\mathbb{E}\big[\bar{z}_{\tau+1}(k)\big] = \frac{1}{N}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_\tau}\mathbb{E}\Big[\widehat{\ell}_t(i,k)\Big] \leqslant \frac{1}{N}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_\tau} 1 = B$$

and hence $\mathbb{E}\big[\langle q, \bar{\boldsymbol{z}}_{\tau+1}\rangle\big] \leqslant B$. This implies that

$$\alpha\,\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle \boldsymbol{q} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right] \leqslant \alpha\sum_{\tau=1}^{T/B} B = \alpha T = 1 \tag{18}$$

where we used $\alpha = \frac{1}{T}$ in the last equality. Finally, plugging in Eq. (17) and Eq. (18) into the regret decomposition proves the statement. ■

B.3. Omitted Proof Details for Lemma 4

In this section, we present the proof for Lemma 4, which provides the concrete regret guarantee for Algorithm 2 with base OLO algorithm $\mathcal{B}$ set to Algorithm 3, using the entropy regularizer and η defined in Theorem 2.

Lemma 4 *Under the same assumptions as Theorem 2,*

$$\max_{i\in V}\max_{k\in[K]}\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle \boldsymbol{p}'_\tau(i)-\boldsymbol{e}_k,\boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right]\leqslant 2\sqrt{2\left(B+\frac{3K}{N}\right)(\log K)T}+6\,. \tag{7}$$

Proof Let $M\triangleq T/B$ denote the number of blocks and define the parity sets

$$\mathcal{P}_0\triangleq\{\tau\in[M]:\tau\text{ is even}\}\,,\qquad \mathcal{P}_1\triangleq\{\tau\in[M]:\tau\text{ is odd}\}.$$

Let $m_r=|\mathcal{P}_r|$ for $r\in\{0,1\}$, i.e.,

$$m_0=\left\lfloor\frac{M}{2}\right\rfloor\,,\qquad m_1=\left\lceil\frac{M}{2}\right\rceil.$$

Define the local-to-global index maps as follows:

$$\pi_0(s)\triangleq 2s-1,\quad s=1,\dots,m_0\qquad\text{and}\qquad \pi_1(s)\triangleq 2s,\quad s=1,\dots,m_1 \tag{19}$$

so that $\mathcal{P}_r=\{\pi_r(s):s\in[m_r]\}$ for each $r\in\{0,1\}$. Fix an agent $i\in V$ and $k\in[K]$, we decompose the regret as follows:

$$\begin{aligned}&\mathbb{E}\left[\sum_{\tau=1}^{M}\left\langle \boldsymbol{p}'_\tau(i)-\boldsymbol{e}_k,\boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right]\\&=\underbrace{\mathbb{E}\left[\sum_{s=1}^{m_0}\left\langle \boldsymbol{p}'_{\pi_0(s)}(i)-\boldsymbol{e}_k,\boldsymbol{z}^B_{\pi_0(s)+1}(i)\right\rangle\right]}_{\triangleq\mathrm{Reg}_{\mathcal{P}_0}(i)}+\underbrace{\mathbb{E}\left[\sum_{s=1}^{m_1}\left\langle \boldsymbol{p}'_{\pi_1(s)}(i)-\boldsymbol{e}_k,\boldsymbol{z}^B_{\pi_1(s)+1}(i)\right\rangle\right]}_{\triangleq\mathrm{Reg}_{\mathcal{P}_1}(i)}.\end{aligned}$$

We now analyze $\mathrm{Reg}_{\mathcal{P}_1}(i)$, the analysis for $\mathrm{Reg}_{\mathcal{P}_0}(i)$ is analogous.

According to the update rule of $\boldsymbol{q}_t$ in Algorithm 3, we know that $\boldsymbol{p}'_{\pi_1(s)}(i)$ is computed as follows:

$$\boldsymbol{p}'_{\pi_1(s)}(i)=\operatorname*{arg\,min}_{\boldsymbol{q}\in\Delta(K)}\sum_{s'=1}^{s-1}\left\langle \boldsymbol{z}^B_{\pi_1(s')+1}(i),\boldsymbol{q}\right\rangle+\frac{1}{\eta}\sum_{k=1}^{K}q(k)\log\left(q(k)\right) \tag{20}$$

To analyze $\mathrm{Reg}_{\mathcal{P}_1}(i)$, we recall the definition of $\bar{\boldsymbol{z}}_\tau$ defined in Eq. (3)

$$\bar{\boldsymbol{z}}_{\pi_1(s)+1}=\frac{1}{N}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\widehat{\boldsymbol{\ell}}_t(i)$$

and define $\bar{\boldsymbol{p}}_{\pi_1(s)}$ as follows

$$\bar{\boldsymbol{p}}_{\pi_1(s)} \triangleq \operatorname*{arg\,min}_{\boldsymbol{q}\in\Delta(K)} \sum_{s'=1}^{s-1} \left\langle \bar{\boldsymbol{z}}_{\pi_1(s')+1}, \boldsymbol{q} \right\rangle + \frac{1}{\eta} \sum_{k=1}^{K} q(k) \log\left(q(k)\right), \tag{21}$$

which is the strategy output by FTRL with loss vectors $\{\bar{\boldsymbol{z}}_{\pi_1(s')+1}\}_{s'\in[s-1]}$. Now we decompose the regret into the following three terms and bound each term separately.

$$\begin{aligned}
\mathrm{Reg}_{\mathcal{P}_1}(i) &= \mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \boldsymbol{p}'_{\pi_1(s)}(i) - \boldsymbol{e}_k, \boldsymbol{z}^B_{\pi_1(s)+1}(i) \right\rangle\right] \\
&= \underbrace{\mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \boldsymbol{p}'_{\pi_1(s)}(i) - \bar{\boldsymbol{p}}_{\pi_1(s)}, \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle\right]}_{\clubsuit} \\
&\quad + \underbrace{\mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \boldsymbol{p}'_{\pi_1(s)}(i) - \boldsymbol{e}_k, \boldsymbol{z}^B_{\pi_1(s)+1}(i) - \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle\right]}_{\heartsuit} \\
&\quad + \underbrace{\mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle\right]}_{\spadesuit} .
\end{aligned} \tag{22}$$

Bounding the term ♣. By Cauchy-Schwarz inequality, for each $s \in [m_1]$, we have

$$\left\langle \boldsymbol{p}'_{\pi_1(s)}(i) - \bar{\boldsymbol{p}}_{\pi_1(s)}, \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle \leqslant \|\boldsymbol{p}'_{\pi_1(s)}(i) - \bar{\boldsymbol{p}}_{\pi_1(s)}\|_2 \cdot \|\bar{\boldsymbol{z}}_{\pi_1(s)+1}\|_2 .$$

Since $\boldsymbol{p}'_{\pi_1(s)}(i)$ and $\bar{\boldsymbol{p}}_{\pi_1(s)}$ follow the update rule of Eq. (20) and Eq. (21) and using $\psi(\boldsymbol{q}) = \sum_{k=1}^{K} q(k)\log(q(k))$ is 1-strongly convex w.r.t. ℓ_2-norm, Lemma 12 yields that

$$\|\boldsymbol{p}'_{\pi_1(s)}(i) - \bar{\boldsymbol{p}}_{\pi_1(s)}\|_2 \leqslant \frac{2\eta B}{T^4 K^5} .$$

Moreover, since each $\widehat{\boldsymbol{\ell}}_t(i)$ has at most one non-zero coordinate and $p_{\pi_1(s)}(i,j) \geqslant \frac{\alpha}{K} = 1/(KT)$ for all $j \in [K]$, we know that $\|\widehat{\boldsymbol{\ell}}_t(i)\|_2 \leqslant KT$, and hence

$$\|\bar{\boldsymbol{z}}_{\pi_1(s)+1}\|_2 = \left\| \frac{1}{N} \sum_{t\in\mathcal{T}_{\pi_1(s)}} \sum_{i=1}^{N} \widehat{\boldsymbol{\ell}}_t(i) \right\|_2 \leqslant \frac{1}{N} \sum_{t\in\mathcal{T}_{\pi_1(s)}} \sum_{i=1}^{N} \|\widehat{\boldsymbol{\ell}}_t(i)\|_2 \leqslant BKT .$$

Therefore,

$$\clubsuit \leqslant \sum_{s=1}^{m_1} \mathbb{E}\Big[\|\boldsymbol{p}'_{\pi_1(s)}(i) - \bar{\boldsymbol{p}}_{\pi_1(s)}\|_2 \cdot \|\bar{\boldsymbol{z}}_{\pi_1(s)+1}\|_2\Big] \leqslant \sum_{s=1}^{m_1} \frac{2\eta B}{T^4 K^5} \cdot BKT = \frac{2\eta\, m_1\, B^2}{T^3 K^4} \leqslant 2 \tag{23}$$

where the last inequality uses $m_1 \leqslant T/B$, $B \leqslant T$ and $\eta \leqslant 1$.

Bounding the term ♡. By Cauchy-Schwarz inequality, for each $s \in [m_1]$, we have

$$\left\langle \boldsymbol{p}'_{\pi_1(s)}(i) - \boldsymbol{e}_k, \boldsymbol{z}^B_{\pi_1(s)+1}(i) - \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle \leqslant \left\| \boldsymbol{p}'_{\pi_1(s)}(i) - \boldsymbol{e}_k \right\|_2 \cdot \left\| \boldsymbol{z}^B_{\pi_1(s)+1}(i) - \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\|_2 .$$

Since $\boldsymbol{p}'_{\pi_1(s)}(i) \in \Delta(K)$ and $\boldsymbol{e}_k$ is a vertex of the simplex, we have $\|\boldsymbol{p}'_{\pi_1(s)}(i) - \boldsymbol{e}_k\|_2 \leqslant \sqrt{2}$. Moreover, according to Lemma 1, we know that

$$\left\| \boldsymbol{z}^B_{\pi_1(s)+1}(i) - \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\|_2 \leqslant \frac{2B}{T^5 K^5} .$$

Therefore,

$$\heartsuit = \mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \boldsymbol{p}'_{\pi_1(s)}(i) - \boldsymbol{e}_k, \boldsymbol{z}^B_{\pi_1(s)+1}(i) - \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle \right] \leqslant \sum_{s=1}^{m_1} \frac{2\sqrt{2}B}{T^5 K^5} \leqslant \frac{3 m_1 B}{T^5 K^5} \leqslant 1, \tag{24}$$

where the last inequality uses $m_1 \leqslant T$ and $B \leqslant T$.

Bounding the term ♠. Recall that the definitions of $\bar{\boldsymbol{z}}_{\pi_1(s)+1}$ and $\widehat{\ell}_t(i,j)$ are as follows

$$\bar{\boldsymbol{z}}_{\pi_1(s)+1} = \frac{1}{N} \sum_{t \in \mathcal{T}_{\pi_1(s)}} \sum_{i=1}^{N} \widehat{\boldsymbol{\ell}}_t(i), \quad \widehat{\ell}_t(i,j) = \frac{\ell_t(i,j)}{p_{\pi_1(s)}(i,j)} \mathbb{1}\{A_t(i) = j\} .$$

Since $\bar{\boldsymbol{p}}_{\pi_1(s)}$ follows the update rule shown in Eq. (21), using the standard analysis of FTRL (e.g. Theorem 5.2 in Hazan (2016)), we obtain

$$\begin{aligned} \spadesuit &= \mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle \right] \\ &\leqslant \frac{\log K}{\eta} + 2\eta \sum_{s=1}^{m_1} \sum_{j=1}^{K} \mathbb{E}\left[\bar{p}_{\pi_1(s)}(j) \left(\frac{1}{N} \sum_{t \in \mathcal{T}_{\pi_1(s)}} \sum_{i=1}^{N} \widehat{\ell}_t(i,j) \right)^2 \right] \end{aligned} \tag{25}$$

Let $\mathcal{F}_{\pi_1(s)}$ be the filtration of all random events observed up to the beginning of block $\pi_1(s)$. Since $\bar{p}_{\pi_1(s)}(j)$ is $\mathcal{F}_{\pi_1(s)}$-measurable, we know that

$$\begin{aligned} \mathbb{E}\Big[\bar{p}_{\pi_1(s)}(j) \bar{z}^2_{\pi_1(s)+1}(j) \Big] &= \mathbb{E}\Big[\mathbb{E}\big[\bar{p}_{\pi_1(s)}(j) \bar{z}^2_{\pi_1(s)+1}(j) \mid \mathcal{F}_{\pi_1(s)} \big] \Big] \\ &= \mathbb{E}\Big[\bar{p}^2_{\pi_1(s)}(j) \mathbb{E}\big[\bar{z}_{\pi_1(s)+1}(j) \mid \mathcal{F}_{\pi_1(s)} \big] \Big] . \end{aligned} \tag{26}$$

Conditioned on $\mathcal{F}_{\pi_1(s)}$, for all $t \in \mathcal{T}_{\pi_1(s)}$, direct calculation shows that

$$\mathbb{E}\left[\widehat{\ell}_t(i,j) \mid \mathcal{F}_{\pi_1(s)} \right] = \ell_t(i,j) \qquad \text{and} \qquad \mathbb{E}\left[\widehat{\ell}_t(i,j)^2 \mid \mathcal{F}_{\pi_1(s)} \right] = \frac{\ell_t^2(i,j)}{p_{\pi_1(s)}(i,j)} .$$

Since $A_t(i)$ are independently drawn across $t \in \mathcal{T}_{\pi_1(s)}$ and $i \in V$ conditioned on $\mathcal{F}_{\pi_1(s)}$, direct calculation shows that

$$\mathbb{E}\left[\left(\sum_{t \in \mathcal{T}_{\pi_1(s)}} \sum_{i=1}^{N} \widehat{\ell}_t(i,j) \right)^2 \middle| \mathcal{F}_{\pi_1(s)} \right]$$

$$
\begin{aligned}
&= \left(\mathbb{E}\left[\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\widehat{\ell}_t(i,j)\,\middle|\,\mathcal{F}_{\pi_1(s)}\right]\right)^2 + \mathbb{E}\left[\left(\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}(\widehat{\ell}_t(i,j)-\ell_t(i,j))\right)^2\,\middle|\,\mathcal{F}_{\pi_1(s)}\right] \\
&= \left(\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\ell_t(i,j)\right)^2 + \sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\mathbb{E}\left[\left(\widehat{\ell}_t(i,j)-\ell_t(i,j)\right)^2 \mid \mathcal{F}_{\pi_1(s)}\right] \\
&= \left(\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\ell_t(i,j)\right)^2 + \sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\left(\frac{\ell_t^2(i,j)}{p_{\pi_1(s)}(i,j)}-\ell_t^2(i,j)\right) \qquad (27)\\
&\leqslant (NB)^2 + \sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\frac{1}{p_{\pi_1(s)}(i,j)}
\end{aligned}
$$

where the second equality uses $\mathbb{E}\left[\widehat{\ell}_t(i,j) \mid \mathcal{F}_{\pi_1(s)}\right] = \ell_t(i,j)$ and the conditional independence of $A_t(i)$ among $t \in \mathcal{T}_{\pi_1(s)}$ and $i \in V$, and the last inequality holds because $\ell_t(i,j) \in [0,1]$. Hence,

$$
\begin{aligned}
\mathbb{E}\left[z_{\pi_1(s)+1}(j)^2 \mid \mathcal{F}_{\pi_1(s)}\right] &= \frac{1}{N^2}\,\mathbb{E}\left[\left(\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\widehat{\ell}_t(i,j)\right)^2\,\middle|\,\mathcal{F}_{\pi_1(s)}\right] \\
&\leqslant B^2 + \frac{1}{N^2}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\frac{1}{p_{\pi_1(s)}(i,j)}\,. \qquad (28)
\end{aligned}
$$

Combining Eq. (26) and Eq. (28) yields

$$
\begin{aligned}
\spadesuit &\leqslant \mathbb{E}\left[\sum_{s=1}^{m_1}\left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle\right] \\
&\leqslant \frac{\log K}{\eta} + 2\eta\sum_{s=1}^{m_1}\sum_{j=1}^{K}\bar{p}_{\pi_1(s)}(j)\left(B^2 + \frac{1}{N^2}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\frac{1}{p_{\pi_1(s)}(i,j)}\right) \\
&\leqslant \frac{\log K}{\eta} + 2\eta\left(m_1B^2 + \frac{3KBm_1}{N}\right) \qquad (29)
\end{aligned}
$$

where the last inequality uses $\sum_{j=1}^{K}\bar{p}_s(j) = 1$ and the fact that for all $s \in [m_1],\ i \in [N],\ j \in [K]$, with $\alpha = \frac{1}{T}$,

$$
\frac{\bar{p}_{\pi_1(s)}(j)}{p_{\pi_1(s)}(i,j)} = \frac{\bar{p}_{\pi_1(s)}(j)}{(1-\alpha)p'_{\pi_1(s)}(i,j) + \alpha/K} \leqslant 3, \qquad (30)
$$

where the equality is by the definition of $p_{\pi_1(s)}(i,j)$, and the inequality is due to Lemma 12.

