Collaborative Pure Exploration in Kernel Bandit

YIHAN DU, IIIS, Tsinghua University, China

WEI CHEN, Microsoft Research, China

YUKO KUROKI, The University of Tokyo / RIKEN, Japan

LONGBO HUANG, IIIS, Tsinghua University, China

In this paper, we formulate a Collaborative Pure Exploration in Kernel Bandit problem (CoPE-KB), which provides a novel model for multi-agent multi-task decision making under limited communication and general reward functions, and is applicable to many online learning tasks, e.g., recommendation systems and network scheduling. We consider two settings of CoPE-KB, i.e., Fixed-Confidence (FC) and Fixed-Budget (FB), and design two optimal algorithms CoopKernelFC (for FC) and CoopKernelFB (for FB). Our algorithms are equipped with innovative and efficient kernelized estimators to simultaneously achieve computation and communication efficiency. Matching upper and lower bounds under both the statistical and communication metrics are established to demonstrate the optimality of our algorithms. The theoretical bounds successfully quantify the influences of task similarities on learning acceleration and only depend on the effective dimension of the kernelized feature space. Our analytical techniques, including data dimension decomposition, linear structured instance transformation and (communication) round-speedup induction, are novel and applicable to other bandit problems. Empirical evaluations are provided to validate our theoretical results and demonstrate the performance superiority of our algorithms.

Additional Key Words and Phrases: collaborative pure exploration, kernel bandit, communication round, learning speedup

ACM Reference Format:

Yihan Du, Wei Chen, Yuko Kuroki, and Longbo Huang. 2021. Collaborative Pure Exploration in Kernel Bandit. In *Woodstock '21: ACM Symposium on Neural Gaze Detection, June 03–05, 2021, Woodstock, NY*. ACM, New York, NY, USA, 44 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Pure exploration [3, 9, 12, 18, 21, 24] is a fundamental online learning problem in multi-armed bandits, where an agent sequentially chooses options (often called arms) and observes random feedback, with the objective of identifying the best option (arm). This problem finds various applications such as recommendation systems [31], online advertising [37] and neural architecture search [17]. However, the traditional single-agent pure exploration problem [3, 9, 12, 18, 21, 24] cannot be directly applied to many real-world distributed online learning platforms, which often face a large volume of user requests and need to coordinate multiple *distributed* computing devices to process the requests, e.g., geographically distributed data centers [30] and Web servers [44]. These computing devices communicate with each other to exchange information in order to attain globally optimal performance.

To handle such distributed pure exploration problem, prior works [20, 22, 38] have developed the Collaborative Pure Exploration (CoPE) model, where there are multiple agents that communicate and cooperate in order to identify the best arm with learning speedup. Yet, existing results only investigate the classic multi-armed bandit (MAB) setting [3, 18, 21], and focus only on the *fully-collaborative* setting, i.e., the agents aim to solve a common task. However, in many real-world applications such as recommendation systems [31], it is often the case that different computing devices face *different but correlated* recommendation tasks. Moreover, there usually exists some *structured dependency* of user utilities on the recommended items. In such applications, it is important to develop a more general CoPE model that allows heterogeneous tasks and complex reward structures, and quantitatively investigate how task similarities impact learning acceleration.

Motivated by the above facts, we propose a novel *Collaborative Pure Exploration in Kernel Bandit (CoPE-KB)* problem, which generalizes traditional single-task CoPE problems [20, 22, 38] to the multi-task setting. It also generalizes the classic MAB model to allow general (linear or nonlinear) reward structures via the powerful kernel representation. Specifically, each agent is given a set of arms, and the expected reward of each arm is generated by a task-dependent reward function with a low norm in a high-dimensional (possibly infinite-dimensional) Reproducing Kernel Hilbert Space (RKHS) [33, 42], by which we can represent real-world nonlinear reward dependency as some linear function in a high-dimensional space, and can go beyond linear rewards as commonly done in the literature, e.g., [10, 13, 14, 35, 40]. Each agent sequentially chooses arms to sample and observes noisy outcomes. The agents can broadcast and receive messages to/from others in communication rounds, so that they can exploit the task similarity and collaborate to expedite learning processes. The task of each agent is to find the best arm that maximizes the expected reward among her arm set.

Our CoPE-KB formulation can handle different tasks in parallel and characterize the dependency of rewards on options, which provides a more general and flexible model for real-world applications. For example, in distributed recommendation systems [31], different computing devices can face different tasks, and it is inefficient to learn the reward of each option individually. Instead, CoPE-KB enables us to directly learn the relationship between option features and user utilities, and exploit the similarity of such relationship among different tasks to accelerate learning. There are also many other applications, such as clinical trials [41], where we conduct multiple clinical trials in parallel and utilize the common useful information to accelerate drug development, and neural architecture search [17], where we simultaneously run different tests of neural architectures under different environmental setups to expedite search processes.

We consider two important pure exploration settings under the CoPE-KB model, i.e., *Fixed-Confidence (FC)*, where agents aim to minimize the number of used samples under a given confidence, and *Fixed-Budget (FB)*, where the goal is to minimize the error probability under a given sample budget. Note that due to the high dimension (possibly infinite) of the RKHS, it is highly non-trivial to simplify the burdensome computation and communication in the RKHS, and to derive theoretical bounds only dependent on the effective dimension of the kernelized feature space. To tackle the above challenges, we adopt efficient kernelized estimators and design novel algorithms CoopKernelFC and CoopKernelFB for the FC and FB settings, respectively, which only cost $\mathrm{Poly}(nV)$ computation and communication complexity instead of $\mathrm{Poly}(\dim(\mathcal{H}_K))$ as in [10, 43], where n is the number of arms, V is the number of agents, and $\mathcal{H}_K$ is the high-dimensional RKHS. We also establish matching upper and lower bounds in terms of sampling and communication complexity to demonstrate the optimality of our algorithms (within logarithmic factors).

Our work distinguishes itself from prior CoPE works, e.g., [20, 22, 38], in the following aspects: (i) Prior works [20, 22, 38] only consider the classic MAB setting, while we adopt a high-dimensional RKHS to allow more general real-world reward dependency on option features. (ii) Unlike [20, 22, 38] which restrict tasks (given arm sets and rewards) among agents to be the same, we allow different tasks for different agents, and explicitly quantify how task similarities impact learning acceleration. (iii) In lower bound analysis, prior works [20, 38] mainly focus on a 2-armed case, whereas we derive a novel lower bound analysis for general multi-armed cases with high-dimensional linear reward structures. Moreover, when reducing CoPE-KB to prior CoPE with classic MAB setting (all agents are solving the same classic MAB task) [20, 38], our lower and upper bounds also match the existing state-of-the-art results in [38].

The contributions of this paper are summarized as follows:

- We formulate a novel Collaborative Pure Exploration in Kernel Bandit (CoPE-KB) problem, which models distributed multi-task decision making problems with general reward functions, and finds applications in many real-world online learning tasks, and study two settings of CoPE-KB, i.e., CoPE-KB with fixed-confidence (FC) and CoPE-KB with fixed-budget (FB).
- For CoPE-KB with fixed-confidence (FC), we propose an algorithm CoopKernelFC, which adopts an efficient kernelized estimator to significantly reduce computation and communication complexity from existing $\text{Poly}(\dim(\mathcal{H}_K))$ to only $\text{Poly}(nV)$. We derive matching upper and lower bounds of sample complexity $\tilde{O}(\frac{\rho^*}{V}\log\delta^{-1})$ and communication rounds $O(\log\Delta_{\min}^{-1})$. Here ρ^* is the problem hardness (defined in Section 4.2), and $\Delta_{\min}^{-1}$ is the minimum reward gap.
- For CoPE-KB with fixed-budget (FB), we design an efficient algorithm CoopKernelFB with error probability $\tilde{O}\left(\exp\left(-\frac{TV}{\rho^*}\right)n^2V\right)$ and communication rounds $O(\log(\omega(\tilde{\mathcal{X}})))$. A matching lower bound of communication rounds is also established to validate the communication optimality of CoopKernelFB (within double-logarithmic factors). Here T is the sample budget, $\tilde{\mathcal{X}}$ is the set of arms, and $\omega(\tilde{\mathcal{X}})$ is the principle dimension of data projections in $\tilde{\mathcal{X}}$ to RKHS (defined in Section 5.1.1).
- Our results explicitly quantify the impacts of task similarities on learning acceleration. Our novel analytical techniques, including data dimension decomposition, linear structured instance transformation and round-speedup induction, can be of independent interests and are applicable to other bandit problems.