Substituting Eq. (29), Eq. (23), and Eq. (24) into the decomposition Eq. (22), we obtain

$$
\mathrm{Reg}_{\mathcal{P}_1(i)} \leqslant \frac{\log K}{\eta} + 2\eta\left(m_1B^2 + \frac{3KBm_1}{N}\right) + 3.
$$

The analysis for $\mathrm{Reg}_{\mathcal{P}_0}(i)$ is identical. Summing the bounds over the two parity subsequences and using $m_1 + m_2 = M = T/B$ yields

$$\mathbb{E}\left[\sum_{\tau=1}^{M}\left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right] \leqslant \frac{\log K}{\eta} + 2\eta\left(TB + \frac{3KT}{N}\right) + 6$$
$$\leqslant 2\sqrt{2\log K\left(B + \frac{3K}{N}\right)T} + 6 \tag{31}$$

where the second inequality follows by choosing

$$\eta = \sqrt{\frac{\log K}{2(TB + \frac{3KT}{N})}}.$$

■

B.4. Omitted Proof Details for Lemma 12

Lemma 12 *Let $\mathcal{A}$ be Algorithm 2 and $\mathcal{B}$ be an instance of Algorithm 3 with a regularizer ψ that is 1-strongly convex w.r.t. the ℓ_2-norm. Suppose that each agent uses an instance of Algorithm 1 with κ and B defined in Eq. (4). Define $\bar{\boldsymbol{q}}^{(1)}_s$ and $\boldsymbol{q}^{(1)}_s(i)$ for all $i \in [N]$ as follows*

$$\boldsymbol{q}^{(1)}_s(i) = \operatorname*{arg\,min}_{\boldsymbol{q}\in\Delta(K)}\left\{\sum_{s'=1}^{s-1}\left\langle \boldsymbol{z}^B_{\pi_1(s')+1}(i), \boldsymbol{q}\right\rangle + \frac{1}{\eta}\psi(\boldsymbol{q})\right\}$$
$$\bar{\boldsymbol{q}}^{(1)}_s = \operatorname*{arg\,min}_{\boldsymbol{q}\in\Delta(K)}\left\{\sum_{s'=1}^{s-1}\left\langle \bar{\boldsymbol{z}}_{\pi_1(s')+1}, \boldsymbol{q}\right\rangle + \frac{1}{\eta}\psi(\boldsymbol{q})\right\}$$

where $\bar{\boldsymbol{z}}_\tau$ is defined in Eq. (3) and π_1 is defined in Eq. (19). Then we have

$$\|\bar{\boldsymbol{q}}^{(1)}_s - \boldsymbol{q}^{(1)}_s(i)\|_2 \leqslant \frac{2\eta B}{T^4K^5} \qquad \textit{for all } i \in [N]. \tag{32}$$

and

$$\frac{\bar{q}^{(1)}_s(k)}{(1-\alpha)q^{(1)}_s(i,k) + \alpha/K} \leqslant 3 \quad \textit{for all } i \in [N] \textit{ and } k \in [K]$$

where $\alpha = \frac{1}{T} \leqslant \frac{1}{2}$.

Proof Since ψ is 1-strongly convex w.r.t. $\|\cdot\|_2$, we have

$$\left\|\bar{\boldsymbol{q}}^{(1)}_s - \boldsymbol{q}^{(1)}_s(i)\right\|_2 \leqslant \eta\left\|\sum_{s'=1}^{s-1}\bar{\boldsymbol{z}}_{\pi_1(s')+1} - \sum_{s'=1}^{s-1}\boldsymbol{z}^B_{\pi_1(s')+1}(i)\right\|_2 \qquad \text{(Lemma 10)}$$
$$\leqslant \eta\sum_{s'=1}^{s-1}\left\|\bar{\boldsymbol{z}}_{\pi_1(s')+1} - \boldsymbol{z}^B_{\pi_1(s')+1}(i)\right\|_2 \qquad \text{(triangle inequality)}$$
$$\leqslant \frac{2\eta B}{T^4K^5} \qquad \text{(Lemma 1)}$$

$$\leqslant \frac{2}{KT} \,. \qquad (B \leqslant T)$$

To prove the second inequality, using $\|\cdot\|_\infty \leqslant \|\cdot\|_2$ we obtain

$$\left|\bar{q}_s^{(1)}(k) - q_s^{(1)}(i,k)\right| \leqslant \frac{2}{KT} \qquad \text{for all } k \in [K].$$

Now direct calculation shows that

$$\begin{aligned}
\frac{\bar{q}_s^{(1)}(k)}{(1-\alpha)q_s^{(1)}(i,k) + \alpha/K} &\leqslant \frac{q_s^{(1)}(i,k) + \frac{2}{KT}}{(1-\alpha)q_s^{(1)}(i,k) + \alpha/K} \\
&= 1 + \frac{\alpha\, q_s^{(1)}(i,k) + \frac{2}{KT} - \alpha/K}{(1-\alpha)q_s^{(1)}(i,k) + \alpha/K} \\
&\leqslant 1 + \max\left\{\frac{\alpha}{1-\alpha}, \frac{2}{\alpha T}\right\} = 3
\end{aligned}$$

where the last inequality uses the fact that $\alpha = 1/T \leqslant \frac{1}{2}$. ■

Appendix C. Omitted Details for Adaptive Bounds for Distributed K-armed Bandits

C.1. Omitted Proof Details for Theorem 5

In this section, we present the proof for Theorem 5, which provides the small- loss guarantee. For completeness, we restate Theorem 5 and provide its proof as follows.

Theorem 5 *Let $\mathcal{A}$ be Algorithm 2 run with base OLO algorithm $\mathcal{B}$ set to Algorithm 3 with $\psi_t(\boldsymbol{q}) = \frac{1}{\eta}\sum_{i=1}^K q(i)\log q(i) - \frac{1}{\gamma}\sum_{k=1}^K \log q(k)$ for all $t \in [T]$ where $\eta = \min\left\{\frac{1}{4B}, \sqrt{\frac{\log K}{BL^\star}}\right\}$ and $\gamma = \min\left\{\frac{N}{12}, \sqrt{\frac{KN\log T}{L^\star}}\right\}$. Assume all agents $i \in V$ run Algorithm 1 with bandit algorithm $\mathcal{A}$, gossip matrix W, and κ and B defined in Eq. (4). Then*

$$\begin{aligned}
\mathrm{Reg}_T &= \mathcal{O}\Big(\sqrt{BL^\star \log K} + \sqrt{\frac{KL^\star \log T}{N}} + B\log K + \frac{K\log T}{N}\Big) \\
&= \tilde{\mathcal{O}}\Big(\sqrt{\rho^{-1/2}L^\star} + \sqrt{\frac{KL^\star}{N}} + \rho^{-1/2} + \frac{K}{N}\Big) \,.
\end{aligned}$$

Proof We only show the steps that differ from the proof of Theorem 2, or more specifically Lemma 3 and Lemma 4. Note that the only algorithmic change compared to Theorem 2 is to use Algorithm 3 with a different regularizer as the subroutine $\mathcal{B}$. Since Lemma 1 and Lemma 3 is independent of the choice of $\mathcal{B}$, the conclusion from Lemma 1 and Lemma 3 still hold. To analyze the three terms ♣, ♡, and ♠ shown in the regret decomposition Eq. (22), since ψ is still 1-strongly convex with respect to ℓ_2-norm and $\boldsymbol{z}_{\pi_1(s)+1}(i)$ still follows the accelerated gossip dynamic, we still have ♣ $\leqslant 2$ and ♡ $\leqslant 1$. Therefore, we have

$$\mathrm{Reg}_{\mathcal{P}_1}(i) \leqslant \underbrace{\mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle\right]}_{\spadesuit} + 3 \,. \qquad (33)$$

It remains to analyze ♠ term $\mathbb{E}\left[\sum_{s=1}^{m_1}\left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle\right]$. Let $D_\psi(\cdot,\cdot)$ be the Bregman divergence based on ψ, and define $\widetilde{\boldsymbol{p}}_{\pi_1(s+1)} \triangleq \arg\min_{\boldsymbol{q}\in\mathbb{R}^d_{\geqslant 0}} \left\langle \boldsymbol{q}, \bar{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle + D_\psi(\boldsymbol{q}, \bar{\boldsymbol{p}}_{\pi_1(s)})$. Picking $\boldsymbol{p}^\star = \left(1-\frac{1}{T}\right)\boldsymbol{e}_k + \frac{1}{TK}\mathbf{1}$ and using Lemma 7.16 in Orabona (2025), we obtain

$$
\begin{aligned}
\spadesuit &= \mathbb{E}\left[\sum_{s=1}^{m_1}\left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle\right] \\
&= \mathbb{E}\left[\sum_{s=1}^{m_1}\left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k + \boldsymbol{p}^\star - \boldsymbol{p}^\star, \bar{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle\right] \\
&\leqslant \mathbb{E}\left[\sum_{s=1}^{m_1}\left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{p}^\star, \bar{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle\right] + 2 \\
&\leqslant \psi(\boldsymbol{p}^\star) - \psi(\bar{\boldsymbol{p}}_{\pi_1(1)}) + 2 \\
&\quad + \sum_{s=1}^{m_1}\sum_{j=1}^{K}\left(\frac{\eta\gamma y^2_{\pi_1(s)}(j)}{\gamma y_{\pi_1(s)}(j)+\eta}\right)\mathbb{E}\left[\bar{\boldsymbol{z}}^2_{\pi_1(s)+1}\right] \\
&\leqslant \frac{\log K}{\eta} + \frac{K\log T}{\gamma} + 2 \\
&\quad + \sum_{s=1}^{m_1}\sum_{j=1}^{K}\left(\frac{\eta\gamma y^2_{\pi_1(s)}(j)}{\gamma y_{\pi_1(s)}(j)+\eta}\right)\mathbb{E}\left[\bar{\boldsymbol{z}}^2_{\pi_1(s)+1}\right]
\end{aligned}
\tag{34}
$$

where $\boldsymbol{y}_{\pi_1(s)}$ lies on the line segment between $\bar{\boldsymbol{p}}_{\pi_1(s)}$ and $\widetilde{\boldsymbol{p}}_{\pi_1(s+1)}$, and the last inequality is because $\bar{\boldsymbol{p}}_{\pi_1(1)}$ is the uniform distribution. According to the update rule of $\widetilde{\boldsymbol{p}}_{\pi_1(s+1)}$, we know $\widetilde{\boldsymbol{p}}_{\pi_1(s+1)}$ satisfies that for all $j\in[K]$,

$$
\widetilde{p}_{\pi_1(s+1)}(j) + \log\widetilde{p}_{\pi_1(s+1)}(j) - \frac{\eta/\gamma}{\widetilde{p}_{\pi_1(s+1)}(j)} = \bar{p}_{\pi_1(s)}(j) + \log\bar{p}_{\pi_1(s)}(j) - \frac{\eta/\gamma}{\bar{p}_{\pi_1(s)}(j)} - \eta\bar{z}_{\pi_1(s)+1}(j),
$$

Since $x\mapsto x+\log x - c/x$ with $c>0$ is strictly increasing on $(0,\infty)$, $\bar{z}_{\pi_1(s)+1}(j)\geqslant 0$ implies that for all $j\in[K]$,

$$
\widetilde{p}_{\pi_1(s+1)}(j) \leqslant \bar{p}_{\pi_1(s)}(j),
$$

meaning that $\boldsymbol{y}_{\pi_1(s)}(j) \leqslant \bar{p}_{\pi_1(s)}(j)$. Plugging this inequality to Eq. (34) and using the fact that $\frac{x^2}{\gamma x+\eta}$ is increasing in x for $x\in(0,1)$, we have that

$$
\begin{aligned}
\spadesuit &\leqslant \frac{\log K}{\eta} + \frac{K\log T}{\gamma} + 2 + \sum_{s=1}^{m_1}\sum_{j=1}^{K}\left(\frac{\eta\gamma\bar{p}^2_{\pi_1(s)}(j)}{\gamma\bar{p}_{\pi_1(s)}(j)+\eta}\right)\mathbb{E}\left[\bar{\boldsymbol{z}}^2_{\pi_1(s)+1}\right] \\
&\leqslant \frac{\log K}{\eta} + \frac{K\log T}{\gamma} + 2 \qquad\qquad (35) \\
&\quad + \mathbb{E}\left[\frac{1}{N^2}\sum_{s=1}^{m_1}\sum_{j=1}^{K}\left(\frac{\eta\gamma\bar{p}^2_{\pi_1(s)}(j)}{\gamma\bar{p}_{\pi_1(s)}(j)+\eta}\right)\left(\left(\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\ell_t(i,j)\right)^2 + \sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\left(\frac{\ell_t^2(i,j)}{p_{\pi_1(s)}(i,j)}\right)\right)\right]
\end{aligned}
$$

(using Eq. (27))

$$
\begin{aligned}
&\leqslant \frac{\log K}{\eta} + \frac{K\log T}{\gamma} + 2 + \mathbb{E}\left[\frac{B}{N}\sum_{s=1}^{m_1}\sum_{j=1}^{K}\eta\bar{p}_{\pi_1(s)}(j)\left(\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\ell_t(i,j)\right)\right] \\
&\quad + \mathbb{E}\left[\sum_{s=1}^{m_1}\sum_{j=1}^{K}\gamma\bar{p}^2_{\pi_1(s)}(j)\left(\frac{1}{N^2}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\frac{\ell_t^2(i,j)}{p_{\pi_1(s)}(i,j)}\right)\right] \\
&\leqslant \frac{\log K}{\eta} + \frac{K\log T}{\gamma} + 2 + \mathbb{E}\left[\frac{B}{N}\sum_{s=1}^{m_1}\sum_{j=1}^{K}\eta\bar{p}_{\pi_1(s)}(j)\left(\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\ell_t(i,j)\right)\right] \\
&\quad + 3\mathbb{E}\left[\sum_{s=1}^{m_1}\sum_{j=1}^{K}\gamma\bar{p}_{\pi_1(s)}(j)\left(\frac{1}{N^2}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^{N}\ell_t(i,j)\right)\right] \\
&\leqslant \frac{\log K}{\eta} + \frac{K\log T}{\gamma} + 2 + \left(B\eta + \frac{3\gamma}{N}\right)\cdot\mathbb{E}\left[\sum_{s=1}^{m_1}\left\langle\bar{p}_{\pi_1(s)}, \bar{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle\right], \qquad (36)
\end{aligned}
$$

where the fourth inequality holds by using Eq. (30) induced by Lemma 12. Rearranging Eq. (36) and using definition of $\bar{\boldsymbol{z}}_{\pi_1(s)+1}$, we have

$$
\begin{aligned}
&(1 - B\eta - \frac{3\gamma}{N})\mathbb{E}\left[\sum_{s=1}^{m_1}\left\langle\bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle\right] \\
&\qquad \leqslant \frac{\log K}{\eta} + \frac{K\log T}{\gamma} + 2 + \left(B\eta + \frac{3\gamma}{N}\right)\sum_{s=1}^{m_1}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\bar{\ell}_t(k).
\end{aligned}
$$

Picking $\eta \leqslant \frac{1}{4B}$ and $\gamma \leqslant \frac{N}{12}$ and rearranging the terms lead to

$$
\mathbb{E}\left[\sum_{s=1}^{m_1}\left\langle\bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle\right] \leqslant \frac{2\log K}{\eta} + \frac{2K\log T}{\gamma} + 4 + \left(2B\eta + \frac{6\gamma}{N}\right)\sum_{s=1}^{m_1}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\bar{\ell}_t(k).
$$

The analysis for $\mathrm{Reg}_{\mathcal{P}_0}(i)$ is identical. Summing the bounds over the two parity subsequences, using $m_1 + m_2 = M = T/B$ and picking $\eta = \min\left\{\frac{1}{4B}, \sqrt{\frac{\log K}{BL^\star}}\right\}$ and $\gamma = \min\left\{\frac{N}{12}, \sqrt{\frac{KN\log T}{L^\star}}\right\}$, we obtain that

$$
\begin{aligned}
\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle\bar{\boldsymbol{p}}_\tau - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right] &\leqslant \left(\frac{4\log K}{\eta} + \frac{4K\log T}{\gamma} + 8\right) + 2\left(B\eta + \frac{3\gamma}{N}\right)\sum_{t=1}^{T}\bar{\ell}_t(k) \\
&\leqslant \mathcal{O}\left(\sqrt{BL^\star\log K} + \sqrt{\frac{KL^\star\log T}{N}} + B\log K + \frac{K\log T}{N}\right).
\end{aligned}
$$

Combining the bounds for ♡ and ♣, we have

$$
\max_{i\in V}\max_{k\in[K]}\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle\boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right]
$$

$$\leqslant \mathcal{O}\left(\sqrt{BL^\star \log K} + \sqrt{\frac{KL^\star \log T}{N}} + B\log K + \frac{K\log T}{N}\right).$$

Combining Lemma 3 with the above inequality finishes the proof. ■

C.2. Omitted Proof Details for Theorem 6

In this section, we present the proof for Theorem 6, which provides the Best of Both Worlds guarantee. For completeness, we restate Theorem 6 and provide its proof as follows.