Due to space limit, we defer all the proofs to Appendix.

2 RELATED WORK

This work falls in the literature of multi-armed bandits [4, 8, 27, 28, 39]. Here we mainly review three most related lines of research, i.e., collaborative pure exploration and kernel bandit.

Collaborative Pure Exploration (CoPE). The collaborative pure exploration literature is initiated by [20], which considers the classic MAB and fully-collaborative settings, and designs fixed-confidence algorithms based on majority vote with upper bound analysis. Tao et al. [38] further develop a fixed-budget algorithm by calling conventional single-agent fixed-confidence algorithms, and completes the analysis of round-speedup lower bounds. Karpov et al. [22] extend the formulation of [20, 38] to the best m arm identification problem, and designs fixed-confidence and fixed-budget algorithms with tight round-speedup upper and lower bounds, which give a strong separation between best arm identification and the extended best m arm identification. Our CoPE-KB model encompasses the classic MAB and fully-collaborative settings in the above works [20, 22, 38], but faces unique challenges on computation and communication efficiency due to the high-dimensional reward structures.

Collaborative Regret Minimization. There are other works studying collaborative (distributed) bandit with the regret minimization objective. Bistritz and Leshem [5], Liu and Zhao [29], Rosenski et al. [32] study the multi-player bandit with collisions motivated by cognitive radio networks, where multiple players simultaneously choose arms from the same set and receive no reward if more than one player choose the same arm (i.e., a collision happens). Bubeck and Budzinski [6], Bubeck et al. [7] investigate a variant multi-player bandit problem where players cannot communicate but have access to shared randomness, and they propose algorithms that achieve nearly optimal regrets without collisions. Chakraborty et al. [11] introduce another distributed bandit problem, where each agent decides either to pull an arm or to broadcast a message in order to maximize the total reward. Korda et al. [25], Szorenyi et al. [36] adapt bandit algorithms to peer-to-peer networks, where the peers pick arms from the same set and can only communicate with a few random others along network links. The above works consider different learning objectives and communication protocols from ours, and do not involve the challenges of simultaneously handling multiple different tasks and analyzing the relationship between communication rounds and learning speedup.

Kernel Bandit. There are a number of works for kernel bandit with the regret minimization objective. Srinivas et al. [35] study the Gaussian process bandit problem with RKHS, which is the Bayesian version of kernel bandits, and designs an Upper Confidence Bound (UCB) style algorithm. Chowdhury and Gopalan [13] further improve the regret results of [35] by constructing tighter kernelized confidence intervals. Valko et al. [40] consider kernel bandit from the frequentist perspective and provides an alternative regret analysis based on effective dimension. Deshmukh et al. [14], Krause and Ong [26] study the multi-task kernel bandits, where the kernel function of RKHS is constituted by two compositions from task similarities and arm features. Dubey et al. [16] investigate the multi-agent kernel bandit with a local communication protocol, with the learning objective being to reduce the average regret suffered by per agent. For kernel bandit with the pure exploration objective, there are only two works [10, 43] to our best knowledge. Camilleri et al. [10] design a single-agent algorithm which uses a robust inverse propensity score estimator to reduce the sample complexity incurred by rounding procedures. Zhu et al. [43] propose a variant of [10] which applies neural networks to approximate nonlinear reward functions. All of these works consider either regret minimization or single-agent setting, which largely differs from our problem, and they do not investigate the distributed decision making and (communication) round-speedup trade-off. Thus, their algorithms and analysis cannot be applied to solve our CoPE-KB problem.

3 COLLABORATIVE PURE EXPLORATION IN KERNEL BANDIT (COPE-KB)

In this section, we present the formal formulation of the *Collaborative Pure Exploration in Kernel Bandit (CoPE-KB)*, and discuss the two important settings under CoPE-KB that will be investigated.

Agents and rewards. There are V agents $[V] = \{1, \dots, V\}$, who collaborate to solve different but possibly related instances (tasks) of the pure exploration in kernel bandit (PE-KB) problem. For each agent $v \in [V]$, she is given a set of n arms $\mathcal{X}_v = \{x_{v,1}, \dots, x_{v,n}\} \subseteq \mathbb{R}^{d_{\mathcal{X}}}$, where $d_{\mathcal{X}}$ is the dimension of arm feature vectors. The expected reward of each arm $x \in \mathcal{X}_v$ is $f_v(x)$, where $f_v : \mathcal{X}_v \mapsto \mathbb{R}$ is an unknown reward function. Let $\mathcal{X} = \cup_{v\in[V]}\mathcal{X}_v$. Following the literature in kernel bandits [14, 26, 35, 40], we assume that for any $v \in [V]$, f_v has a bounded norm in a Reproducing Kernel Hilbert Space (RKHS) specified by kernel $K_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ (see below for more details). At each timestep t, each agent v pulls an arm $x_{v,t} \in \mathcal{X}_v$ and observes a random reward $y_{v,t} = f(x_{v,t}) + \eta_{v,t}$, where $\eta_{v,t}$ is an independent and zero-mean 1-sub-Gaussian noise (without loss of generality).[1] We assume that the best arms $x_{v,*} = \mathrm{argmax}_{x\in\mathcal{X}_v} f_v(x)$ are unique for all $v \in [V]$, which is a common assumption in the pure exploration literature, e.g., [3, 12, 18, 24].

Multi-Task Kernel Composition. We assume that the functions f_v are parametric functionals of a global function $F : \mathcal{X} \times \mathcal{Z} \mapsto \mathbb{R}$, which satisfies that, for each agent $v \in [V]$, there exists a task feature vector $z_v \in \mathcal{Z}$ such that

$$f_v(x) = F(x, z_v), \ \forall x \in \mathcal{X}_v. \tag{1}$$

Here $\mathcal{X}$ and $\mathcal{Z}$ denote the arm feature space and task feature space, respectively. Eq. (1) allows tasks to be different for agents, whereas prior CoPE works [20, 22, 38] restrict the tasks ($\mathcal{X}_v$ and f_v) to be the same for all agents $v \in [V]$.

[1] A random variable η is called 1-sub-Gaussian if it satisfies that $\mathbb{E}[\exp(\lambda\eta - \lambda\mathbb{E}[\eta])] \le \exp(\lambda^2/2)$ for any $\lambda \in \mathbb{R}$.