Theorem 6 *Let $\mathcal{A}$ be Algorithm 2 run with base OLO algorithm $\mathcal{B}$ set to Algorithm 3 with $\psi_t(\boldsymbol{q}) = \frac{1}{\eta_t}\sum_{i=1}^K q(i)\log q(i) - \frac{2}{\gamma_t}\sum_{k=1}^K \sqrt{q(k)}$ where $\eta_t = \min\left\{\frac{1}{B}, \sqrt{\frac{\log K}{tB^2}}\right\}$ and $\gamma_t = \sqrt{\frac{N}{tB}}$ and. Assume all agents $i \in V$ run Algorithm 1 with bandit algorithm $\mathcal{A}$, gossip matrix W, and parameters κ and B defined in Eq. (4), Then, in the adversarial environment, the agents' regret is bounded as*

$$\mathrm{Reg}_T = \mathcal{O}\Big(\sqrt{BT\log K} + \sqrt{\frac{KT}{N}} + B\log K\Big) = \widetilde{\mathcal{O}}\Big(\sqrt{\rho^{-1/2}T} + \sqrt{\frac{KT}{N}} + \rho^{-1/2}\Big).$$

In the stochastic environment, the regret of the agents is bounded as

$$\mathrm{Reg}_T = \mathcal{O}\Big(\frac{B}{N}\sum_{k\neq k^\star}\frac{\log(T/B)}{\delta(k)} + \sum_{k\neq k^\star}\frac{B^2}{\delta(k)\log K} + B^2\log K\Big).$$

Proof We only show the steps that differ from the worst-case bound for distributed bandits. Changing $\mathcal{B}$ does not affect Lemma 1 and Lemma 3. Because $\psi_t(\boldsymbol{q})$ is 1-strongly convex w.r.t. the ℓ_2-norm, all the proof steps in Lemma 4 are the same , except ♠ terms in $\mathrm{Reg}_{\mathcal{P}_0}$ and $\mathrm{Reg}_{\mathcal{P}_1}$. We consider

$$\bar{\boldsymbol{p}}_{\pi_1(s)} \triangleq \operatorname*{arg\,min}_{\boldsymbol{q}\in\Delta(K)} \sum_{s'=1}^{s-1} \left\langle \widetilde{\boldsymbol{z}}_{\pi_1(s')+1}, \boldsymbol{q}\right\rangle + \psi_t(\boldsymbol{q})$$

where

$$\widetilde{z}_{\pi_1(s')+1}(k) = \frac{1}{N}\sum_{i=1}^N \sum_{t\in\mathcal{T}_{\pi_1(s)}} \left(\widehat{\ell}_t(i,k) - \ell_t(i,k^\star)\right) \tag{37}$$

and

$$\psi_t(\boldsymbol{q}) = \frac{1}{\eta_{\pi_1(s)}}\sum_{k=1}^K q(k)\log q(k) - \frac{2}{\gamma_t}\sum_{k=1}^K \sqrt{q(k)}.$$

Note that $\widetilde{\boldsymbol{z}}_{\pi_1(s')+1}(k) \geqslant -B$ for all $k \in [K]$ and $s' \in [m_1]$. This centering does not change the argmin, hence we also have

$$\bar{\boldsymbol{p}}_{\pi_1(s)+1} = \operatorname*{arg\,min}_{\boldsymbol{q}\in\Delta(K)} \sum_{s'=1}^{s-1} \left\langle \bar{\boldsymbol{z}}_{\pi_1(s')+1}, \boldsymbol{q}\right\rangle + \psi_t(\boldsymbol{q}).$$

We now analyze $\mathrm{Reg}_{\mathcal{P}_1}(i)$, and the analysis for $\mathrm{Reg}_{\mathcal{P}_0}(i)$ is analogous. Let $D_\psi(\cdot,\cdot)$ be the Bregman divergence based on ψ, and define $\widetilde{\boldsymbol{p}}_{\pi_1(s+1)} \triangleq \arg\min_{\boldsymbol{q}\in\mathbb{R}^d_{\geqslant 0}} \left\langle \boldsymbol{q}, \widetilde{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle + D_\psi\left(\boldsymbol{q}, \bar{\boldsymbol{p}}_{\pi_1(s)}\right)$. Using Lemma 7.16 in Orabona (2025), we obtain

$$
\begin{aligned}
\spadesuit &= \mathbb{E}\left[\sum_{s=1}^{m_1}\left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \widetilde{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle\right] && (38)\\
&= \mathbb{E}\left[\sum_{s=1}^{m_1}\left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1}\right\rangle\right]\\
&\leqslant \psi_{\pi_1(m_1)}(\boldsymbol{e}_k) - \psi_{\pi_1(1)}(\bar{p}_{\pi_1(1)}) + \sum_{s=1}^{m_1-1}\left(\psi_{\pi_1(s)}\left(\bar{\boldsymbol{p}}_{\pi_1(s+1)}\right) - \psi_{\pi_1(s+1)}\left(\bar{\boldsymbol{p}}_{\pi_1(s+1)}\right)\right)\\
&\quad + \sum_{s=1}^{m_1}\sum_{j=1}^{K}\left(\frac{2\eta_{\pi_1(s)}\gamma_t y^{3/2}_{\pi_1(s)}(j)}{\gamma_t\sqrt{y_{\pi_1(s)}(j)} + \eta_{\pi_1(s)}}\right)\mathbb{E}\left[\widetilde{\boldsymbol{z}}^2_{\pi_1(s)+1}(j)\right] && (39)
\end{aligned}
$$

where $\boldsymbol{y}_{\pi_1(s)}$ lies on the line segment between $\bar{\boldsymbol{p}}_{\pi_1(s)}$ and $\widetilde{\boldsymbol{p}}_{\pi_1(s+1)}$. Moreover, $\boldsymbol{y}_{\pi_1(s)}$ has the following update expression:

$$
\log y_{\pi_1(s)}(j) - \frac{\eta_{\pi_1(s)}/\gamma_{\pi_1(s)}}{\sqrt{y_{\pi_1(s)}(j)}} = \log \bar{p}_{\pi_1(s)}(j) - \frac{\eta_{\pi_1(s)}/\gamma_{\pi_1(s)}}{\sqrt{\bar{p}_{\pi_1(s)}(j)}} - \eta_{\pi_1(s)}\widetilde{z}_{\pi_1(s)+1}(j), \quad \text{for all } j\in[K].
$$

Rearranging the terms leads to

$$
\log y_{\pi_1(s)}(j) - \log \bar{p}_{\pi_1(s)}(j) \leqslant \frac{\eta_{\pi_1(s)}/\gamma_{\pi_1(s)}}{\sqrt{y_{\pi_1(s)}(j)}} - \frac{\eta_{\pi_1(s)}/\gamma_{\pi_1(s)}}{\sqrt{\bar{p}_{\pi_1(s)}(j)}} - \eta_{\pi_1(s)}\widetilde{z}_{\pi_1(s)+1}(j).
$$

We claim that this implies $y_{\pi_1(s)}(j) \leqslant 3\bar{p}_{\pi_1(s)}(j)$. To see this, we consider the following two cases. If $y_{\pi_1(s)}(j) \leqslant \bar{p}_{\pi_1(s)}(j)$, then we surely have $y_{\pi_1(s)}(j) \leqslant 3\bar{p}_{\pi_1(s)}(j)$. Otherwise, we know that $\log y_{\pi_1(s)}(j) - \log \bar{p}_{\pi_1(s)}(j) \leqslant -\eta_{\pi_1(s)}\widetilde{z}_{\pi_1(s)+1}(j) \leqslant \eta_{\pi_1(s)}B$, leading to the following

$$
y_{\pi_1(s)}(j) \leqslant e^{\eta_{\pi_1(s)}B}\cdot \bar{p}_{\pi_1(s)}(j) \leqslant 3\bar{p}_{\pi_1(s)}(j)
$$

where the last inequality holds because $\eta_{\pi_1(s)} \leqslant 1/B$. This concludes that $y_{\pi_1(s)}(j) \leqslant 3\bar{p}_{\pi_1(s)}(j)$. Plugging this inequality to Eq. (39), we have

$$
\begin{aligned}
\spadesuit &\leqslant \psi_{\pi_1(m_1)}(\boldsymbol{e}_k) - \psi_{\pi_1(1)}(\bar{\boldsymbol{p}}_{\pi_1(1)}) + \sum_{s=1}^{m_1-1}\left(\psi_{\pi_1(s)}\left(\bar{\boldsymbol{p}}_{\pi_1(s+1)}\right) - \psi_{\pi_1(s+1)}\left(\bar{\boldsymbol{p}}_{\pi_1(s+1)}\right)\right)\\
&\quad + 6\sum_{s=1}^{m_1}\sum_{j=1}^{K}\left(\frac{\eta_{\pi_1(s)}\gamma_t \bar{p}^{3/2}_{\pi_1(s)}(j)}{\gamma_t\sqrt{\bar{p}_{\pi_1(s)}(j)} + \eta_{\pi_1(s)}}\right)\mathbb{E}\left[\widetilde{z}^2_{\pi_1(s)+1}(j).\right] && (40)
\end{aligned}
$$

Next, we derive our regret guarantees for the adversarial and the stochastic environment separately.

Adversarial environment guarantees. We first analyze the first summation of Eq. (40).

$$
\begin{aligned}
&\psi_{\pi_1(m_1)}(\boldsymbol{e}_k) - \psi_{\pi_1(1)}(\bar{\boldsymbol{p}}_{\pi_1(1)}) + \sum_{s=1}^{m_1-1} \left(\psi_{\pi_1(s)}\left(\bar{\boldsymbol{p}}_{\pi_1(s+1)}\right) - \psi_{\pi_1(s+1)}\left(\bar{\boldsymbol{p}}_{\pi_1(s+1)}\right)\right) \\
&\leqslant \frac{\log K}{\eta_{\pi_1(1)}} + \frac{2\sqrt{K}}{\gamma_{\pi_1(1)}} - \frac{2}{\gamma_{\pi_1(m_1)}} + \sum_{s=1}^{m_1-1} \left(\psi_{\pi_1(s)}\left(\bar{\boldsymbol{p}}_{\pi_1(s+1)}\right) - \psi_{\pi_1(s+1)}\left(\bar{\boldsymbol{p}}_{\pi_1(s+1)}\right)\right) \\
&\leqslant \frac{\log K}{\eta_{\pi_1(1)}} + \frac{2\sqrt{K}-2}{\gamma_{\pi_1(1)}} \\
&\quad + 2\sum_{s=1}^{m_1-1} \left(\frac{1}{\gamma_{\pi_1(s+1)}} - \frac{1}{\gamma_{\pi_1(s)}}\right) \left(\sum_{j=1}^{K} \sqrt{\bar{p}_{\pi_1(s+1)}(j)} - 1\right) \\
&\quad - \sum_{s=2}^{m_1} \left(\frac{1}{\eta_{\pi_1(s)}} - \frac{1}{\eta_{\pi_1(s-1)}}\right) \sum_{j=1}^{K} \bar{p}_{\pi_1(s)}(j) \log \bar{p}_{\pi_1(s)}(j). \qquad (41)
\end{aligned}
$$

We first consider the first summation in Eq. (41). Recall the definition of γ_t and $\pi_1(s) = 2s$ for all $s \in [m_1]$, we have

$$
\frac{1}{\gamma_{\pi_1(s+1)}} - \frac{1}{\gamma_{\pi_1(s)}} \leqslant \sqrt{\frac{B}{N}} \cdot \frac{\pi_1(s+1) - \pi_1(s)}{\sqrt{\pi_1(s+1)}} \leqslant \sqrt{\frac{B}{N}} \cdot \frac{2}{\sqrt{\pi_1(s+1)}} \,.
$$

and

$$
\frac{1}{\eta_{\pi_1(s+1)}} - \frac{1}{\eta_{\pi_1(s)}} \leqslant \frac{B}{\sqrt{\log K}} \cdot \frac{\pi_1(s+1) - \pi_1(s)}{\sqrt{\pi_1(s+1)}} \leqslant \frac{B}{\sqrt{\log K}} \cdot \frac{2}{\sqrt{\pi_1(s+1)}} \,.
$$

Hence, we have

$$
\begin{aligned}
&2\sum_{s=1}^{m_1-1} \left(\frac{1}{\gamma_{\pi_1(s+1)}} - \frac{1}{\gamma_{\pi_1(s)}}\right) \left(\sum_{k=1}^{K} \sqrt{\bar{p}_{\pi_1(s+1)}(k)} - 1\right) \\
&\leqslant \frac{4\sqrt{B}}{\sqrt{N}} \sum_{s=1}^{m_1-1} \frac{1}{\sqrt{\pi_1(s+1)}} \left(\sum_{k=1}^{K} \sqrt{\bar{p}_{\pi_1(s+1)}(k)} - 1\right) \\
&\leqslant \frac{4\sqrt{B}}{\sqrt{N}} \sum_{s=1}^{m_1} \frac{1}{\sqrt{\pi_1(s)}} \left(\sum_{k=1}^{K} \sqrt{\bar{p}_{\pi_1(s)}(k)} - 1\right) \qquad (42) \\
&\leqslant \mathcal{O}\left(\sqrt{\frac{TK}{N}}\right). \qquad (43)
\end{aligned}
$$

Regarding the second summation in Eq. (41), we have

$$
\begin{aligned}
&-\sum_{s=2}^{m_1} \left(\frac{1}{\eta_{\pi_1(s)}} - \frac{1}{\eta_{\pi_1(s-1)}}\right) \sum_{j=1}^{K} \bar{p}_{\pi_1(s)}(j) \log \bar{p}_{\pi_1(s)}(j) \\
&\leqslant -\sum_{s=1}^{m_1-1} \frac{B}{\sqrt{\log K}} \cdot \frac{2}{\sqrt{\pi_1(s)}} \sum_{j=1}^{K} \bar{p}_{\pi_1(s)}(j) \log \bar{p}_{\pi_1(s)}(j) \qquad (44)
\end{aligned}
$$

$$\leqslant \mathcal{O}\left(\sqrt{BT\log K}\right), \tag{45}$$

where the last inequality is because $-\sum_{j=1}^{K}\bar{p}_{\pi_1(s)}(j)\log\bar{p}_{\pi_1(s)}(j)\leqslant\log K$. Plugging Eq. (43) and Eq. (45) to Eq. (41), we have

$$\begin{aligned}
&\psi_{\pi_1(m_1)}(\boldsymbol{e}_k)-\psi_{\pi_1(1)}(\bar{\boldsymbol{p}}_{\pi_1(1)})+\sum_{s=1}^{m_1-1}\left(\psi_{\pi_1(s)}\left(\bar{\boldsymbol{p}}_{\pi_1(s+1)}\right)-\psi_{\pi_1(s+1)}\left(\bar{\boldsymbol{p}}_{\pi_1(s+1)}\right)\right)\\
&\leqslant\frac{\log K}{\eta_{\pi_1(1)}}+\frac{2\sqrt{K}-2}{\gamma_{\pi_1(1)}}+\mathcal{O}\left(\sqrt{BT\log K}+\sqrt{\frac{TK}{N}}\right)\\
&\leqslant\mathcal{O}\left(B\log K+\sqrt{\frac{BK}{N}}+\sqrt{\frac{TK}{N}}+\sqrt{BT\log K}\right)\\
&\leqslant\mathcal{O}\left(B\log K+\sqrt{\frac{TK}{N}}+\sqrt{BT\log K}\right),
\end{aligned}\tag{46}$$

where the last inequality uses $B\leqslant T$.