Denote $\tilde{\mathcal{X}} = \mathcal{X} \times \mathcal{Z}$ and $\tilde{x} = (x, z_v)$. As a standard assumption in kernel bandits [14, 16, 26, 35], we assume that F has a bounded norm in a global RKHS $\mathcal{H}_K$ specified by kernel $K : \tilde{\mathcal{X}} \times \tilde{\mathcal{X}} \mapsto \mathbb{R}$, and there exists a feature mapping $\phi : \tilde{\mathcal{X}} \mapsto \mathcal{H}_K$ and an unknown parameter $\theta^* \in \mathcal{H}_K$ such that

$$F(\tilde{x}) = \phi(\tilde{x})^\top \theta^*, \ \forall \tilde{x} \in \tilde{\mathcal{X}}, \quad \text{and} \quad K(\tilde{x}, \tilde{x}') = \phi(\tilde{x})^\top \phi(\tilde{x}'), \ \forall \tilde{x}, \tilde{x}' \in \tilde{\mathcal{X}}.$$

Here $\|\theta^*\| := \sqrt{\theta^{*\top}\theta^*} \leq B$ for some known constant $B > 0$. $K : \tilde{\mathcal{X}} \times \tilde{\mathcal{X}} \mapsto \mathbb{R}$ is a product composite kernel, which satisfies that for any $z, z' \in \mathcal{Z}, x, x' \in \mathcal{X}$,

$$K((x, z), (x', z')) = K_{\mathcal{Z}}(z, z') \cdot K_{\mathcal{X}}(x, x'),$$

where $K_{\mathcal{X}}$ is the arm feature kernel that depicts the feature structure of arms, and $K_{\mathcal{Z}}$ is the task feature kernel that measures the similarity of functions f_v. For example, in the fully-collaborative setting, all agents solve a common task, and we have that $z_v = 1$ for all $v \in [V]$, $K_{\mathcal{Z}}(z, z') = 1$ for all $z, z' \in \mathcal{Z}$, and $K = K_{\mathcal{X}}$. On the contrary, if all tasks are different, then $\text{rank}(K_{\mathcal{Z}}) = V$. $K_{\mathcal{Z}}$ allows us to characterize the influences of task similarities ($1 \leq \text{rank}(K_{\mathcal{Z}}) \leq V$) on learning.

We give a simple 2-agent (2-task) illustrating example in Figure 1. Agent 1 is given Items 1,2 with the expected rewards μ_1, μ_2, respectively, denoted by $\mathcal{X}_1 = \{x_{1,1}, x_{1,2}\}$. Agent 2 is given Items 2,3 with the expected rewards μ_2, μ_3, respectively, denoted by $\mathcal{X}_2 = \{x_{2,1}, x_{2,2}\}$. Here $x_{1,2} = x_{2,1}$ is the same item. In this case, $\phi(\tilde{x}_{1,1}) = [1, 0, 0]^\top$, $\phi(\tilde{x}_{1,2}) = \phi(\tilde{x}_{2,1}) = [0, 1, 0]^\top$, $\phi(\tilde{x}_{2,2}) = [0, 0, 1]^\top$, and $\theta^* = [\mu_1, \mu_2, \mu_3]^\top$. The two agents can share the learned information on the second dimension of θ^* to accelerate learning processes.

Note that the RKHS $\mathcal{H}_K$ can have infinite dimensions, and any direct operation on $\mathcal{H}_K$, e.g., the calculation of $\phi(\tilde{x})$ and explicit expression of the estimate of θ^*, is impracticable. In this paper, all our algorithms only query the kernel function $K(\cdot, \cdot)$ instead of directly operating on $\mathcal{H}_K$, and $\phi(\tilde{x})$ and θ^* are only used in our theoretical analysis, which is different from existing works, e.g., [10, 43].

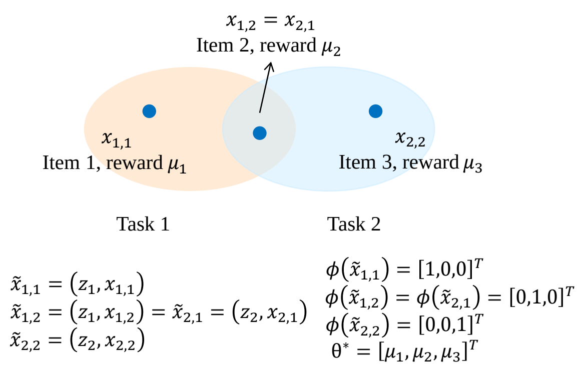


Fig. 1. Illustrating example.

Communication. Following the popular communication protocol in existing CoPE works [20, 22, 38], we allow these V agents to exchange information via communication rounds, in which each agent can broadcast and receive messages from others. While we do not restrict the exact length of a message, for practical implementation it should be bounded by $O(n)$ bits. Here n is the number of arms for each agent, and we consider the number of bits for representing a real number as a constant.

In the CoPE-KB problem, our goal is to design computation and communication efficient algorithms to coordinate different agents to simultaneously complete multiple tasks in collaboration and characterize how the take similarities impact the learning speedup.

Fixed-Confidence and Fixed-Budget. We consider two versions of the CoPE-KB problem, one with *fixed-confidence (FC)* and the other with *fixed-budget (FB)*. Specifically, in the FC setting, given a confidence parameter $\delta \in (0, 1)$, the agents aim to identify $x_{v,*}$ for all $v \in [V]$ with probability at least $1 - \delta$ and minimize the average number of samples used by each agent. In the FB setting, on the other hand, the agents are given an overall $T \cdot V$ sample budget (T average samples per agent), and aim to use at most $T \cdot V$ samples to identify $x_{v,*}$ for all $v \in [V]$ and minimize the error probability. In both FC and FB settings, agents are requested to minimize the number of communication rounds.

To evaluate the learning acceleration of our algorithms, following the CoPE literature, e.g., [20, 22, 38], we also define the speedup metric of our algorithms. For a CoPE-KB instance $\mathcal{I}$, let $T_{\mathcal{A}_M,\mathcal{I}}$ denote the average number of samples used by each agent in multi-agent algorithm $\mathcal{A}_M$ to identify $x_{v,*}$ for all $v \in [V]$, and let $T_{\mathcal{A}_S,\mathcal{I}}$ denote the average number of samples used by each task for a single-agent algorithm $\mathcal{A}_S$ to sequentially (without communication) identify $x_{v,*}$ for all $v \in [V]$. Then, the speedup of $\mathcal{A}_M$ on instance $\mathcal{I}$ is formally defined as

$$\beta_{\mathcal{A}_M,\mathcal{I}} = \inf_{\mathcal{A}_S} \frac{T_{\mathcal{A}_S,\mathcal{I}}}{T_{\mathcal{A}_M,\mathcal{I}}}. \tag{2}$$

It can be seen that $1 \leq \beta_{\mathcal{A}_M,\mathcal{I}} \leq V$, where $\beta_{\mathcal{A}_M,\mathcal{I}} = 1$ for the case where all tasks are different and $\beta_{\mathcal{A}_M,\mathcal{I}}$ can approach V for a fully-collaborative instance. By taking $T_{\mathcal{A}_M,\mathcal{I}}$ and $T_{\mathcal{A}_S,\mathcal{I}}$ as the smallest numbers of samples needed to meet the confidence constraint, the definition of $\beta_{\mathcal{A}_M,\mathcal{I}}$ can be similarly defined for error probability results.

In particular, when all agents $v \in [V]$ have the same arm set $\mathcal{X}_v = \mathcal{X} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ (i.e., standard bases in $\mathbb{R}^n$) and the same reward function $f_v(x) = f(x) = x^\top \theta^*$ for any $x \in \mathcal{X}$, all agents are solving a common classic MAB task, and then the task feature $\mathcal{Z} = \{1\}$ and $K_{\mathcal{Z}}(z, z') = 1$ for any $z, z' \in \mathcal{Z}$. In this case, our CoPE-KB problem reduces to prior CoPE with classic MAB setting [20, 38].

4 FIXED-CONFIDENCE COPE-KB

We start with the fixed-confidence (FC) setting and propose the CoopKernelFC algorithm. We explicitly quantify how task similarities impact learning acceleration, and provide sample complexity and round-speedup lower bounds to demonstrate the optimality of CoopKernelFC.