Next, we analyze the second summation of Eq. (40). Let $\mathcal{F}_{\pi_1(s)}$ be the filtration of all random events observed up to the beginning of block $\pi_1(s)$. Since $\bar{p}_{\pi_1(s)}(j)$ is $\mathcal{F}_{\pi_1(s)}$-measurable, we know that

$$\begin{aligned}
\mathbb{E}\left[\tilde{z}^2_{\pi_1(s)+1}(k)\middle|\mathcal{F}_{\pi_1(s)}\right]&=\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{\pi_1(s)}}(\widehat{\ell}_t(i,k)-\ell_t(i,k^\star))\right)^2\middle|\mathcal{F}_{\pi_1(s)}\right]\\
&=\frac{1}{N^2}\mathbb{E}\left[\left(\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{\pi_1(s)}}(\widehat{\ell}_t(i,k)-\ell_t(i,k))\right)^2\middle|\mathcal{F}_{\pi_1(s)}\right]\\
&\quad+\frac{1}{N^2}\left(\mathbb{E}\left[\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{\pi_1(s)}}(\widehat{\ell}_t(i,k)-\ell_t(i,k^\star))\middle|\mathcal{F}_{\pi_1(s)}\right]\right)^2\\
&\qquad\qquad\text{(using }\mathbb{E}[x^2]=\mathbb{E}[(x-\mathbb{E}[x])^2]+\mathbb{E}[x]^2\text{)}\\
&=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\mathbb{E}\left[(\widehat{\ell}_t(i,k)-\ell_t(i,k))^2\middle|\mathcal{F}_{\pi_1(s)}\right]\\
&\quad+\frac{1}{N^2}\left(\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{\pi_1(s)}}(\ell_t(i,k)-\ell_t(i,k^\star))\right)^2\\
&\qquad\qquad\text{(since }A_t(i)\text{ are drawn independently over }i\in[N]\text{)}\\
&=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\ell_t^2(i,k)\left(\frac{1}{p_{\pi_1(s)}(i,k)}-1\right)\\
&\quad+\frac{1}{N^2}\left(\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{\pi_1(s)}}(\ell_t(i,k)-\ell_t(i,k^\star))\right)^2
\end{aligned}$$

$$= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_{\pi_1(s)}} \left(\frac{\ell_t^2(i,k)}{p_{\pi_1(s)}(i,k)} - 2\ell_t(i,k)\ell_t(i,k^\star) + \ell_t^2(i,k^\star) \right)$$
$$+ \frac{1}{N^2} \sum_{\substack{(i,t)\neq(j,t') \\ t,t' \in \mathcal{T}_{\pi_1(s)} \\ i,j \in [N]}} \left(\ell_t(i,k) - \ell_t(i,k^\star)\right)\left(\ell_{t'}(j,k) - \ell_{t'}(j,k^\star)\right) . \tag{47}$$

Plugging Eq. (47) and Eq. (46) in Eq. (40), we obtain

$$\spadesuit \leqslant \mathcal{O}\left(B \log K + \sqrt{\frac{TK}{N}} + \sqrt{BT \log K} \right)$$
$$+ \frac{6}{N^2} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{k=1}^{K} \bar{p}_{\pi_1(s)}^{3/2}(k) \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_{\pi_1(s)}} \left(\frac{\ell_t^2(i,k)}{p_{\pi_1(s)}(i,k)} - 2\ell_t(i,k)\ell_t(i,k^\star) + \ell_t^2(i,k^\star) \right)$$
$$+ \mathbb{E}\left[\frac{6}{N^2} \sum_{s=1}^{m_1} \eta_{\pi_1(s)} \sum_{k=1}^{K} \bar{p}_{\pi_1(s)}^{3/2}(k) \sum_{\substack{(i,t)\neq(j,t') \\ t,t' \in \mathcal{T}_{\pi_1(s)} \\ i,j \in [N]}} \left(\ell_t(i,k) - \ell_t(i,k^\star)\right)\left(\ell_{t'}(j,k) - \ell_{t'}(j,k^\star)\right) \right] . \tag{48}$$

Regarding the first summation in Eq. (48), we have

$$\frac{6}{N^2} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{k=1}^{K} \bar{p}_{\pi_1(s)}^{3/2}(k) \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_{\pi_1(s)}} \left(\frac{\ell_t^2(i,k)}{p_{\pi_1(s)}(i,k)} - 2\ell_t(i,k)\ell_t(i,k^\star) + \ell_t^2(i,k^\star) \right)$$
$$= \frac{6}{N^2} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_{\pi_1(s)}} \left(\bar{p}_{\pi_1(s)}^{3/2}(k) \frac{\ell_t^2(i,k)}{p_{\pi_1(s)}(i,k)} - \bar{p}_{\pi_1(s)}^{3/2}(k)\ell_t(i,k)\ell_t(i,k^\star) \right)$$
$$+ \frac{6}{N^2} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_{\pi_1(s)}} \bar{p}_{\pi_1(s)}^{3/2}(k) \left(\ell_t^2(i,k^\star) - \ell_t(i,k)\ell_t(i,k^\star) \right) . \tag{49}$$

For the first summation of Eq. (49), the terms corresponding to $k \neq k^\star$ are together bounded by

$$\frac{18B}{N} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{k \neq k^\star} \sqrt{\bar{p}_{\pi_1(s)}(k)} \tag{50}$$

by ignoring the second negative term and using the fact that $\frac{\bar{p}_{\pi_1(s)}(j)}{p_{\pi_1(s)}(i,j)} \leqslant 3$ according to Lemma 12 and the fact that $\ell_t(i,k) \in [0,1]$; the term corresponding to $k = k^\star$ is bounded as follows:

$$\frac{6}{N^2} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_{\pi_1(s)}} \left(\bar{p}_{\pi_1(s)}^{3/2}(k^\star) \frac{\ell_t^2(i,k^\star)}{p_{\pi_1(s)}(i,k^\star)} - \bar{p}_{\pi_1(s)}^{3/2}(k^\star)\ell_t^2(i,k^\star) \right)$$

$$
\begin{aligned}
&= \frac{6}{N^2} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_{\pi_1(s)}} \frac{\bar{p}_{\pi_1(s)}^{3/2}(k^\star)}{p_{\pi_1(s)}(i,k^\star)} \left(1 - p_{\pi_1(s)}(i,k^\star)\right) \ell_t^2(i,k^\star) \\
&\leqslant \frac{18B}{N^2} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{i=1}^{N} \left(1 - p_{\pi_1(s)}(i,k)\right) \qquad \text{(according to Lemma 12 and } |\mathcal{T}_{\pi_1(s)}| \leqslant B) \\
&= \frac{18B}{N^2} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{k \neq k^\star} \sum_{i=1}^{N} p_{\pi_1(s)}(i,k) \\
&\leqslant \frac{18B}{N} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{k \neq k^\star} \left(\bar{p}_{\pi_1(s)}(k) + \frac{2\eta B}{K^4 T^5} \right) \qquad \text{(using Lemma 12)} \\
&\leqslant \frac{18B}{N} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{k \neq k^\star} \sqrt{\bar{p}_{\pi_1(s)}(k)} + 36 \,. \qquad (51)
\end{aligned}
$$

Similarly, for the second summation of Eq. (49), the terms corresponding to $k \neq k^\star$ are together bounded by

$$
\frac{6B}{N} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{k \neq k^\star} \sqrt{\bar{p}_{\pi_1(s)}(k)} \qquad (52)
$$

again by ignoring the second negative term; the term corresponding to $k = k^\star$ is simply 0.

Regarding the second summation in Eq. (48), using $\ell_t(i,k) \in [0,1]$ we have

$$
\begin{aligned}
&\mathbb{E}\left[\frac{6}{N^2} \sum_{s=1}^{m_1} \eta_{\pi_1(s)} \sum_{k=1}^{K} \bar{p}_{\pi_1(s)}^{3/2}(k) \sum_{\substack{(i,t) \neq (j,t') \\ t,t' \in \mathcal{T}_{\pi_1(s)} \\ i,j \in [N]}} \left(\ell_t(i,k) - \ell_t(i,k^\star)\right) \left(\ell_{t'}(j,k) - \ell_{t'}(j,k^\star)\right) \right] \\
&\leqslant 6B^2 \sum_{s=1}^{m_1} \eta_{\pi_1(s)} \sum_{k \neq k^*} \bar{p}_{\pi_1(s)}(k) \,. \qquad (53)
\end{aligned}
$$

Plugging Eq. (49), Eq. (50), Eq. (51), Eq. (52) and Eq. (53) in Eq. (48), we have

$$
\begin{aligned}
\spadesuit &= \mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle \right] \\
&\leqslant \mathcal{O}\left(B \log K + \sqrt{\frac{TK}{N}} + \sqrt{BT \log K} \right) \\
&\quad + \frac{42B}{N} \sum_{s=1}^{m_1} \gamma_{\pi_1(s)} \sum_{k \neq k^\star} \sqrt{\bar{p}_{\pi_1(s)}(k)} + 6B^2 \sum_{s=1}^{m_1} \eta_{\pi_1(s)} \sum_{k \neq k^*} \bar{p}_{\pi_1(s)}(k) + 36 \,. \qquad (54)
\end{aligned}
$$

The analysis for $\mathrm{Reg}_{\mathcal{P}_0}(i)$ is identical. Summing the bounds over the two parity subsequences, using $m_1 + m_2 = M = T/B$ and using $\eta_t = \min\left\{\frac{1}{B}, \sqrt{\frac{\log K}{tB^2}}\right\}$ and $\gamma_t = \sqrt{\frac{N}{tB}}$, we obtain

$$\begin{aligned}
\mathbb{E}\left[\sum_{s=1}^{T/B} \langle \bar{\boldsymbol{p}}_{\pi_s} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{s+1}\rangle\right] &\leqslant \mathcal{O}\left(B\log K + \sqrt{\frac{TK}{N}} + \sqrt{BT\log K}\right) \\
&\quad + \frac{42B}{N}\sum_{s=1}^{T/B}\gamma_s \sum_{k\neq k^\star}\sqrt{\bar{p}_s(k)} + 6B^2\sum_{s=1}^{T/B}\eta_s\sum_{k\neq k^*}\bar{p}_s(k) \\
&\leqslant \mathcal{O}\left(\sqrt{BT\log K} + \sqrt{\frac{K}{N}T} + B\log K\right) , \qquad (55)
\end{aligned}$$

where the last inequality holds by bounding $\sum_k \sqrt{\bar{p}_s(k)}$ by $\sqrt{K}$. Combining the bounds for $\heartsuit$ and $\clubsuit$, we have

$$\max_{i\in V}\max_{k\in[K]}\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right] \leqslant \mathcal{O}\left(\sqrt{BT\log K} + \sqrt{\frac{K}{N}T} + B\log K\right)$$

Combining Lemma 3 with the above inequality finishes the proof of the adversarial bound.

Stochastic environment guarantees. When the losses are stochastic, we still analyze the terms in Eq. (40). For the first three terms in Eq. (40), according to Eq. (42), we know that

$$\begin{aligned}
&2\sum_{s=1}^{m_1-1}\left(\frac{1}{\gamma_{\pi_1(s+1)}} - \frac{1}{\gamma_{\pi_1(s)}}\right)\left(\sum_{k=1}^{K}\sqrt{\bar{p}_{\pi_1(s+1)}(k)} - 1\right) \\
&\leqslant \frac{4\sqrt{B}}{\sqrt{N}}\sum_{s=1}^{m_1}\frac{1}{\sqrt{\pi_1(s)}}\left(\sum_{k=1}^{K}\sqrt{\bar{p}_{\pi_1(s)}(k)} - 1\right) \\
&\leqslant \frac{4\sqrt{B}}{\sqrt{N}}\sum_{s=1}^{m_1}\frac{1}{\sqrt{\pi_1(s)}}\sum_{k\neq k^*}\sqrt{\bar{p}_{\pi_1(s)}(k)} . \qquad (56)
\end{aligned}$$

We also have the following according to Eq. (44)

$$\begin{aligned}
&-\sum_{s=2}^{m_1}\left(\frac{1}{\eta_{\pi_1(s)}} - \frac{1}{\eta_{\pi_1(s-1)}}\right)\sum_{j=1}^{K}\bar{p}_{\pi_1(s)}(j)\log\bar{p}_{\pi_1(s)}(j) \\
&\leqslant -\sum_{s=1}^{m_1-1}\frac{B}{\sqrt{\log K}}\cdot\frac{2}{\sqrt{\pi_1(s)}}\sum_{j=1}^{K}\bar{p}_{\pi_1(s)}(j)\log\bar{p}_{\pi_1(s)}(j) \\
&\leqslant \sum_{s=1}^{m_1} -\frac{B}{\sqrt{\log K}}\cdot\frac{2}{\sqrt{\pi_1(s)}}\sum_{k\neq k^\star}\bar{p}_{\pi_1(s)}(k)\log\bar{p}_{\pi_1(s)}(k) \\
&\quad + \sum_{s=1}^{m_1}\frac{B}{\sqrt{\log K}}\cdot\frac{2}{\sqrt{\pi_1(s)}}\sum_{k\neq k^\star}\bar{p}_{\pi_1(s)}(k) , \qquad (57)
\end{aligned}$$

where the second inequality holds using

$$-\bar{p}_{\pi_1(s)}(k^\star)\log\bar{p}_{\pi_1(s)}(k^\star) \leqslant (1-\bar{p}_{\pi_1(s)}(k^\star)) = \sum_{k\neq k^\star}\bar{p}_{\pi_1(s)}(k^\star)\,.$$

Therefore, we know that

$$\begin{aligned}
&\psi_{\pi_1(m_1)}(\boldsymbol{e}_k) - \psi_{\pi_1(1)}(\bar{\boldsymbol{p}}_{\pi_1(1)}) + \sum_{s=1}^{m_1-1}\left(\psi_{\pi_1(s)}\left(\bar{\boldsymbol{p}}_{\pi_1(s+1)}\right) - \psi_{\pi_1(s+1)}\left(\bar{\boldsymbol{p}}_{\pi_1(s+1)}\right)\right)\\
&\leqslant \frac{\log K}{\eta_{\pi_1(1)}} + \frac{2\sqrt{K}-2}{\gamma_{\pi_1(1)}} + \frac{4\sqrt{B}}{\sqrt{N}}\sum_{s=1}^{m_1}\frac{1}{\sqrt{\pi_1(s)}}\sum_{k\neq k*}\sqrt{\bar{p}_{\pi_1(s)}(k)}\\
&\quad + \sum_{s=1}^{m_1} -\frac{B}{\sqrt{\log K}}\cdot\frac{2}{\sqrt{\pi_1(s)}}\sum_{k\neq k^\star}\bar{p}_{\pi_1(s)}(k)\log\bar{p}_{\pi_1(s)}(k)\\
&\quad + \sum_{s=1}^{m_1}\frac{B}{\sqrt{\log K}}\cdot\frac{2}{\sqrt{\pi_1(s)}}\sum_{k\neq k^\star}\bar{p}_{\pi_1(s)}(k)\,. \qquad (58)
\end{aligned}$$