4.1 Algorithm CoopKernelFC

4.1.1 Algorithm. CoopKernelFC has three key components: (i) maintain alive arm sets for all agents, (ii) perform sampling individually according to the globally optimal sample allocation, and (iii) exchange the distilled observation information to estimate reward gaps and eliminate sub-optimal arms, via efficient kernelized computation and communication schemes.

The procedure of CoopKernelFC (Algorithm 1) for each agent v is as follows. Agent v maintains alive arm sets $\mathcal{B}_{v'}^{(t)}$ for all $v' \in [V]$ by successively eliminating sub-optimal arms in each phase. In phase t, she solves a global min-max optimization, which takes into account the objectives and available arm sets of all agents, to obtain the optimal sample allocation $\lambda_t^* \in \triangle_{\tilde{\mathcal{X}}}$ and optimal value ρ_t^* (Line 4). Here $\triangle_{\tilde{\mathcal{X}}}$ is the collection of all distributions on $\mathcal{X}$. ξ_t is a regularization parameter such that

$$\sqrt{\xi_t} \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{X}}_v, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{\left(\xi_t I + \sum_{x \in \mathcal{X}} \frac{1}{nV} \phi(x)\phi(x)^\top\right)^{-1}} \leq \frac{1}{(1+\varepsilon)B \cdot 2^{t+1}}, \tag{3}$$

which ensures the estimation bias for reward gap to be bounded by $2^{-(t+1)}$ and can be efficiently computed by kernelized transformation (specified in Section 4.1.2). Then, agent v uses ρ_t^* to compute the number of required samples $N^{(t)}$, which guarantees that the confidence radius of estimation for reward gaps is within 2^{-t} (Line 5). In algorithm CoopKernelFC, we use a rounding procedure $\mathrm{ROUND}_\varepsilon(\lambda, N)$ with approximation parameter ϵ from [2, 10], which rounds the sample

Algorithm 1: Distributed Algorithm CoopKernelFC: for Agent v ($v \in [V]$)

Input: $\delta, \tilde{\mathcal{X}}_1, \dots, \tilde{\mathcal{X}}_V, K(\cdot,\cdot) : \tilde{\mathcal{X}} \times \tilde{\mathcal{X}} \mapsto \mathbb{R}$, B, rounding procedure $\mathrm{ROUND}_\varepsilon(\cdot,\cdot)$ with approximation parameter ε.

1 **Initialization:** $\mathcal{B}^{(1)}_{v'} \leftarrow \mathcal{X}_{v'}$ for all $v' \in [V]$. $t \leftarrow 1$; `// initialize alive arm sets` $\mathcal{B}^{(1)}_{v'}$

2 **while** $\exists v' \in [V], |\mathcal{B}^{(t)}_{v'}| > 1$ **do**

3 $\quad \delta_t \leftarrow \frac{\delta}{2t^2}$;

4 $\quad$ Let λ^*_t and ρ^*_t be the optimal solution and optimal value of $\min_{\lambda \in \triangle_{\tilde{\mathcal{X}}}} \max_{\tilde{x}_i, \tilde{x}_j \in \mathcal{B}^{(t)}_{v'}, v' \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{(\xi_t I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda_{\tilde{x}} \phi(\tilde{x}) \phi(\tilde{x})^\top)^{-1}}$, where ξ_t is a regularization parameter that satisfies Eq. (3) ; `// compute the optimal sample allocation`

5 $\quad N^{(t)} \leftarrow \lceil 8(2^t)^2 (1+\varepsilon)^2 \rho^*_t \log\left(2n^2 V / \delta_t\right) \rceil$; `// compute the number of required samples`

6 $\quad (\tilde{s}_1, \dots, \tilde{s}_{N^{(t)}}) \leftarrow \mathrm{ROUND}_\varepsilon(\lambda^*_t, N^{(t)})$;

7 $\quad$ Let $\tilde{\boldsymbol{s}}^{(t)}_v$ be the sub-sequence of $(\tilde{s}_1, \dots, \tilde{s}_{N^{(t)}})$ which only contains the arms in $\tilde{\mathcal{X}}_v$; `// generate the sample sequence for agent` v

8 $\quad$ Pull arms $\tilde{\boldsymbol{s}}^{(t)}_v$ and observe random rewards $\boldsymbol{y}^{(t)}_v$;

9 $\quad$ Broadcast $\{(N^{(t)}_{v,i}, \bar{y}^{(t)}_{v,i})\}_{i \in [n]}$, where $N^{(t)}_{v,i}$ is the number of samples and $\bar{y}^{(t)}_{v,i}$ is the average observed reward on arm $\tilde{x}_{v,i}$;

10 $\quad$ Receive $\{(N^{(t)}_{v',i}, \bar{y}^{(t)}_{v',i})\}_{i \in [n]}$ from all other agents $v' \in [V] \setminus \{v\}$;

11 $\quad$ For notational simplicity, we combine the subscripts v', i in $\tilde{x}_{v',i}, N^{(t)}_{v',i}, \bar{y}^{(t)}_{v',i}$ by using $\tilde{x}_{(v'-1)n+i}, N^{(t)}_{(v'-1)n+i}, \bar{y}^{(t)}_{(v'-1)n+i}$, respectively. Then, $k_t(\tilde{x}) \leftarrow [\sqrt{N^{(t)}_1} K(\tilde{x}, \tilde{x}_1), \dots, \sqrt{N^{(t)}_{nV}} K(\tilde{x}, \tilde{x}_{nV})]^\top$ for any $\tilde{x} \in \tilde{\mathcal{X}}$. $K^{(t)} \leftarrow [\sqrt{N^{(t)}_i N^{(t)}_j} K(\tilde{x}_i, \tilde{x}_j)]_{i,j \in [nV]}$. $\bar{\boldsymbol{y}}^{(t)} \leftarrow [\sqrt{N^{(t)}_1} \bar{y}^{(t)}_1, \dots, \sqrt{N^{(t)}_{nV}} \bar{y}^{(t)}_{nV}]^\top$; `// organize overall observation information`

12 $\quad$ **for** all $v' \in [V]$ **do**

13 $\quad\quad \hat{\Delta}_t(\tilde{x}_i, \tilde{x}_j) \leftarrow (k_t(\tilde{x}_i) - k_t(\tilde{x}_j))^\top (K^{(t)} + N^{(t)} \xi_t I)^{-1} \bar{\boldsymbol{y}}^{(t)}, \ \ \forall \tilde{x}_i, \tilde{x}_j \in \mathcal{B}^{(t)}_{v'}$; `// estimate the reward gap between` $\tilde{x}_i$ `and` $\tilde{x}_j$

14 $\quad\quad \mathcal{B}^{(t+1)}_{v'} \leftarrow \mathcal{B}^{(t)}_{v'} \setminus \{\tilde{x} \in \mathcal{B}^{(t)}_{v'} | \exists \tilde{x}' \in \mathcal{B}^{(t)}_{v'} : \hat{\Delta}_t(\tilde{x}', \tilde{x}) \geq 2^{-t}\}$; `// discard sub-optimal arms`

15 $\quad t \leftarrow t + 1$;

16 E **return** $\mathcal{B}^{(t)}_1, \dots, \mathcal{B}^{(t)}_V$;

allocation $\lambda \in \triangle_{\tilde{\mathcal{X}}}$ into the integer numbers of samples $\kappa \in \mathbb{N}^{|\tilde{\mathcal{X}}|}$, such that $\sum_{\tilde{x} \in \tilde{\mathcal{X}}} \kappa_{\tilde{x}} = N$ and

$$\begin{aligned} &\max_{\tilde{x}_i, \tilde{x}_j \in \mathcal{B}^{(t)}_{v'}, v' \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{(N\xi_t I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \kappa_{\tilde{x}} \phi(\tilde{x}) \phi(\tilde{x})^\top)^{-1}} \\ \leq &(1+\epsilon) \max_{\tilde{x}_i, \tilde{x}_j \in \mathcal{B}^{(t)}_{v'}, v' \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{(N\xi_t I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} N\lambda_{\tilde{x}} \phi(\tilde{x}) \phi(\tilde{x})^\top)^{-1}}. \end{aligned}$$