Plugging Eq. (49), Eq. (50), Eq. (51), Eq. (52), Eq. (53), Eq. (58) and Eq. (47) in Eq. (40), we obtain

$$\begin{aligned}
\spadesuit &= \frac{\log K}{\eta_{\pi_1(1)}} + \frac{2\sqrt{K}-2}{\gamma_{\pi_1(1)}} + \frac{4\sqrt{B}}{\sqrt{N}}\sum_{s=1}^{m_1}\frac{1}{\sqrt{\pi_1(s)}}\sum_{k\neq k*}\sqrt{\bar{p}_{\pi_1(s)}(k)}\\
&\quad + \sum_{s=1}^{m_1} -\frac{B}{\sqrt{\log K}}\cdot\frac{2}{\sqrt{\pi_1(s)}}\sum_{k\neq k^\star}\bar{p}_{\pi_1(s)}(k)\log\bar{p}_{\pi_1(s)}(k)\\
&\quad + \sum_{s=1}^{m_1}\frac{B}{\sqrt{\log K}}\cdot\frac{2}{\sqrt{\pi_1(s)}}\sum_{k\neq k^\star}\bar{p}_{\pi_1(s)}(k) + \frac{42B}{N}\sum_{s=1}^{m_1}\gamma_{\pi_1(s)}\sum_{k\neq k^\star}\sqrt{\bar{p}_{\pi_1(s)}(k)}\\
&\quad + \mathbb{E}\left[\frac{6}{N^2}\sum_{s=1}^{m_1}\eta_{\pi_1(s)}\sum_{k=1}^{K}\bar{p}^{3/2}_{\pi_1(s)}(k)\sum_{\substack{(i,t)\neq(j,t')\\ t,t'\in\mathcal{T}_{\pi_1(s)}\\ i,j\in[N]}}\left(\ell_t(i,k)-\ell_t\left(i,k^\star\right)\right)\left(\ell_{t'}(j,k)-\ell_{t'}\left(j,k^\star\right)\right)\right]
\end{aligned} \qquad (59)$$

Then we analyze the last summation of Eq. (59), we have

$$\begin{aligned}
&\mathbb{E}\left[\frac{6}{N^2}\sum_{s=1}^{m_1}\eta_{\pi_1(s)}\sum_{k=1}^{K}\bar{p}^{3/2}_{\pi_1(s)}(k)\sum_{\substack{(i,t)\neq(j,t')\\ t,t'\in\mathcal{T}_{\pi_1(s)}\\ i,j\in[N]}}\left(\ell_t(i,k)-\ell_t\left(i,k^\star\right)\right)\left(\ell_{t'}(j,k)-\ell_{t'}\left(j,k^\star\right)\right)\right]\\
&\leqslant \frac{6}{N^2}\sum_{s=1}^{m_1}\eta_{\pi_1(s)}\sum_{k=1}^{K}\bar{p}^{3/2}_{\pi_1(s)}(k)\sum_{\substack{(i,t)\neq(j,t')\\ t,t'\in\mathcal{T}_{\pi_1(s)}\\ i,j\in[N]}}\left(\mu(i,k)-\mu\left(i,k^\star\right)\right)\left(\mu(j,k)-\mu\left(j,k^\star\right)\right)
\end{aligned}$$

$$= \frac{6}{N^2} \sum_{s=1}^{m_1} \eta_{\pi_1(s)} \sum_{k=1}^{K} \bar{p}_{\pi_1(s)}^{3/2}(k) \Bigg[\Big(\sum_{t \in \mathcal{T}_{\pi_1(s)}} \sum_{i=1}^{N} \big(\mu(i,k) - \mu(i,k^\star)\big) \Big)^2 \tag{60}$$

$$- \sum_{t \in \mathcal{T}_{\pi_1(s)}} \sum_{i=1}^{N} \big(\mu(i,k) - \mu(i,k^\star)\big)^2 \Bigg]$$

$$= \frac{6}{N^2} \sum_{s=1}^{m_1} \eta_{\pi_1(s)} \sum_{k=1}^{K} \bar{p}_{\pi_1(s)}^{3/2}(k) \Bigg[|\mathcal{T}_{\pi_1(s)}|^2 \Big(\sum_{i=1}^{N} \big(\mu(i,k) - \mu(i,k^\star)\big) \Big)^2 \tag{61}$$

$$- |\mathcal{T}_{\pi_1(s)}| \sum_{i=1}^{N} \big(\mu(i,k) - \mu(i,k^\star)\big)^2 \Bigg]$$

$$\leqslant \frac{6}{N^2} \sum_{s=1}^{m_1} \eta_{\pi_1(s)} \sum_{k=1}^{K} \bar{p}_{\pi_1(s)}^{3/2}(k) \, |\mathcal{T}_{\pi_1(s)}|^2 \left(\sum_{i=1}^{N} \big(\mu(i,k) - \mu(i,k^\star)\big) \right)^2$$

$$= \frac{6}{N^2} \sum_{s=1}^{m_1} \eta_{\pi_1(s)} \sum_{k=1}^{K} \bar{p}_{\pi_1(s)}^{3/2}(k) |\mathcal{T}_{\pi_1(s)}|^2 N^2 \delta(k)^2$$

$$\leqslant 6B^2 \sum_{s=1}^{m_1} \eta_{\pi_1(s)} \sum_{k \neq k^\star} \bar{p}_{\pi_1(s)}(k) \delta(k) \tag{62}$$

Plugging Eq. (62) in Eq. (59), we have

$$\begin{aligned}
\spadesuit &= \mathbb{E}\left[\sum_{s=1}^{m_1} \big\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1} \big\rangle \right] \\
&\leqslant \frac{\log K}{\eta_{\pi_1(1)}} + \frac{2\sqrt{K} - 2}{\gamma_{\pi_1(1)}} + \frac{46\sqrt{B}}{\sqrt{N}} \sum_{s=1}^{m_1} \frac{1}{\sqrt{\pi_1(s)}} \sum_{k \neq k^\star} \sqrt{\bar{p}_{\pi_1(s)}(k)} + 36 \\
&\quad + \sum_{s=1}^{m_1} -\frac{B}{\sqrt{\log K}} \cdot \frac{2}{\sqrt{\pi_1(s)}} \sum_{k \neq k^\star} \bar{p}_{\pi_1(s)}(k) \log \bar{p}_{\pi_1(s)}(k) \\
&\quad + \sum_{s=1}^{m_1} \frac{B}{\sqrt{\log K}} \cdot \frac{2}{\sqrt{\pi_1(s)}} \sum_{k \neq k^\star} \bar{p}_{\pi_1(s)}(k) \\
&\quad + 6B \sum_{s=1}^{m_1} \sqrt{\frac{\log K}{\pi_1(s)}} \sum_{k \neq k^\star}^{K} \bar{p}_{\pi_1(s)}(k) \delta(k)
\end{aligned} \tag{63}$$

The analysis for $\mathrm{Reg}_{\mathcal{P}_0}(i)$ is identical. Summing the bounds over the two parity subsequences, using $m_1 + m_2 = M = T/B$, we obtain that

$$\mathbb{E}\left[\sum_{s=1}^{T/B} \langle \bar{\boldsymbol{p}}_s - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{s+1} \rangle \right] \leqslant \frac{\log K}{\eta_{\pi_0(1)}} + \frac{2\sqrt{K} - 2}{\gamma_{\pi_0(1)}} + \frac{\log K}{\eta_{\pi_1(1)}} + \frac{2\sqrt{K} - 2}{\gamma_{\pi_1(1)}} + 72$$

$$+ \frac{C}{2} \Bigg(\frac{\sqrt{B}}{\sqrt{N}} \sum_{s=1}^{T/B} \frac{1}{\sqrt{s}} \sum_{k \neq k^\star} \sqrt{\bar{p}_{\pi_1(s)}(k)}$$

$$
\begin{aligned}
&+ \sum_{s=1}^{T/B} -\frac{B}{\sqrt{\log K}} \cdot \frac{1}{\sqrt{s}} \sum_{k \neq k^\star} \bar{p}_s(k) \log \bar{p}_s(k) \\
&+ \sum_{s=1}^{T/B} \frac{B}{\sqrt{\log K}} \cdot \frac{1}{\sqrt{s}} \sum_{k \neq k^\star} \bar{p}_s(k) \\
&+ B \sum_{s=1}^{T/B} \sqrt{\frac{\log K}{s}} \sum_{k \neq k^\star} \bar{p}_s(k) \delta(k) \Bigg),
\end{aligned}
$$

where $C > 0$ is a universal constant. Note that by definition of $\delta(k)$, we have

$$
\mathbb{E}\left[\sum_{s=1}^{T/B} \langle \bar{\boldsymbol{p}}_s - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{s+1} \rangle\right] = \sum_{s=1}^{T/B} \sum_{k \neq k^\star} \bar{p}_s(k) \delta(k) \ .
$$

Therefore, we are able to rewrite the regret as follows:

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{s=1}^{T/B} \langle \bar{\boldsymbol{p}}_s - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{s+1} \rangle\right] \qquad (64) \\
&= 2\mathbb{E}\left[\sum_{s=1}^{T/B} \langle \bar{\boldsymbol{p}}_s - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{s+1} \rangle\right] - \mathbb{E}\left[\sum_{s=1}^{T/B} \langle \bar{\boldsymbol{p}}_s - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{s+1} \rangle\right] \\
&\leqslant \mathcal{O}\left(B \log K + \sqrt{\frac{BK}{N}}\right) \\
&\quad + C \frac{\sqrt{B}}{\sqrt{N}} \sum_{s=1}^{T/B} \frac{1}{\sqrt{s}} \sum_{k \neq k^\star} \sqrt{\bar{p}_{\pi_1(s)}(k)} - \frac{1}{4} \sum_{s=1}^{T/B} \sum_{k \neq k^\star} \bar{p}_s(k) \delta(k) \\
&\quad + C \sum_{s=1}^{T/B} -\frac{B}{\sqrt{\log K}} \cdot \frac{1}{\sqrt{s}} \sum_{k \neq k^\star} \bar{p}_s(k) \log \bar{p}_s(k) - \frac{1}{4} \sum_{s=1}^{T/B} \sum_{k \neq k^\star} \bar{p}_s(k) \Delta(k) \\
&\quad + C \sum_{s=1}^{T/B} \frac{B}{\sqrt{\log K}} \cdot \frac{1}{\sqrt{s}} \sum_{k \neq k^\star} \bar{p}_s(k) - \frac{1}{4} \sum_{s=1}^{T/B} \sum_{k \neq k^\star} \bar{p}_s(k) \delta(k) \\
&\quad + CB \sum_{s=1}^{T/B} \sqrt{\frac{\log K}{s}} \sum_{k \neq k^\star}^{K} \bar{p}_s(k) \delta(k) - \frac{1}{4} \sum_{s=1}^{T/B} \sum_{k \neq k^\star} \bar{p}_s(k) \delta(k). \qquad (65)
\end{aligned}
$$

Next, we analyze the four summations in Eq. (65) separately.

Bounding the first summation in Eq. (65). Direct calculation shows that

$$
\begin{aligned}
\sum_{s=1}^{T/B} \sum_{k \neq k^\star} \left(C \sqrt{\frac{B}{N}} \frac{1}{\sqrt{s}} \sqrt{\bar{p}_{\pi_1(s)}(k)} - \frac{1}{4} \bar{p}_{\pi_1(s)}(k) \delta(k) \right) &\leqslant \sum_{s=1}^{T/B} \sum_{k \neq k^\star} \frac{BC^2}{Ns\delta(k)} \\
&\leqslant \mathcal{O}\left(\frac{B}{N} \ln\left(\frac{T}{B}\right) \sum_{k \neq k^\star} \frac{1}{\delta(k)} \right) \ . \qquad (66)
\end{aligned}
$$

where the first inequality holds by using $2\sqrt{xy} - y \leqslant x$ for all $x, y \geqslant 0$.

Bounding the second summation in Eq. (65). We define

$$b_s \triangleq \frac{CB}{4\sqrt{s\log K}}\,.$$

Fix $k \neq k^\star$ and $\delta(k) > 0$. Define $g(z) \triangleq -b_s z \log z - \delta(k) z$. Direct calculation shows that $g(z)$ is concave and $z^\star \triangleq \arg\min_z g(z) = \exp(-\delta(k)/b_s - 1)$, hence

$$-b_s z \log z \leqslant \delta(k) z + g(z^\star) = \delta(k) z + b_s \exp\left(-\frac{\delta(k)}{b_s} - 1\right).$$

Summing over all $s \leqslant T/B$ shows that

$$\begin{aligned}
&\sum_{s=1}^{T/B} -b_s z \log z - \delta(k) z \\
&\leqslant \sum_{s=1}^{T/B} \sum_{k\neq k^\star} b_s \exp\Big(-\frac{\delta(k)}{b_s} - 1\Big) \\
&= \sum_{s=1}^{T/B} \sum_{k\neq k^\star} \frac{BC}{4\sqrt{s\log K}} \exp\Big(-\frac{4\delta(k)\sqrt{s\log K}}{BC} - 1\Big) \\
&\leqslant \sum_{k\neq k^\star} \int_0^\infty \frac{BC}{4\sqrt{x\log K}} \exp\Big(-\frac{4\delta(k)\sqrt{x\log K}}{BC} - 1\Big) dx \\
&\leqslant \mathcal{O}\left(\sum_{k\neq k^\star} \frac{B^2}{\delta(k)\log K}\right) \qquad (67)
\end{aligned}$$

where the last inequality holds by using $\int_0^\infty \frac{a}{\sqrt{x}} \exp(-\frac{\delta\sqrt{x}}{a} - 1)\, dx = \frac{2a^2}{e\delta}$ with $a = \frac{BC}{4\sqrt{\log K}}$ and $\delta = \delta(k)$.

Bounding the third summation in Eq. (65). Define $s^\star(k) \triangleq \left\lceil \frac{16C^2B^2}{\delta(k)^2 \log K} \right\rceil$. Then, we know that for all $s \geqslant s^\star(k)$,

$$C \cdot \frac{B}{\sqrt{\log K}} \cdot \frac{1}{\sqrt{s}} \bar{p}_s(k) - \frac{1}{4} \bar{p}_s(k) \delta(k) \leqslant 0$$

Therefore, it suffices to bound the third summation for s from 1 to $s^\star(k)$ for each $k \neq k^\star$. Then direct calculation shows that

$$\begin{aligned}
&\sum_{k\neq k^\star} \sum_{s=1}^{s^\star(k)} \left(\frac{CB}{\sqrt{\log K}} \cdot \frac{1}{\sqrt{s}} - \frac{1}{4} \bar{p}_s(k)\delta(k)\right) \\
&\leqslant \sum_{k\neq k^\star} \sum_{s=1}^{s^\star(k)} \frac{CB}{\sqrt{\log K}} \cdot \frac{1}{\sqrt{s}} \\
&= \sum_{k\neq k^\star} \sum_{s=1}^{\left\lceil \frac{16C^2B^2}{\delta(k)^2 \log K} \right\rceil} \frac{CB}{\sqrt{\log K}} \cdot \frac{1}{\sqrt{s}} \leqslant \mathcal{O}\left(\sum_{k\neq k^\star} \frac{B^2}{\delta(k)\log K}\right). \qquad (68)
\end{aligned}$$

Bounding the fourth summation in Eq. (65). Define $s^\star = \lceil 16C^2B^2 \log K \rceil$. Since when $s \geqslant s^\star$, we have

$$BC\sqrt{\frac{\log K}{s}}\bar{p}_s(k)\delta(k) - \frac{1}{4}\bar{p}_s(k)\delta(k) \leqslant 0 \,.$$

Therefore, it suffices to bound the fourth summation for s from 1 to $s^\star$. Direct calculation shows that

$$\begin{aligned}
&\sum_{s=1}^{s^\star}\sum_{k\neq k^\star}\left(BC\sqrt{\frac{\log K}{s}}\bar{p}_s(k)\delta(k) - \frac{1}{4}\bar{p}_s(k)\delta(k)\right) \\
&\leqslant \sum_{s=1}^{\lceil 16C^2B^2\log K\rceil}\sum_{k\neq k^\star} BC\sqrt{\frac{\log K}{s}}\bar{p}_s(k)\delta(k) \\
&\leqslant \sum_{s=1}^{\lceil 16C^2B^2\log K\rceil} BC\sqrt{\frac{\log K}{s}} \\
&\leqslant \mathcal{O}\left(B^2\log K\right). \qquad (69)
\end{aligned}$$

Plugging Eq. (66), Eq. (67), Eq. (68) and Eq. (69) in Eq. (65) and combining the bounds for ♡ and ♣, we have

$$\begin{aligned}
&\max_{i\in V}\max_{k\in[K]}\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right] \\
&\qquad \leqslant \mathcal{O}\left(\frac{B}{N}\ln\left(\frac{T}{B}\right)\sum_{k\neq k^\star}\frac{1}{\delta(k)} + \sum_{k\neq k^\star}\frac{B^2}{\delta(k)\log K} + B^2\log K\right)
\end{aligned}$$

Combining Lemma 3 with the above inequality finishes the proof.

■

Appendix D. Omitted Proof Details for Distributed Linear Bandit

D.1. Omitted Algorithm Description

In this section, we provide the omitted details regarding the description of the base algorithm $\mathcal{B}$ in Theorem 7. We utilize an instance of FTRL that first reconstructs the estimated loss for every original action $\boldsymbol{a}_k \in \Omega$ using the spanner decomposition coefficients $\boldsymbol{\lambda}^{(k)}$, and subsequently performs a standard FTRL update with entropy regularization.

D.2. Omitted Proof Details for Lemma 13

We begin by proving a consensus error bound for Algorithm 4, analogous to Lemma 1 in the MAB setting. Specifically, we first demonstrate that the local gossip vector at the end of each block concentrates around the network-wide averaged loss defined on the volumetric spanners. This result then implies a corresponding concentration for the reconstructed losses of the original actions.

Algorithm 5 FTRL for Linear Bandits

Input: Regularizer ψ, K-sized action set Ω, volumetric spanner $\mathcal{S} \subset [K]$ of Ω
Initialize: $\boldsymbol{q}_1 = \frac{1}{K}\mathbf{1} \in \Delta(K)$.
for $t = 1, \cdots T$ **do**
  Output $\boldsymbol{q}_t$ and receive $\boldsymbol{z}_t \in \mathbb{R}^{|\mathcal{S}|}$
  Construct $\widetilde{\boldsymbol{z}}_t \in \mathbb{R}^K$ where $\widetilde{z}_t(k) = \sum_{j=1}^{|\mathcal{S}|} \lambda^{(k)}(j) z_t(j)$ for all $k \in [K]$
  Update $\boldsymbol{q}_{t+1} = \arg\min_{\boldsymbol{q} \in \Delta(K)} \sum_{\tau \leqslant t} \langle \widetilde{\boldsymbol{z}}_\tau, \boldsymbol{q} \rangle + \psi(\boldsymbol{q})$
end

Lemma 13 *Assume all agents $i \in V$ run Algorithm 4 with an arbitrary linear bandits algorithm $\mathcal{A}$. Then, under the same assumptions as Theorem 7,*

$$\max_{i \in V} \max_{\tau \in [T/B]} \left\| \boldsymbol{z}_\tau^B(i) - \bar{\boldsymbol{z}}_\tau^{\mathcal{S}} \right\|_2 \leqslant \frac{2}{T^2 K^3} , \tag{70}$$

where $\bar{z}_\tau^{\mathcal{S}}(k) = \frac{1}{N} \sum_{i=1}^N \sum_{t \in \mathcal{T}_{\tau-1}} \left\langle \boldsymbol{b}_k, \widehat{\boldsymbol{\theta}}_t(i) \right\rangle$ for all $k \in [|\mathcal{S}|]$. Moreover, we also have

$$\max_{i \in V} \max_{\tau \in [T/B]} \left\| \widetilde{\boldsymbol{z}}_\tau^B(i) - \bar{\boldsymbol{z}}_\tau \right\|_2 \leqslant \frac{1}{K^{1.5} T^2} ,$$

where we recall that $\widetilde{z}_\tau^B(i,k) = \sum_{j=1}^{|\mathcal{S}|} \lambda^{(k)}(j)\, z_\tau^B(i,j)$ and with an abuse of notation, we define

$$\bar{z}_\tau(k) \triangleq \sum_{j=1}^{|\mathcal{S}|} \lambda^{(k)}(j) \bar{z}_\tau^{\mathcal{S}}(j), \tag{71}$$

for all $k \in [K]$.

Proof Following the analysis of Lemma 1 and Eq. (13), we know that

$$\left\| \boldsymbol{z}_\tau^B(i) - \bar{\boldsymbol{z}}_\tau^{\mathcal{S}} \right\|_2 \leqslant \frac{2\sqrt{\sum_{i=1}^N \left\| \sum_{t \in \mathcal{T}_{\tau-1}} \widehat{\ell}_t(i) \right\|_2^2}}{T^6 K^6 \sqrt{N}} .$$

According to Lemma 17, we have

$$\widehat{\ell}(i,k) = \left| \left\langle \boldsymbol{b}_k, \widehat{\boldsymbol{\theta}}_t(i) \right\rangle \right| \leqslant \frac{|\mathcal{S}|}{\beta} \qquad \text{for all } k \in [|\mathcal{S}|]. \tag{72}$$

Using $|\mathcal{S}| \leqslant K$, we have

$$\left\| \boldsymbol{z}_\tau^B(i) - \bar{\boldsymbol{z}}_\tau^{\mathcal{S}} \right\|_2 \leqslant \frac{2BK}{T^6 K^6 \beta} \leqslant \frac{2B}{T^6 K^5 \beta} \leqslant \frac{2}{T^2 K^3}.$$

For the last inequality, it suffices to show that $\beta \geqslant \frac{B}{2T^4K^2}$. Since we pick $\beta = 3Bd\eta$ and $\eta = \min\left\{ \frac{1}{6Bd}, \sqrt{\log K/(dTB + \frac{dT}{N})} \right\}$, using $d \leqslant K$, $B \leqslant T$, $N \geqslant 1$, and $K \geqslant 2$ (so $\log K \geqslant \log 2$), we have

$$\beta = 3Bd\sqrt{\frac{\log K}{dTB + \frac{dT}{N}}} \geqslant 3Bd\sqrt{\frac{\log 2}{KT^2 + TK}}.$$

Moreover, for $T \geqslant 2$ and $K \geqslant 2$, we have $\sqrt{\frac{\log 2}{KT^2+TK}} \geqslant \frac{\log 2}{2T^2K}$, hence $\beta \geqslant \frac{3Bd\log 2}{2T^2K} \geqslant \frac{B}{2T^4K^2}$.

Next, we bound the reconstruction error on the original action space. Recall from Algorithm 5 that

$$\widetilde{z}_\tau^B(i,k) = \sum_{j=1}^{|\mathcal{S}|} \lambda^{(k)}(j)\, z_\tau^B(i,j), \qquad \forall k \in [K]\,.$$

Note that $\widetilde{\boldsymbol{z}}_\tau^B(i) \in \mathbb{R}^K$ while $\boldsymbol{z}_\tau^B(i) \in \mathbb{R}^{|\mathcal{S}|}$. Then, for each $k \in [K]$,

$$\begin{aligned}
\bar{z}_\tau(k) &= \sum_{j=1}^{|\mathcal{S}|} \lambda^{(k)}(j)\, \bar{z}_\tau^{\mathcal{S}}(j) \\
&= \sum_{j=1}^{|\mathcal{S}|} \lambda^{(k)}(j) \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_{\tau-1}} \left\langle \boldsymbol{b}_j, \widehat{\boldsymbol{\theta}}_t(i) \right\rangle \right) \\
&= \frac{1}{N} \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_{\tau-1}} \left\langle \sum_{j=1}^{|\mathcal{S}|} \lambda^{(k)}(j) \boldsymbol{b}_j, \widehat{\boldsymbol{\theta}}_t(i) \right\rangle \\
&= \frac{1}{N} \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_{\tau-1}} \left\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i) \right\rangle. \qquad (73)
\end{aligned}$$

Consequently, for any $k \in [K]$,

$$\begin{aligned}
\left|\bar{z}_\tau(k) - \widetilde{z}_\tau^B(i,k)\right| &= \left| \sum_{j=1}^{|\mathcal{S}|} \lambda^{(k)}(j) \bar{z}_\tau^{\mathcal{S}}(j) - \sum_{j=1}^{|\mathcal{S}|} \lambda^{(k)}(j) z_\tau^B(i,j) \right| \\
&= \left| \sum_{j=1}^{|\mathcal{S}|} \lambda^{(k)}(j) \left(\bar{z}_\tau^{\mathcal{S}}(j) - z_\tau^B(i,j) \right) \right| \\
&= \left| \left\langle \boldsymbol{\lambda}^{(k)}, \bar{\boldsymbol{z}}_\tau^{\mathcal{S}} - \boldsymbol{z}_\tau^B(i) \right\rangle \right| \\
&\leqslant \|\boldsymbol{\lambda}^{(k)}\|_2 \, \left\| \bar{\boldsymbol{z}}_\tau^{\mathcal{S}} - \boldsymbol{z}_\tau^B(i) \right\|_2 \\
&\leqslant \left\| \bar{\boldsymbol{z}}_\tau^{\mathcal{S}} - \boldsymbol{z}_\tau^B(i) \right\|_2 \leqslant \frac{2}{T^2K^3}\,,
\end{aligned}$$

where we used Cauchy–Schwarz and the volumetric spanner property $\|\boldsymbol{\lambda}^{(k)}\|_2 \leqslant 1$. Therefore,

$$\left\| \bar{\boldsymbol{z}}_\tau - \widetilde{\boldsymbol{z}}_\tau^B(i) \right\|_2 \leqslant \sqrt{K} \cdot \max_{k \in [K]} \left| \bar{z}_\tau(k) - \widetilde{z}_\tau^B(i,k) \right| \leqslant \frac{2\sqrt{K}}{T^2K^3} = \frac{2}{T^2K^{5/2}} \leqslant \frac{1}{K^{1.5}T^2}\,,$$

where the last inequality holds by using $K \geqslant 2$. ■

D.3. Omitted Proof Details for Lemma 14

In this subsection, we provide the proof of Lemma 14, which is the analog of Lemma 3 in the MAB setting.

Lemma 14 *Assume all agents $i \in V$ run Algorithm 4 with a delayed linear bandits algorithm $\mathcal{A}$ whose predictions $\boldsymbol{p}'_1(i), \dots, \boldsymbol{p}'_{T/B}(i)$ for each agent $i \in V$ satisfy*

$$\max_{i \in V} \max_{k \in [K]} \mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \widetilde{\boldsymbol{z}}^B_{\tau+1}(i) \right\rangle\right] \leqslant \mathsf{R}^{\mathrm{del}} \tag{74}$$

for some $\mathsf{R}^{\mathrm{del}} > 0$. Then, under the same assumptions as Theorem 7, the agents' regret satisfies

$$\max_{i \in V} \mathrm{Reg}_T(i) \leqslant \mathsf{R}^{\mathrm{del}} + 6\sqrt{dBT \log K} + 3\,.$$

Proof Recall that Algorithm 4 uses $\boldsymbol{p}_\tau(i) = (1-\alpha-\beta)\boldsymbol{p}'_\tau(i) + \frac{\alpha}{K}\mathbf{1} + \frac{\beta}{|\mathcal{S}|}\mathbf{1}_{\mathcal{S}}$, where $\boldsymbol{p}'_\tau(i) \in \Delta(K)$ is the distribution output by $\mathcal{A}$. Fix an agent $i \in V$, and let

$$k \in \operatorname*{arg\,max}_{k' \in [K]} \mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{e}_{k'}, \sum_{t \in \mathcal{T}_\tau} \bar{\boldsymbol{\ell}}_t \right\rangle\right].$$

Then we can write

$$\begin{aligned}
\mathrm{Reg}_T(i) &= \mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}_\tau(i) - \boldsymbol{e}_k, \sum_{t \in \mathcal{T}_\tau} \bar{\boldsymbol{\ell}}_t \right\rangle\right] \\
&= \mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} \right\rangle\right] && \text{(by Eq. (3))} \\
&= \mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle (1-\alpha-\beta)\,\boldsymbol{p}'_\tau(i) + \frac{\alpha}{K}\mathbf{1} + \frac{\beta}{|\mathcal{S}|}\mathbf{1}_{\mathcal{S}} - \boldsymbol{e}_k, \widetilde{\boldsymbol{z}}^B_{\tau+1}(i) + \bar{\boldsymbol{z}}_{\tau+1} - \widetilde{\boldsymbol{z}}^B_{\tau+1}(i) \right\rangle\right] \\
&= (1-\alpha-\beta)\,\mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \widetilde{\boldsymbol{z}}^B_{\tau+1}(i) \right\rangle\right] \\
&\quad + (1-\alpha-\beta)\,\mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} - \widetilde{\boldsymbol{z}}^B_{\tau+1}(i) \right\rangle\right] \\
&\quad + \alpha\mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \frac{1}{K}\mathbf{1} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} \right\rangle\right] \\
&\quad + \beta\mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \frac{1}{|\mathcal{S}|}\mathbf{1}_{\mathcal{S}} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} \right\rangle\right] \\
&\leqslant \mathsf{R}^{\mathrm{del}} && \text{(by Eq. (74))} \\
&\quad + (1-\alpha-\beta)\,\mathbb{E}\left[\sum_{\tau=1}^{T/B} \left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} - \widetilde{\boldsymbol{z}}^B_{\tau+1}(i) \right\rangle\right] && (75)
\end{aligned}$$

$$+ \alpha \mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle \frac{1}{K}\mathbf{1} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right] \tag{76}$$

$$+ \beta \mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle \frac{1}{|\mathcal{S}|}\mathbf{1}_{\mathcal{S}} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right]. \tag{77}$$

Bounding Eq. (75). By the Cauchy–Schwarz inequality, for each block τ,

$$\left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} - \widetilde{\boldsymbol{z}}^B_{\tau+1}(i)\right\rangle \leqslant \left\|\boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k\right\|_2 \left\|\bar{\boldsymbol{z}}_{\tau+1} - \widetilde{\boldsymbol{z}}^B_{\tau+1}(i)\right\|_2.$$

Since $\boldsymbol{p}'_\tau(i) \in \Delta(K)$, we have $\|\boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k\|_2 \leqslant \sqrt{2}$. Using Lemma 13,

$$\left\|\bar{\boldsymbol{z}}_{\tau+1} - \widetilde{\boldsymbol{z}}^B_{\tau+1}(i)\right\|_2 \leqslant \frac{1}{K^{1.5}T^2} \leqslant \frac{1}{KT}.$$

Therefore,

$$\left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} - \widetilde{\boldsymbol{z}}^B_{\tau+1}(i)\right\rangle \leqslant \frac{\sqrt{2}}{KT}.$$

Taking expectations and summing over $\tau = 1, \ldots, T/B$ yields

$$(1-\alpha-\beta)\,\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1} - \widetilde{\boldsymbol{z}}^B_{\tau+1}(i)\right\rangle\right] \leqslant \frac{T}{B}\cdot\frac{\sqrt{2}}{KT} = \frac{\sqrt{2}}{BK} \leqslant 1, \tag{78}$$

where the last inequality uses $K \geqslant 2$ and $B \geqslant 1$.

Bounding Eq. (76) and Eq. (77). Taking expectations and using unbiasedness of $\widehat{\boldsymbol{\theta}}_t(i)$ gives

$$\mathbb{E}[\bar{z}_{\tau+1}(k)] = \frac{1}{N}\sum_{t\in\mathcal{T}_\tau}\sum_{i=1}^{N}\langle \boldsymbol{a}_k, \theta_t(i)\rangle = \frac{1}{N}\sum_{t\in\mathcal{T}_\tau}\sum_{i=1}^{N}\ell_t(i,k).$$

Since $\ell_t(i,k) \in [-1,1]$ and $|\mathcal{T}_\tau| = B$, we have

$$-B \leqslant \mathbb{E}[\bar{z}_{\tau+1}(k)] \leqslant B, \quad \forall k \in [K].$$

Then, by linearity of expectation,

$$\mathbb{E}\left[\left\langle \frac{1}{K}\mathbf{1} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right] = \left\langle \frac{1}{K}\mathbf{1} - \boldsymbol{e}_k, \mathbb{E}[\bar{\boldsymbol{z}}_{\tau+1}]\right\rangle = \left\langle \frac{1}{K}\mathbf{1}, \mathbb{E}[\bar{\boldsymbol{z}}_{\tau+1}]\right\rangle - \langle \boldsymbol{e}_k, \mathbb{E}[\bar{\boldsymbol{z}}_{\tau+1}]\rangle.$$

Hence, we have

$$\mathbb{E}\left[\left\langle \frac{1}{K}\mathbf{1} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right] \leqslant 2B.$$

The analysis for Eq. (77) is similar, and we obtain

$$\mathbb{E}\left[\left\langle \frac{1}{|\mathcal{S}|}\mathbf{1}_{\mathcal{S}} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right] \leqslant 2B.$$

Combining the bounds for both terms, we know that

$$\alpha\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle\frac{1}{K}\mathbf{1}-\boldsymbol{e}_k,\bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right]+\beta\mathbb{E}\left[\sum_{\tau=1}^{T/B}\left\langle\frac{1}{|\mathcal{S}|}\mathbf{1}_{\mathcal{S}}-\boldsymbol{e}_k,\bar{\boldsymbol{z}}_{\tau+1}\right\rangle\right]\leqslant 2\alpha T+2\beta T$$
$$\leqslant 2+6\sqrt{dBT\log K}, \quad (79)$$

where we used $\alpha=\frac{1}{T}$ and $\beta=3Bd\eta=3Bd\cdot\min\left\{\frac{1}{6Bd},\sqrt{\frac{\log K}{dTB+\frac{dT}{N}}}\right\}$ in the last inequality. Plugging Eq. (78) and Eq. (79) into the decomposition finishes the proof.