By calling $\mathrm{ROUND}_\varepsilon(\lambda^*_t, N^{(t)})$, agent v generates an overall sample sequence $(\tilde{s}_1, \dots, \tilde{s}_{N^{(t)}})$ according to λ^*_t, and extracts a sub-sequence $\tilde{\boldsymbol{s}}^{(t)}_v$ that only contains the arms in $\tilde{\mathcal{X}}_v$ to sample (Lines 6-8). After sampling, she only communicates the number of samples $N^{(t)}_{v,i}$ and average observed reward $\bar{y}^{(t)}_{v,i}$ for each arm with other agents (Lines 9-10). With the overall observation information, she estimates the reward gap $\hat{\Delta}_t(\tilde{x}_i, \tilde{x}_j)$ between any arm pair $\tilde{x}_i, \tilde{x}_j \in \mathcal{B}^{(t)}_{v'}$ for all $v' \in [V]$ and discards sub-optimal arms (Lines 13-14).

4.1.2 Computation and Communication Efficiency. Here we explain the efficiency of CoopKernelFC. Note that in CoPE-KB, due to its high-dimensional reward structures, directly using the empirical mean to estimate rewards will cause loose sample complexity, and naively calculating and transmitting infinite-dimensional parameter θ^* will incur huge computation and communication costs. As a result, we cannot directly compute and communicate scalar empirical rewards as in prior CoPE with classic MAB works [20, 22, 38].

Computation Efficiency. CoopKernelFC uses three efficient kernelized operations, i.e., optimization solver (Line 4), condition for regularization parameter ξ_t (Eq. (3)) and estimator of reward gaps (Line 13). Unlike prior kernel bandit algorithms [10, 43] which explicitly compute $\phi(\tilde{x})$ and maintain the estimate of θ^* on the infinite-dimensional RKHS, CoopKernelFC only queries kernel function $K(\cdot,\cdot)$ and significantly reduces the computation (memory) costs from $\mathrm{Poly}(\dim(\mathcal{H}_K))$ to only $\mathrm{Poly}(nV)$.

Below we give the formal expressions of these operations and defer their detailed derivation to Appendix A.1.

Kernelized Estimator. We first introduce the kernelized estimator of reward gaps (Line 13). Following the standard estimation procedure in linear/kernel bandits [10, 19, 23, 43], we consider the following regularized least square estimator of underlying reward parameter θ^*

$$\hat{\theta}_t = \left(N^{(t)} \xi_t I + \sum_{j=1}^{N^{(t)}} \phi(\tilde{s}_j) \phi(\tilde{s}_j)^\top \right)^{-1} \sum_{j=1}^{N^{(t)}} \phi(\tilde{s}_j) y_j$$

Note that this form of $\hat{\theta}_t$ has $N^{(t)}$ terms in the summation, which are cumbersome to compute and communicate. Since the samples $(\tilde{s}_1, \dots, \tilde{s}_{N^{(t)}})$ are composed by arms $\tilde{x}_1, \dots, \tilde{x}_{nV}$, we merge repetitive computations for same arms in the summation and obtain (for notational simplicity, we combine the subscripts v', i in $\tilde{x}_{v',i}, N^{(t)}_{v',i}, \bar{y}^{(t)}_{v',i}$ by using $\tilde{x}_{(v'-1)n+i}, N^{(t)}_{(v'-1)n+i}, \bar{y}^{(t)}_{(v'-1)n+i}$, respectively)

$$\begin{aligned} \hat{\theta}_t &\overset{(a)}{=} \left(N^{(t)} \xi_t I + \sum_{i=1}^{nV} N_i^{(t)} \phi(\tilde{x}_i) \phi(\tilde{x}_i)^\top \right)^{-1} \sum_{i=1}^{nV} N_i^{(t)} \phi(\tilde{x}_i) \bar{y}_i^{(t)} \\ &\overset{(b)}{=} \Phi_t^\top \left(N^{(t)} \xi_t I + K^{(t)} \right)^{-1} \bar{y}^{(t)}. \end{aligned} \tag{4}$$

Here $N_i^{(t)}$ is the number of samples and $\bar{y}_i^{(t)}$ is the average observed reward on arm $\tilde{x}_i$ for any $i \in [nV]$. $\Phi_t = [\sqrt{N_1^{(t)}} \phi(\tilde{x}_1)^\top; \dots; \sqrt{N_{nV}^{(t)}} \phi(\tilde{x}_{nV})^\top]$ is the empirically weighted feature vector, $K^{(t)} = \Phi_t \Phi_t^\top = [\sqrt{N_i^{(t)} N_j^{(t)}} K(\tilde{x}_i, \tilde{x}_j)]_{i,j \in [nV]}$ is the kernel matrix, and $\bar{y}^{(t)} = [\sqrt{N_1^{(t)}} \bar{y}_1^{(t)}, \dots, \sqrt{N_{nV}^{(t)}} \bar{y}_{nV}^{(t)}]^\top$ is the average observations. Equality (a) rearranges the summation according to different chosen arms, and (b) follows from kernel transformation.

Then, by multiplying $(\phi(\tilde{x}_i) - \phi(\tilde{x}_j))^\top$, we obtain the estimator of reward gaps $f(\tilde{x}_i) - f(\tilde{x}_j)$ as

$$\hat{\Delta}(\tilde{x}_i, \tilde{x}_j) = (\phi(\tilde{x}_i) - \phi(\tilde{x}_j))^\top \hat{\theta}_t = (k_t(\tilde{x}_i) - k_t(\tilde{x}_j))^\top \left(N^{(t)} \xi_t I + K^{(t)} \right)^{-1} \bar{y}^{(t)}, \tag{5}$$

where $k_t(\tilde{x}) = \Phi_t \phi(\tilde{x}) = [\sqrt{N_1^{(t)}} K(\tilde{x}, \tilde{x}_1), \dots, \sqrt{N_{nV}^{(t)}} K(\tilde{x}, \tilde{x}_{nV})]^\top$ for any $\tilde{x} \in \tilde{\mathcal{X}}$. This estimator not only transforms heavy operations on the infinite-dimensional RKHS to efficient ones that only query the kernel function, but also merges repetitive computations for same arms (equality (a)) and only requires calculations dependent on nV.

Kernelized Optimization Solver/Condition for Regularization Parameter. Now we introduce the optimization solver (Line 4) and condition for regularization parameter ξ_t (Eq. (3)).

For the kernelized optimization solver, we solve the min-max optimization in Line 4 by projected gradient descent, which follows the procedure in [10]. Specifically, let $A(\xi, \lambda) = \xi I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda_{\tilde{x}} \phi(\tilde{x}) \phi(\tilde{x})^\top$ for any $\xi > 0, \lambda \in \triangle_{\tilde{\mathcal{X}}}$. We define function $h(\lambda) = \max_{\tilde{x}_i, \tilde{x}_j \in \mathcal{B}_v^{(t)}, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{A(\xi_t, \lambda)^{-1}}$, and denote the optimal solution of $h(\lambda)$ by $\tilde{x}_i^*(\lambda), \tilde{x}_j^*(\lambda)$. Then, the gradient of $h(\lambda)$ is given by

$$[\nabla_\lambda h(\lambda)]_{\tilde{x}} = -\left(\left(\phi(\tilde{x}_i^*(\lambda)) - \phi(\tilde{x}_j^*(\lambda))\right)^\top A(\xi_t, \lambda)^{-1} \phi(\tilde{x})\right)^2, \ \forall \tilde{x} \in \tilde{\mathcal{X}}, \tag{6}$$

which can be efficiently calculated by the following kernel transformation

$$\begin{aligned}&\left(\phi(\tilde{x}_i^*(\lambda)) - \phi(\tilde{x}_j^*(\lambda))\right)^\top A(\xi_t, \lambda)^{-1} \phi(\tilde{x}) \\ =&\xi_t^{-1}\left(K(\tilde{x}_i^*(\lambda), \tilde{x}) - K(\tilde{x}_j^*(\lambda), \tilde{x}) - \left(k_\lambda(\tilde{x}_i^*(\lambda)) - k_\lambda(\tilde{x}_j^*(\lambda))\right)^\top (\xi_t I + K_\lambda)^{-1} k_\lambda(\tilde{x})\right),\end{aligned} \tag{7}$$

where $k_\lambda(\tilde{x}) = [\frac{1}{\sqrt{\lambda_1}} K(\tilde{x}, \tilde{x}_1), \dots, \frac{1}{\lambda_{nV}} K(\tilde{x}, \tilde{x}_{nV})]^\top$ and $K_{\lambda_u} = [\frac{1}{\sqrt{\lambda_i \lambda_j}} K(\tilde{x}_i, \tilde{x}_j)]_{i,j \in [nV]}$.

For condition Eq. (3) on the regularization parameter ξ_t, we can transform it to

$$\max_{\substack{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{X}}_v \\ v \in [V]}} \sqrt{\left(K(\tilde{x}_i, \tilde{x}_i) + K(\tilde{x}_j, \tilde{x}_j) - 2K(\tilde{x}_i, \tilde{x}_j)\right) - \|k_{\lambda_u}(\tilde{x}_i) - k_{\lambda_u}(\tilde{x}_j)\|^2_{(\xi_t I + K_{\lambda_u})^{-1}}} \leq \frac{1}{(1+\varepsilon) B \cdot 2^{t+1}}, \tag{8}$$

where $\lambda_u = [\frac{1}{nV}, \dots, \frac{1}{nV}]^\top$ is the uniform distribution on $\tilde{\mathcal{X}}$, $k_{\lambda_u}(\tilde{x}) = [\frac{1}{\sqrt{nV}} K(\tilde{x}, \tilde{x}_1), \dots, \frac{1}{\sqrt{nV}} K(\tilde{x}, \tilde{x}_{nV})]^\top$ and $K_{\lambda_u} = [\frac{1}{nV} K(\tilde{x}_i, \tilde{x}_j)]_{i,j \in [nV]}$.

Both the kernelized optimization solver and condition for ξ_t avoid inefficient operations directly on infinite-dimensional RKHS by querying the kernel function, and only cost $\text{Poly}(nV)$ computation (memory) complexity (Eqs. (7),(8) only contains scalar $K(\tilde{x}_i, \tilde{x}_j)$, nV-dimensional vector k_λ and $nV \times nV$-dimensional matrix K_λ).

Communication Efficiency. By taking advantage of the kernelized estimator (Eq. (5)), CoopKernelFC merges repetitive computations for the same arms and only transmits nV scalar tuples $\{(N_{v,i}^{(t)}, \bar{y}_{v,i}^{(t)})\}_{i \in [n], v \in [V]}$ among agents instead of transmitting all $N^{(t)}$ samples as in [16]. This significantly reduces the communication cost from $O(N^{(t)})$ bits to $O(nV)$ bits (Lines 9-10).

4.2 Theoretical performance of CoopKernelFC

Define the problem hardness of identifying the best arms $\tilde{x}_v^*$ for all $v \in [V]$ as

$$\rho^* = \min_{\lambda \in \triangle_{\tilde{\mathcal{X}}}} \max_{\tilde{x} \in \tilde{\mathcal{X}}_v, v \in [V]} \frac{\|\phi(\tilde{x}_v^*) - \phi(\tilde{x})\|^2_{(\xi_* I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda_{\tilde{x}} \phi(\tilde{x}) \phi(\tilde{x})^\top)^{-1}}}{(f(\tilde{x}_v^*) - f(\tilde{x}))^2}, \tag{9}$$

where $\xi_* = \min_{t \geq 1} \xi_t$. ρ^* is the information-theoretic lower bound of the CoPE-KB problem, which is adapted from linear/kernel bandit pure exploration [10, 19, 23, 43]. Let S denote the per-agent sample complexity, i.e., average number of samples used by each agent in algorithm CoopKernelFC.

The sample complexity and number of communication rounds of CoopKernelFC are as follows.

Theorem 1 (Fixed-Confidence Upper Bound). *With probability at least $1-\delta$, algorithm CoopKernelFC returns the correct answers $\tilde{x}_v^*$ for all $v \in [V]$, with per-agent sample complexity*

$$S = O\left(\frac{\rho^*}{V} \cdot \log \Delta_{\min}^{-1} \left(\log\left(\frac{nV}{\delta}\right) + \log\log \Delta_{\min}^{-1}\right)\right)$$

and communication rounds $O(\log \Delta_{\min}^{-1})$.

Remark 1. ρ^* is comprised of two sources of problem hardness, one due to handling different tasks and the other due to distinguishing different arms (We will decompose the sample complexity into these two parts in Corollary 1(c)). We see that the sample complexity of CoopKernelFC matches the lower bound (up to logarithmic factors). For fully-collaborative instances where single-agent algorithms [19, 23] have $\tilde{O}(\rho^* \log \delta^{-1})$ sample complexity, our CoopKernelFC achieves the maximum V-speedup (i.e., enjoys $\tilde{O}(\frac{\rho^*}{V} \log \delta^{-1})$ sample complexity) using only logarithmic communication rounds.

Interpretation. We further interpret Theorem 1 via standard expressive tools in kernel bandits [14, 35, 40], i.e., effective dimension and maximum information gain, to characterize the relationship between sample complexity and data structures, and demonstrate how task similarity influences learning performance.

To this end, define the maximum information gain over all sample allocation $\lambda \in \triangle_{\tilde{\mathcal{X}}}$ as

$$\Upsilon = \max_{\lambda \in \triangle_{\tilde{\mathcal{X}}}} \log\det\left(I + {\xi_*}^{-1} K_\lambda\right).$$

Denote $\lambda^* = \mathrm{argmax}_{\lambda \in \triangle_{\tilde{\mathcal{X}}}} \log\det\left(I + {\xi_*}^{-1} K_\lambda\right)$ and $\alpha_1 \geq \cdots \geq \alpha_{nV}$ the eigenvalues of K_{λ^*}, and define the effective dimension of K_{λ^*} as

$$d_{\text{eff}} = \min\left\{ j : j\xi_* \log(nV) \geq \sum_{i=j+1}^{nV} \alpha_i \right\}.$$

We then have the following corollary.

Corollary 1. *The per-agent sample complexity of algorithm CoopKernelFC, denoted by S, can also be bounded as follows:*

(a) $S = O\left(\frac{\Upsilon}{\Delta_{\min}^2 V} \cdot g(\Delta_{\min}, \delta)\right)$, *where Υ is the maximum information gain.*

(b) $S = O\left(\frac{d_{\text{eff}}}{\Delta_{\min}^2 V} \cdot \log\left(nV \cdot \left(1 + \frac{\mathrm{Trace}(K_{\lambda^*})}{\xi_* d_{\text{eff}}}\right)\right) g(\Delta_{\min}, \delta)\right)$, *where d_{eff} is the effective dimension.*

(c) $S = O\left(\frac{\mathrm{rank}(K_z) \cdot \mathrm{rank}(K_x)}{\Delta_{\min}^2 V} \cdot \log\left(\frac{\mathrm{Trace}\left(I + \xi_*^{-1} K_{\lambda^*}\right)}{\mathrm{rank}(K_{\lambda^*})}\right) g(\Delta_{\min}, \delta)\right)$.