■

D.4. Omitted Proof Details for Lemma 15

In this subsection, we prove Lemma 15, which adapts the argument of Lemma 4 from MAB setting to the linear bandits setting.

Lemma 15 *Under the same assumptions as in Theorem 7, Algorithm 4 guarantees*

$$\max_{i\in V}\max_{k\in[K]}\mathbb{E}\left[\sum_{\tau=1}^{M}\left\langle\boldsymbol{p}'_\tau(i)-\boldsymbol{e}_k,\tilde{\boldsymbol{z}}^B_{\tau+1}(i)\right\rangle\right]\leqslant\mathcal{O}\left(\sqrt{\log K\left(B+\frac{1}{N}\right)dT}+Bd\log K\right). \quad (80)$$

Proof The analysis is similar to Lemma 4; therefore, we use the same notation, including the parity sets ($\mathcal{P}_0$ and $\mathcal{P}_1$) and the associated index maps ($\pi_0(s)$ and $\pi_1(s)$). Fix an agent $i\in V$ and $k\in[K]$. We decompose the regret as follows:

$$\mathbb{E}\left[\sum_{\tau=1}^{M}\left\langle\boldsymbol{p}'_\tau(i)-\boldsymbol{e}_k,\tilde{\boldsymbol{z}}^B_{\tau+1}(i)\right\rangle\right]$$
$$=\underbrace{\mathbb{E}\left[\sum_{s=1}^{m_0}\left\langle\boldsymbol{p}'_{\pi_0(s)}(i)-\boldsymbol{e}_k,\tilde{\boldsymbol{z}}^B_{\pi_0(s)+1}(i)\right\rangle\right]}_{\triangleq\mathrm{Reg}_{\mathcal{P}_0}(i)}+\underbrace{\mathbb{E}\left[\sum_{s=1}^{m_1}\left\langle\boldsymbol{p}'_{\pi_1(s)}(i)-\boldsymbol{e}_k,\tilde{\boldsymbol{z}}^B_{\pi_1(s)+1}(i)\right\rangle\right]}_{\triangleq\mathrm{Reg}_{\mathcal{P}_1}(i)}.$$

We now analyze $\mathrm{Reg}_{\mathcal{P}_1}(i)$; the analysis for $\mathrm{Reg}_{\mathcal{P}_0}(i)$ is analogous. By the update rule of $\boldsymbol{q}_t$ in Algorithm 5,

$$\boldsymbol{p}'_{\pi_1(s)}(i)=\arg\min_{\boldsymbol{q}\in\Delta(K)}\left\{\sum_{s'=1}^{s-1}\left\langle\tilde{\boldsymbol{z}}^B_{\pi_1(s')+1}(i),\boldsymbol{q}\right\rangle+\frac{1}{\eta}\sum_{k=1}^{K}q(k)\log\left(q(k)\right)\right\}. \quad (81)$$

Recall that $\bar{\boldsymbol{z}}_\tau$ is defined in Eq. (71) and derivation in Eq. (73) shows that

$$\bar{\boldsymbol{z}}_{\pi_1(s)+1}=\frac{1}{N}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\left\langle\boldsymbol{a}_k,\widehat{\boldsymbol{\theta}}_t(i)\right\rangle.$$

Define

$$\bar{\boldsymbol{p}}_{\pi_1(s)} \triangleq \arg\min_{\boldsymbol{q}\in\Delta(K)} \left\{ \sum_{s'=1}^{s-1} \left\langle \bar{\boldsymbol{z}}_{\pi_1(s')+1}, \boldsymbol{q} \right\rangle + \frac{1}{\eta} \sum_{k=1}^{K} q(k) \log\left(q(k)\right) \right\} , \tag{82}$$

which is the strategy output by FTRL when fed with loss vectors $\{\bar{\boldsymbol{z}}_{\pi_1(s')+1}\}_{s'\in[s-1]}$. Similar to the regret decomposition in Lemma 4, we have

$$\begin{aligned}
\mathrm{Reg}_{\mathcal{P}_1}(i) &= \mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \boldsymbol{p}'_{\pi_1(s)}(i) - \boldsymbol{e}_k, \tilde{\boldsymbol{z}}^B_{\pi_1(s)+1}(i) \right\rangle \right] \\
&= \underbrace{\mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \boldsymbol{p}'_{\pi_1(s)}(i) - \boldsymbol{e}_k, \tilde{\boldsymbol{z}}^B_{\pi_1(s)+1}(i) - \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle \right]}_{\heartsuit} \\
&\quad + \underbrace{\mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \boldsymbol{p}'_{\pi_1(s)}(i) - \bar{\boldsymbol{p}}_{\pi_1(s)}, \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle \right]}_{\clubsuit} \\
&\quad + \underbrace{\mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle \right]}_{\spadesuit} .
\end{aligned} \tag{83}$$

Bounding the term ♣. By Cauchy–Schwarz inequality, for each $s \in [m_1]$,

$$\left\langle \boldsymbol{p}'_{\pi_1(s)}(i) - \bar{\boldsymbol{p}}_{\pi_1(s)}, \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle \leqslant \|\boldsymbol{p}'_{\pi_1(s)}(i) - \bar{\boldsymbol{p}}_{\pi_1(s)}\|_2 \cdot \|\bar{\boldsymbol{z}}_{\pi_1(s)+1}\|_2 .$$

Since $\boldsymbol{p}'_{\pi_1(s)}(i)$ and $\bar{\boldsymbol{p}}_{\pi_1(s)}$ follow the update rule of Eq. (81) and Eq. (82) and $\psi(\boldsymbol{q}) = \frac{1}{\eta}\sum_{k=1}^{K} q(k)\log(q(k))$ is $\frac{1}{\eta}$-strongly convex w.r.t. ℓ_2-norm, Lemma 16 yields that

$$\|\boldsymbol{p}'_{\pi_1(s)}(i) - \bar{\boldsymbol{p}}_{\pi_1(s)}\|_2 \leqslant \frac{\eta}{K^{1.5}T^2} .$$

Moreover, since $\left|\widehat{\ell}_t(i,k)\right| \leqslant \frac{|\mathcal{S}|}{\beta}$ for all $k \in [|\mathcal{S}|]$ by Lemma 17, we have $|\bar{z}_{\pi_1(s)+1}(k)| \leqslant \frac{B|\mathcal{S}|}{\beta}$ for all $k \in [K]$. Hence,

$$\|\bar{\boldsymbol{z}}_{\pi_1(s)+1}\|_2 \leqslant \frac{B|\mathcal{S}|\sqrt{K}}{\beta} .$$

Therefore,

$$\clubsuit \leqslant \sum_{s=1}^{m_1} \mathbb{E}\Big[\|\boldsymbol{p}'_{\pi_1(s)}(i) - \bar{\boldsymbol{p}}_{\pi_1(s)}\|_2 \cdot \|\bar{\boldsymbol{z}}_{\pi_1(s)+1}\|_2 \Big] \leqslant m_1 \cdot \frac{\eta}{K^{1.5}T^2} \cdot \frac{B|\mathcal{S}|\sqrt{K}}{\beta} \leqslant 1 \tag{84}$$

where we use $\beta = 3Bd\eta$, $|\mathcal{S}| \leqslant 3d$ and $m_1 \leqslant T$.

Bounding the term ♡. By Cauchy-Schwarz inequality, for each $s \in [m_1]$, we have

$$\left\langle \boldsymbol{p}'_{\pi_1(s)}(i) - \boldsymbol{e}_k, \widetilde{\boldsymbol{z}}^B_{\pi_1(s)+1}(i) - \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle \leqslant \left\| \boldsymbol{p}'_{\pi_1(s)}(i) - \boldsymbol{e}_k \right\|_2 \cdot \left\| \widetilde{\boldsymbol{z}}^B_{\pi_1(s)+1}(i) - \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\|_2 .$$

Since $\boldsymbol{p}'_{\pi_1(s)}(i) \in \Delta(K)$, we have $\|\boldsymbol{p}'_{\pi_1(s)}(i) - \boldsymbol{e}_k\|_2 \leqslant \sqrt{2}$. Moreover, according to Lemma 13, we know that

$$\left\| \widetilde{\boldsymbol{z}}^B_{\pi_1(s)+1}(i) - \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\|_2 \leqslant \frac{1}{K^{1.5}T^2} .$$

Therefore,

$$\heartsuit = \mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \boldsymbol{p}'_{\pi_1(s)}(i) - \boldsymbol{e}_k, \widetilde{\boldsymbol{z}}^B_{\pi_1(s)+1}(i) - \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle \right] \leqslant \frac{\sqrt{2}m_1}{K^{1.5}T^2} \leqslant 2 , \tag{85}$$

where the last inequality uses $m_1 \leqslant T$.

Bounding the term ♠. Recall that $\bar{z}_{\pi_1(s)+1} \in \mathbb{R}^K$ and for all $k \in [K]$,

$$\bar{z}_{\pi_1(s)+1}(k) = \frac{1}{N} \sum_{t \in \mathcal{T}_{\pi_1(s)}} \sum_{i=1}^{N} \left\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i) \right\rangle ,$$

where

$$\widehat{\boldsymbol{\theta}}_t(i) = M_{\pi_1(s)}(i)^{-1} \boldsymbol{a}_{A_t(i)} \left\langle \boldsymbol{a}_{A_t(i)}, \boldsymbol{\theta}_t(i) \right\rangle , \quad M_{\pi_1(s)}(i) = \sum_{k \in [K]} p_{\pi_1(s)}(i,k) \boldsymbol{a}_k \boldsymbol{a}_k^\top .$$

Using Lemma 17, for all $k \in [K]$ we know that $|\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i) \rangle| \leqslant \frac{|\mathcal{S}|}{\beta}$, where $\beta = 3Bd\eta$. Hence we have $\eta \bar{z}_{\pi_1(s)+1}(k) \in [-1, 1]$. Since $\bar{\boldsymbol{p}}_{\pi_1(s)}$ follows the update rule shown in Eq. (82), using the standard analysis of FTRL (e.g. Theorem 5.2 in Hazan (2016)), we can obtain that

$$\begin{aligned} \spadesuit &= \mathbb{E}\left[\sum_{s=1}^{m_1} \left\langle \bar{\boldsymbol{p}}_{\pi_1(s)} - \boldsymbol{e}_k, \bar{\boldsymbol{z}}_{\pi_1(s)+1} \right\rangle \right] \\ &\leqslant \frac{\log K}{\eta} + 2\eta \sum_{s=1}^{m_1} \mathbb{E}\left[\sum_{k=1}^{K} \bar{p}_{\pi_1(s)}(k) \left(\frac{1}{N} \sum_{t \in \mathcal{T}_{\pi_1(s)}} \sum_{i=1}^{N} \widehat{\ell}_t(i,k) \right)^2 \right] . \end{aligned} \tag{86}$$

Define $\mathcal{F}_{\pi_1(s)}$ to be the filtration generated by all random variables revealed up to the beginning of $\pi_1(s)$. Since $\bar{p}_{\pi_1(s)}(k)$ is $\mathcal{F}_{\pi_1(s)}$-measurable, by the tower property,

$$\begin{aligned} \mathbb{E}\left[\bar{p}_{\pi_1(s)}(k) \bar{z}^2_{\pi_1(s)+1}(k) \right] &= \mathbb{E}\left[\mathbb{E}\left[\bar{p}_{\pi_1(s)}(k) \bar{z}^2_{\pi_1(s)+1}(k) \mid \mathcal{F}_{\pi_1(s)} \right] \right] \\ &= \mathbb{E}\left[\bar{p}_{\pi_1(s)}(k) \mathbb{E}\left[\bar{z}^2_{\pi_1(s)+1}(k) \mid \mathcal{F}_{\pi_1(s)} \right] \right] . \end{aligned} \tag{87}$$

Conditioned on $\mathcal{F}_{\pi_1(s)}$, for all $t \in \mathcal{T}_{\pi_1(s)}$ we have

$$\mathbb{E}\left[\widehat{\boldsymbol{\theta}}_t(i) \mid \mathcal{F}_{\pi_1(s)} \right] = M^{-1}_{\pi_1(s)}(i) \mathbb{E}\left[\boldsymbol{a}_{A_t(i)} \boldsymbol{a}^\top_{A_t(i)} \mid \mathcal{F}_{\pi_1(s)} \right] \boldsymbol{\theta}_t(i) = \boldsymbol{\theta}_t(i) , \tag{88}$$

and therefore for all $k \in [K]$,

$$\mathbb{E}\left[\left\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i)\right\rangle \mid \mathcal{F}_{\pi_1(s)}\right] = \langle \boldsymbol{a}_k, \boldsymbol{\theta}_t(i)\rangle = \ell_t(i,k)\ .$$

Moreover, since $|\langle \boldsymbol{a}_k, \boldsymbol{\theta}_t(i)\rangle| \leqslant 1$, we have

$$\begin{aligned}
\mathbb{E}\left[\left\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i)\right\rangle^2 \mid \mathcal{F}_{\pi_1(s)}\right] &= \mathbb{E}\left[\left\langle \boldsymbol{a}_k, M_{\pi_1(s)}^{-1}(i)\boldsymbol{a}_{A_t(i)}\right\rangle^2 \left\langle \boldsymbol{a}_{A_t(i)}, \boldsymbol{\theta}_t(i)\right\rangle^2 \mid \mathcal{F}_{\pi_1(s)}\right] \\
&\leqslant \mathbb{E}\left[\left\langle \boldsymbol{a}_k, M_{\pi_1(s)}^{-1}(i)\boldsymbol{a}_{A_t(i)}\right\rangle^2 \mid \mathcal{F}_{\pi_1(s)}\right] \\
&= \boldsymbol{a}_k^\top M_{\pi_1(s)}^{-1}(i)\mathbb{E}\left[\boldsymbol{a}_{A_t(i)}\boldsymbol{a}_{A_t(i)}^\top \mid \mathcal{F}_{\pi_1(s)}\right] M_{\pi_1(s)}^{-1}(i)\boldsymbol{a}_k \\
&= \boldsymbol{a}_k^\top M_{\pi_1(s)}^{-1}(i)\boldsymbol{a}_k\ . \qquad (89)
\end{aligned}$$

Using the conditional independence of $A_t(i)$ across $t \in \mathcal{T}_{\pi_t(s)}$ and $i \in V$, we are able to bound $\mathbb{E}\left[\bar{z}_{\pi_1(s)+1}(k)^2 \mid \mathcal{F}_{\pi_1(s)}\right]$ as follows:

$$\begin{aligned}
&\mathbb{E}\left[\bar{z}_{\pi_1(s)+1}(k)^2 \mid \mathcal{F}_{\pi_1(s)}\right] \\
&= \frac{1}{N^2}\mathbb{E}\left[\left(\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^N \left\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i)\right\rangle\right)^2 \middle| \mathcal{F}_{\pi_1(s)}\right] \\
&= \frac{1}{N^2}\left(\mathbb{E}\left[\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^N \left\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i)\right\rangle \middle| \mathcal{F}_{\pi_1(s)}\right]\right)^2 \\
&\quad + \frac{1}{N^2}\left(\mathbb{E}\left[\left(\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^N \left\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i) - \boldsymbol{\theta}_t(i)\right\rangle\right)^2 \middle| \mathcal{F}_{\pi_1(s)}\right]\right) \\
&= \frac{1}{N^2}\left(\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^N \ell_t(i,k)\right)^2 + \frac{1}{N^2}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^N\left(\mathbb{E}\left[\left\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i)\right\rangle^2 \middle| \mathcal{F}_{\pi_1(s)}\right]\right) \\
&\leqslant B^2 + \frac{1}{N^2}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^N \boldsymbol{a}_k^\top M_{\pi_1(s)}^{-1}(i)\boldsymbol{a}_k, \qquad (90)
\end{aligned}$$

where the last inequality holds because $|\mathcal{T}_{\pi_1(s)}| = B$ and $|\ell_t(i,k)| \leqslant 1$. Combining Eq. (87)–Eq. (89) yields

$$\begin{aligned}
\mathbb{E}\left[\sum_{k=1}^K \bar{p}_{\pi_1(s)}(k)\bar{z}_{\pi_1(s)+1}(k)\right] &\leqslant \mathbb{E}\left[\sum_{k=1}^K \bar{p}_{\pi_1(s)}(k)\left(B^2 + \frac{1}{N^2}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^N \boldsymbol{a}_k^\top M_{\pi_1(s)}^{-1}(i)\boldsymbol{a}_k\right)\right] \\
&= B^2 + \frac{1}{N^2}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^N \mathbb{E}\left[\sum_{k=1}^K \bar{p}_{\pi_1(s)}(k)\boldsymbol{a}_k^\top M_{\pi_1(s)}^{-1}(i)\boldsymbol{a}_k\right] \\
&= B^2 + \frac{1}{N^2}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^N \mathbb{E}\left[\operatorname{tr}\left(M_{\pi_1(s)}^{-1}(i)\overline{M}_{\pi_1(s)}\right)\right], \qquad (91)
\end{aligned}$$

where $\overline{M}_{\pi_1(s)} = \sum_{k=1}^K \bar{p}_{\pi_1(s)}(k)\boldsymbol{a}_k\boldsymbol{a}_k^\top$. Then, according to Lemma 16, we know that

$$\frac{\bar{p}_{\pi_1(s)}(k)}{p_{\pi_1(s)}(i,k)} \leqslant \frac{\bar{p}_{\pi_1(s)}(k)}{(1-\alpha-\beta)p'_{\pi_1(s)}(i,k) + \frac{\alpha}{K} + \frac{\beta}{|\mathcal{S}|}\mathbb{1}\{\boldsymbol{a}_k \in \mathcal{S}\}} \leqslant 6 \qquad \text{for all } i \in [N],\ k \in [K]\,.$$

Therefore,

$$\overline{M}_{\pi_1(s)} = \sum_{k=1}^K \bar{p}_{\pi_1(s)}(k)\boldsymbol{a}_k\boldsymbol{a}_k^\top \preceq 6\sum_{k=1}^K p_{\pi_1(s)}\boldsymbol{a}_k\boldsymbol{a}_k^\top = 6M_{\pi_1(s)}(i),$$

and hence

$$\operatorname{tr}\left(M_{\pi_1(s)}(i)^{-1}\overline{M}_{\pi_1(s)}\right) \leqslant \operatorname{tr}\left(M_{\pi_1(s)}(i)^{-1}\cdot 6M_{\pi_1(s)}(i)\right) = 6\operatorname{tr}(I_d) = 6d.$$

Plugging this into Eq. (91) gives

$$\mathbb{E}\left[\sum_{k=1}^K \bar{p}_{\pi_1(s)}(k)\bar{z}_{\pi_1(s)+1}(k)^2\right] \leqslant B^2 + \frac{1}{N^2}\sum_{t\in\mathcal{T}_{\pi_1(s)}}\sum_{i=1}^N 6d \leqslant B^2 + \frac{6Bd}{N}. \tag{92}$$

Substituting Eq. (84), and Eq. (85), Eq. (86) and Eq. (92) into the decomposition Eq. (83), we obtain

$$\operatorname{Reg}_{\mathcal{P}_1(i)} \leqslant \frac{\log K}{\eta} + 2\eta\left(m_1B^2 + \frac{6KBm_1}{N}\right) + 3\,.$$

The analysis for $\operatorname{Reg}_{\mathcal{P}_2}(i)$ is identical. Summing the bounds over the two parity subsequences and using $m_1 + m_2 = T/B$ yields

$$\begin{aligned}
\mathbb{E}\left[\sum_{\tau=1}^M \left\langle \boldsymbol{p}'_\tau(i) - \boldsymbol{e}_k, \boldsymbol{z}^B_{\tau+1}(i)\right\rangle\right] &\leqslant \frac{2\log K}{\eta} + 2\eta\left(TB + \frac{6dT}{N}\right) + 6 \\
&\leqslant \frac{2\log K}{\eta} + 12\eta\left(TB + \frac{dT}{N}\right) + 6 \\
&\leqslant \mathcal{O}\left(\sqrt{\log K\left(B + \frac{1}{N}\right)dT} + dB\log K\right)
\end{aligned}$$

where the last inequality follows by choosing

$$\eta = \min\left\{\frac{1}{6Bd}, \sqrt{\frac{\log K}{dTB + \frac{dT}{N}}}\right\}.$$

■

D.5. Omitted Proof Details for Lemma 16

Lemma 16 *Let $\mathcal{A}$ be Algorithm 2 and $\mathcal{B}$ be an instance of Algorithm 5 with a regularizer that is $\frac{1}{\eta}$-strongly convex in ℓ_2-norm. Suppose that each agent uses Algorithm 4 with κ and B defined in Eq. (4). Define $\bar{q}_s^{(1)}$ and $q_s^{(1)}(i)$ for all $i \in [N]$*

$$\boldsymbol{q}_s^{(1)}(i) = \operatorname*{arg\,min}_{\boldsymbol{q}\in\Delta(K)}\left\{\sum_{s'=1}^{s-1}\left\langle \tilde{z}^B_{\pi_1(s')+1}(i,\cdot), \boldsymbol{q}\right\rangle + \psi(\boldsymbol{q})\right\},$$

$$\bar{\boldsymbol{q}}_s^{(1)} = \arg\min_{q \in \Delta(K)} \left\{ \sum_{s'=1}^{s-1} \left\langle \bar{\boldsymbol{z}}_{\pi_1(s')+1}, q \right\rangle + \psi(\boldsymbol{q}) \right\},$$

where $\bar{\boldsymbol{z}}_\tau$ is defined in Eq. (3) and π_1 is defined in Eq. (19). Then we have

$$\|\bar{\boldsymbol{q}}_s^{(1)} - \boldsymbol{q}_s^{(1)}(i)\|_2 \leqslant \frac{\eta}{K^{1.5}T^2} \qquad \textit{for all } i \in [N]. \tag{93}$$

and

$$\frac{\bar{q}_s^{(1)}(k)}{(1-\alpha-\beta)q_s^{(1)}(i,k) + \frac{\alpha}{K} + \frac{\beta}{|\mathcal{S}|}\mathbb{1}\{\boldsymbol{a}_k \in \mathcal{S}\}} \leqslant 6 \quad \textit{for all } i \in [N] \textit{ and } k \in [K],$$

where $\alpha = \frac{1}{T}$ and $\beta = 3Bd\eta \leqslant \frac{1}{2}$.

Proof Since $\psi(\boldsymbol{q})$ is 1-strongly convex w.r.t. $\|\cdot\|_2$, using Lemma 10, we know that

$$\begin{aligned}
\|\bar{\boldsymbol{q}}_s^{(1)} - \boldsymbol{q}_s^{(1)}(i)\|_2 &\leqslant \eta \left\| \sum_{s'=1}^{s-1} \tilde{\boldsymbol{z}}_{\pi_1(s')+1} - \sum_{s'=1}^{s-1} \tilde{\boldsymbol{z}}^B_{\pi_1(s')+1}(i) \right\|_2 \\
&\leqslant \eta \sum_{s'=1}^{s-1} \|\bar{\boldsymbol{z}}_{\pi_1(s')+1} - \boldsymbol{z}^B_{\pi_1(s')+1}(i)\|_2 \\
&\leqslant \frac{\eta}{K^{1.5}T},
\end{aligned}$$

where the last inequality is due to Lemma 13 and $s \leqslant T$. As $\eta \leqslant 1$ and $\|a-b\|_\infty \leqslant \|a-b\|_2$ for any $a, b \in \mathbb{R}^K$, we have

$$\bar{q}_s^{(1)}(k) - q_s^{(1)} \leqslant \frac{1}{KT} \qquad \text{for all } k \in [K].$$

Note that $\alpha = \frac{1}{T}$ and $\beta = 3Bd\eta \leqslant \frac{1}{2}$. Direct calculation shows that for $\boldsymbol{a}_k \in \Omega \backslash \mathcal{S}$ and for $T \geqslant 3$,

$$\begin{aligned}
\frac{\bar{q}_s^{(1)}(k)}{(1-\alpha-\beta)q_s^{(1)}(i,k) + \alpha/K} &\leqslant \frac{q_s^{(1)}(i,k) + \frac{1}{KT}}{(1-\alpha-\beta)q_s^{(1)}(i,k) + \alpha/K} \\
&\leqslant \frac{1}{1-\alpha-\beta} \leqslant 6,
\end{aligned}$$

and for $\boldsymbol{a}_k \in \mathcal{S}$ and for $T \geqslant 3$

$$\frac{\bar{q}_s^{(1)}(k)}{(1-\alpha-\beta)q_s^{(1)}(i,k) + \alpha/K + \beta/|\mathcal{S}|} \leqslant \frac{q_s^{(1)}(i,k) + \frac{1}{KT}}{(1-\alpha-\beta)q_s^{(1)}(i,k) + \alpha/K} \leqslant 6\,.$$

■

Lemma 17 *Algorithm 4 guarantees that $\left|\left\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i) \right\rangle\right| \leqslant \frac{|\mathcal{S}|}{\beta}$ for all $k \in [K]$ and $t \in [T]$.*

Proof By definition of $\widehat{\boldsymbol{\theta}}_t(i,k)$ and Cauchy-Schwarz inequality, for all $t \in \mathcal{T}_\tau$, we have

$$\left|\left\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i) \right\rangle\right| \leqslant \|\boldsymbol{a}_k\|_{M_\tau(i)^{-1}} \cdot \|\boldsymbol{a}_{A_t(i)}\|_{M_\tau(i)^{-1}} \cdot \left|\boldsymbol{a}_{A_t(i)}^\top \boldsymbol{\theta}_t(i)\right|. \tag{94}$$

By the definition of $p_\tau(i)$, we know that

$$M_\tau(i) \succeq \frac{\beta}{|\mathcal{S}|} \sum_{k \in [|\mathcal{S}|]} \boldsymbol{b}_k \boldsymbol{b}_k^\top \triangleq \frac{\beta}{|\mathcal{S}|} \Sigma_{\mathcal{S}},$$

so $M_\tau(i)^{-1} \preceq \frac{|\mathcal{S}|}{\beta} \Sigma_{\mathcal{S}}^{-1}$, and therefore for any $j \in [K]$,

$$\|\boldsymbol{a}_j\|_{M_\tau(i)^{-1}}^2 = \boldsymbol{a}_j^\top M_\tau(i)^{-1} \boldsymbol{a}_j \leqslant \frac{|\mathcal{S}|}{\beta} \boldsymbol{a}_j^\top \Sigma_{\mathcal{S}}^{-1} \boldsymbol{a}_j.$$

It remains to show that $\boldsymbol{a}_j^\top \Sigma_{\mathcal{S}}^{-1} \boldsymbol{a}_j \leqslant 1$ for all $j \in [K]$. Let $S := [\boldsymbol{b}_1, \cdots, \boldsymbol{b}_{|\mathcal{S}|}] \in \mathbb{R}^{d \times |\mathcal{S}|}$, so that $\Sigma_{\mathcal{S}} = SS^\top$. Since $\mathcal{S}$ is a volumetric spanner of Ω, for every $j \in [K]$ there exists $\boldsymbol{\lambda}^{(j)} \in \mathbb{R}^{|\mathcal{S}|}$ with $\|\boldsymbol{\lambda}^{(j)}\|_2 \leqslant 1$ such that $\boldsymbol{a}_j = S\boldsymbol{\lambda}^{(j)}$. Hence

$$\boldsymbol{a}_j^\top \Sigma_{\mathcal{S}}^{-1} \boldsymbol{a}_j = \left(\boldsymbol{\lambda}^{(j)}\right)^\top S^\top (SS^\top)^{-1} S \boldsymbol{\lambda}^{(j)} = \left(\boldsymbol{\lambda}^{(j)}\right)^\top P \boldsymbol{\lambda}^{(j)},$$

where $P \triangleq S^\top (SS^\top)^{-1} S$. Because $SS^\top$ is invertible, P is the orthogonal projector, and therefore $0 \preceq P \preceq I$. It follows that

$$\boldsymbol{a}_j^\top \Sigma_{\mathcal{S}}^{-1} \boldsymbol{a}_j = \left(\boldsymbol{\lambda}^{(j)}\right)^\top P \boldsymbol{\lambda}^{(j)} \leqslant \left(\boldsymbol{\lambda}^{(j)}\right)^\top I \boldsymbol{\lambda}^{(j)} = \|\boldsymbol{\lambda}^{(j)}\|_2^2 \leqslant 1\,.$$

Therefore, we have for all $j \in [K]$,

$$\|\boldsymbol{a}_j\|_{M_\tau(i)^{-1}} \leqslant \sqrt{\frac{|\mathcal{S}|}{\beta}}\,.$$

Finally, since $|\boldsymbol{a}_j^\top \boldsymbol{\theta}_t(i)| \leqslant 1$ for all $j \in [K]$, $i \in [N]$, $t \in [T]$. plugging into Eq. (94) gives

$$\left|\left\langle \boldsymbol{a}_k, \widehat{\boldsymbol{\theta}}_t(i) \right\rangle\right| \leqslant \sqrt{\frac{|\mathcal{S}|}{\beta}} \cdot \sqrt{\frac{|\mathcal{S}|}{\beta}} \cdot 1 = \frac{|\mathcal{S}|}{\beta}.$$

This proves the lemma. ■

Appendix E. Omitted Details in Section 6

In this section, we show the omitted proofs in Section 6.

Theorem 8 *For any distributed K-armed bandit algorithm with $K \geqslant 2$, and for any N, T large enough, there exists a communication graph G with N nodes, a gossip matrix W, and a sequence $\ell_1, \ldots, \ell_T$ of loss matrices $\ell_t(\cdot,\cdot) \in [0,1]^{N \times K}$ such that* $\mathrm{Reg}_T = \Omega\left(\sqrt{\rho^{-1/2} T \log K + \frac{KT}{N}}\right)$.

Proof The $\Omega\big(\sqrt{KT/N}\big)$ term in the lower bound is obtained by applying (Seldin et al., 2014, Theorem 2), which holds for the regret of any single agent in the simpler K-armed bandit setting when G is a clique and the local losses are identical for all agents. The other term in the lower bound is obtained by applying the

$$\Omega\left(\left(\frac{d_{\max}}{\lambda_{N-1}(L)}\right)^{1/4}\sqrt{(\log K)T}\right) \tag{95}$$

lower bound proven in (Yi and Vojnovic, 2023, Theorem 3.1) for a certain graph G with Laplacian matrix L, maximum degree $d_{\max}$, and where $\lambda_{N-1}(L)$ is the smallest non-zero eigenvalue of L. In order to relate $\lambda_{N-1}(L)$ to $\rho(W) = 1 - \sigma_2(W)$, we choose $W = I - \frac{\alpha}{2d_{\max}}L$ where $0 < \alpha < 1$. It is easy to check that W is a gossip matrix with second largest singular value $\sigma_2(W) = 1 - \frac{\alpha}{2d_{\max}}\lambda_{N-1}(L)$. Since $0 \leqslant \lambda_i(L) \leqslant 2d_{\max}$ for all $i \in [N]$ (Anderson Jr and Morley, 1985), we have $\sigma_2(W) > 0$. Hence we can write $\frac{1}{2(1-\sigma_2(W))} = \frac{d_{\max}}{\alpha\lambda_{N-1}(L)}$, which, substituted in Eq. (95), delivers the desired other term in the lower bound. ■

Theorem 9 *For any distributed linear bandit algorithm, and for any N, T large enough, there exists an action set $\Omega \subset \mathbb{R}^d$ with $|\Omega| = K \geqslant 2$, a communication graph G with N nodes, a gossip matrix W supported on G, and local loss coefficients $\boldsymbol{\theta}_t(i) \in \mathbb{R}^d$ for $t \in T$ and $i \in V$ such that* $\mathrm{Reg}_T = \Omega\left(\sqrt{(\log K)\left(\rho^{-1/2} + \frac{d}{N}\right)T}\right)$.

Proof To prove the first term we exploit the fact that K-armed bandits are a special case of linear bandits, and invoke Theorem 8. For the second term, we use Ito et al. (2020, Theorem 4) where $\Omega = \{-1, 1\}^d$, so that $\log|\Omega| = d$. ■