Here $g(\Delta_{\min}, \delta) = \log \Delta_{\min}^{-1} \left(\log\left(\frac{nV}{\delta}\right) + \log\log \Delta_{\min}^{-1}\right)$.

Remark 2. Corollary 1(a) shows that, our sample complexity can be bounded by the maximum information gain of any sample allocation on $\tilde{\mathcal{X}}$, which extends conventional information-gain-based results in regret minimization kernel bandits [13, 16, 35] to the pure exploration setting in the view of experimental (allocation) design.

In terms of dimension dependency, it is demonstrated in Corollary 1(b) that our result only depends on the effective dimension of kernel representation, which is the number of principle directions that data projections in RKHS spread.

We also provide a fundamental decomposition of sample complexity into two compositions from task similarities and arm features in Corollary 1(c), which shows that the more tasks are similar, the fewer samples we need for accomplishing all tasks. For example, when tasks are the same (fully-collaborative), i.e., $\mathrm{rank}(K_z) = 1$, each agent only spends a $\frac{1}{V}$ fraction of samples used by single-agent algorithms [10, 43]. Conversely, when the tasks are totally different, i.e., $\mathrm{rank}(K_z) = V$, no advantage can be attained by multi-agent deployments, since the information from neighboring agents is useless for solving local tasks.

4.3 Lower Bound for Fixed-Confidence Setting

We now present lower bounds for the sample complexity and a round-speedup for fully-collaborative instances, using a novel measure transformation techniques. The bounds validate the optimality of CoopKernelFC in both sampling and communication. Specifically, Theorems 2 and 3 below formally present our bounds. In the theorems, we refer to a distributed algorithm $\mathcal{A}$ for CoPE-KB as δ-correct, if it returns the correct answers $\tilde{x}_v^*$ for all $v \in [V]$ with probability at least $1 - \delta$.

Theorem 2 (Fixed-Confidence Sample Complexity Lower Bound). *Consider the fixed-confidence collaborative pure exploration in kernel bandit problem with Gaussian noise $\eta_{v,t}$. Given any $\delta \in (0, 1)$, a δ-correct distributed algorithm $\mathcal{A}$ must have per-agent sample complexity $\Omega(\frac{\rho^*}{V} \log \delta^{-1})$.*

Remark 3. Theorem 2 shows that even if the agents are allowed to share samples without limitation, each agent still requires at least $\tilde{\Omega}(\frac{\rho^*}{V})$ samples on average. Together with Theorem 1, one sees that CoopKernelFC is within logarithmic factors of the optimal sampling.

Theorem 3 (Fixed-Confidence Round-Speedup Lower Bound). *There exists a fully-collaborative instance of the fixed-confidence CoPE-KB problem with multi-armed and linear reward structures, for which given any $\delta \in (0, 1)$, a δ-correct and β-speedup distributed algorithm $\mathcal{A}$ must utilize*

$$\Omega\left(\frac{\log \Delta_{\min}^{-1}}{\log(1 + \frac{V}{\beta}) + \log \log \Delta_{\min}^{-1}}\right)$$

communication rounds in expectation. In particular, when $\beta = V$, $\mathcal{A}$ must require $\Omega(\frac{\log \Delta_{\min}^{-1}}{\log \log \Delta_{\min}^{-1}})$ communication rounds in expectation.

Remark 4. Theorem 3 exhibits that logarithmic communication rounds are indispensable for achieving the full speedup, which validates that CoopKernelFC is near-optimal in communication. Moreover, when CoPE-KB reduces to prior CoPE with classic MAB setting [20, 38], i.e., all agents are solving the same classic MAB task, our upper and lower bounds (Theorems 1 and 3) match the state-of-the-art results in [38].

Novel Analysis for Fixed-Confidence Round-Speedup Lower Bound. We highlight that our round-speedup lower bound for the FC setting analysis has the following novel aspects. (i) Unlike prior CoPE work [38] which focuses on a preliminary 2-armed case without considering reward structures, we investigate multi-armed instances with high-dimensional linear reward structures. (ii) We develop a *linear structured progress lemma* (Lemma 3 in Appendix A.5), which effectively handles the challenges due to different possible sample allocation on multiple arms and derives the required communication rounds under linear reward structures. (iii) We propose *multi-armed measure transformation* and *linear structured instance transformation lemmas* (Lemmas 4,5 in Appendix A.5), which bound the change of probability measures in instance transformation with multiple arms and high-dimensional linear rewards, and serve as basic analytical tools in our proof.

5 FIXED-BUDGET COPE-KB

We now turn to the fixed-budget (FB) setting and design an efficient algorithm CoopKernelFB. We also establish a fixed-budget round-speedup lower bound to validate its communication optimality.

Algorithm 2: Distributed Algorithm CoopKernelFB: for Agent v ($v \in [V]$)

Input: Per-agent budget $T, \tilde{\mathcal{X}}_1, \dots, \tilde{\mathcal{X}}_V, K(\cdot,\cdot) : \tilde{\mathcal{X}} \times \tilde{\mathcal{X}} \mapsto \mathbb{R}$, regularization parameter ξ_*, rounding procedure $\mathsf{ROUND}_\varepsilon(\cdot,\cdot)$ with approximation parameter ε.

1 **Initialization:** $R \leftarrow \lceil \log_2(\omega(\tilde{\mathcal{X}})) \rceil$. $N \leftarrow \lfloor TV/R \rfloor$. $\mathcal{B}_{v'}^{(1)} \leftarrow \mathcal{X}_{v'}$ for all $v' \in [V]$. $t \leftarrow 1$; // pre-determine the number of phases and the number of samples for each phase

2 **while** $t \leq R$ **and** $\exists v' \in [V], |\mathcal{B}_{v'}^{(t)}| > 1$ **do**

3 Let λ_t^* and ρ_t^* be the optimal solution and optimal value of $\min_{\lambda \in \triangle_{\tilde{\mathcal{X}}}} \max_{\tilde{x}_i, \tilde{x}_j \in \mathcal{B}_{v'}^{(t)}, v' \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{(\xi_* I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda_{\tilde{x}} \phi(\tilde{x}) \phi(\tilde{x})^\top)^{-1}}$; // compute the optimal sample allocation

4 $(\tilde{s}_1, \dots, \tilde{s}_{N^{(t)}}) \leftarrow \mathsf{ROUND}_\varepsilon(\lambda_t^*, N^{(t)})$;

5 Let $\tilde{\mathbf{s}}_v^{(t)}$ be the sub-sequence of $(\tilde{s}_1, \dots, \tilde{s}_N)$ which only contains the arms in $\tilde{\mathcal{X}}_v$; // generate the sample sequence for agent v

6 Pull arms $\tilde{\mathbf{s}}_v^{(t)}$ and observe random rewards $\boldsymbol{y}_v^{(t)}$;

7 Broadcast $\{(N_{v,i}^{(t)}, \bar{y}_{v,i}^{(t)})\}_{i \in [n]}$;

8 Receive $\{(N_{v',i}^{(t)}, \bar{y}_{v',i}^{(t)})\}_{i \in [n]}$ from all other agents $v' \in [V] \setminus \{v\}$;

9 $k_t(\tilde{x}) \leftarrow [\sqrt{N_1^{(t)}} K(\tilde{x}, \tilde{x}_1), \dots, \sqrt{N_{nV}^{(t)}} K(\tilde{x}, \tilde{x}_{nV})]^\top$ for any $\tilde{x} \in \tilde{\mathcal{X}}$. $K^{(t)} \leftarrow [\sqrt{N_i^{(t)} N_j^{(t)}} K(\tilde{x}_i, \tilde{x}_j)]_{i,j \in [nV]}$. $\bar{\boldsymbol{y}}^{(t)} \leftarrow [\sqrt{N_1^{(t)}} \bar{y}_1^{(t)}, \dots, \sqrt{N_{nV}^{(t)}} \bar{y}_{nV}^{(t)}]^\top$; // organize overall observation information

10 **for** all $v' \in [V]$ **do**

11 $\hat{f}_t(\tilde{x}) \leftarrow k_t(\tilde{x})^\top (K^{(t)} + N^{(t)} \xi_* I)^{-1} \bar{\boldsymbol{y}}^{(t)}$ for all $\tilde{x} \in \mathcal{B}_{v'}^{(t)}$; // estimate the rewards of alive arms

12 Sort all $\tilde{x} \in \mathcal{B}_{v'}^{(t)}$ by $\hat{f}_t(\tilde{x})$ in decreasing order. Let $\tilde{x}_{(1)}, \dots, \tilde{x}_{(|\mathcal{B}_{v'}^{(t)}|)}$ denote the sorted arm sequence;

13 Let i_{t+1} be the largest index such that $\omega(\{\tilde{x}_{(1)}, \dots, \tilde{x}_{(i_{t+1})}\}) \leq \omega(\mathcal{B}_{v'}^{(t)})/2$;

14 $\mathcal{B}_{v'}^{(t+1)} \leftarrow \{\tilde{x}_{(1)}, \dots, \tilde{x}_{(i_{t+1})}\}$; // cut down the alive arm set to half dimension

15 $t \leftarrow t + 1$;

16 **return** $\mathcal{B}_1^{(t)}, \dots, \mathcal{B}_V^{(t)}$;

5.1 Algorithm CoopKernelFB

5.1.1 Algorithm. CoopKernelFB consists of three key steps: (i) pre-determine the numbers of phases and samples according to data dimension, (ii) maintain alive arm sets for all agents, plan a globally optimal sample allocation, (iii) communicate observation information and cut down alive arms to a half in the dimension sense.

The procedure of CoopKernelFB is given in Algorithm 2. During initialization, we determine the number of phases R and the number of samples for each phase N according to the principle dimension $\omega(\tilde{\mathcal{X}})$ (Line 1), defined as:

$$\omega(\tilde{\mathcal{S}}) = \min_{\lambda \in \triangle_{\tilde{\mathcal{X}}}} \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{S}}} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{(\xi_* I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda_{\tilde{x}} \phi(\tilde{x}) \phi(\tilde{x})^\top)^{-1}}, \quad \forall\, \tilde{\mathcal{S}} \subseteq \tilde{\mathcal{X}}$$

i.e., the principle dimension of data projections in $\tilde{\mathcal{S}}$ to the RKHS. In each phase t, each agent v maintains alive arm sets $\mathcal{B}_{v'}^{(t)}$ for all agents $v' \in [V]$, and solves an integrated optimization to obtain a globally optimal sample allocation λ_t^* (Line 3). Then, she generates a sample sequence $(\tilde{s}_1^{(t)}, \dots, \tilde{s}_N^{(t)})$ according to λ_t^*, and selects the sub-sequence $\tilde{\mathbf{s}}_v^{(t)}$ that only contains her available arms to perform sampling (Lines 4-5). During communication, she only sends and receives the number of samples $N_{v,i}^{(t)}$ and average observed reward $\bar{y}_{v,i}^{(t)}$ for each arm to and from other agents (Lines 7-8). Using the shared information, she estimates rewards of alive arms and only selects the best half of them in the dimension sense to enter the next phase (Lines 11-14).

5.1.2 Computation and Communication Efficiency. CoopKernelFB also adopts the efficient kernelized optimization solver (Eqs. (6),(7)) to solve the min-max optimization in Line 3 and employs the kernelized estimator (Eq. (4)) to estimate the rewards in Line 11. Moreover, CoopKernelFB only spends $\mathrm{Poly}(nV)$ computation time and $O(nV)$-bit communication costs.

5.2 Theoretical performance of CoopKernelFB

We present the error probability of CoopKernelFB in the following theorem, where $\lambda_u = \frac{1}{nV}\mathbf{1}$.

Theorem 4 (Fixed-Budget Upper Bound). *Suppose $\omega(\{\tilde{x}_v^*, \tilde{x}\}) \geq 1$ for any $\tilde{x} \in \mathcal{X}_v \setminus \{\tilde{x}_v^*\}, v \in [V]$ and $T = \Omega(\rho^* \log(\omega(\tilde{\mathcal{X}})))$, and the regularization parameter $\xi_* > 0$ satisfies $\sqrt{\xi_*} \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{X}}_v, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{A(\xi_*, \lambda_u)^{-1}} \leq \frac{\Delta_{\min}}{2(1+\varepsilon)B}$. With at most T samples per agent, CoopKernelFB returns the correct answers $\tilde{x}_v^*$ for all $v \in [V]$, with error probability*

$$Err = O\left(n^2 V \log(\omega(\tilde{\mathcal{X}})) \cdot \exp\left(-\frac{TV}{\rho^* \log(\omega(\tilde{\mathcal{X}}))}\right)\right)$$

and communication rounds $O(\log(\omega(\tilde{\mathcal{X}})))$.

Remark 5. Theorem 4 implies that, to guarantee an error probability δ, CoopKernelFB only requires $O(\frac{\rho^* \log(\omega(\tilde{\mathcal{X}}))}{V} \log(\frac{n^2 V \log(\omega(\tilde{\mathcal{X}}))}{\delta}))$ samples, which matches the sample complexity lower bound (Theorem 2) up to logarithmic factors. In addition, CoopKernelFB attains the maximum V-speedup for fully-collaborative instances with only logarithmic communication rounds, which also matches the round-speedup lower bound (Theorem 5) within double logarithmic factors.

Technical Novelty in Error Probability Analysis. Our analysis extends prior single-agent analysis [23] to the multi-agent setting. The single-agent analysis in [23] only uses a single universal Gaussian-process concentration bound. Instead, we establish novel estimate concentrations and high probability events for each arm pair and each agent to handle the distributed environment, and build a connection between the principle dimension $\omega(\mathcal{B}_v^{(t)})$ and problem hardness ρ^* via elimination rules (Lines 13-14) to guarantee the identification correctness.

Interpretation. Similar to Corollary 1, we can also interpret the error probability result with the standard tools of maximum information gain and effective dimension in kernel bandits [14, 35, 40], and decompose the error probability into two compositions from task similarities and arm features.

Corollary 2. *The error probability of algorithm CoopKernelFC, denoted by Err, can also be bounded as follows:*

(a) $Err = O\left(\exp\left(-\frac{TV\Delta_{\min}^2}{\Upsilon \log(\omega(\tilde{\mathcal{X}}))}\right) \cdot n^2 V \log(\omega(\tilde{\mathcal{X}}))\right)$, *where Υ is the maximum information gain.*

(b) $Err = O\left(\exp\left(-\frac{TV\Delta_{\min}^2}{d_{\mathrm{eff}} \log\left(nV \cdot \left(1 + \frac{\mathrm{Trace}(K_{\lambda^*})}{\xi_* d_{\mathrm{eff}}}\right)\right) \log(\omega(\tilde{\mathcal{X}}))}\right) \cdot n^2 V \log(\omega(\tilde{\mathcal{X}}))\right)$, *where d_{eff} is the effective dimension.*

(c) $Err = O\left(\exp\left(-\frac{TV\Delta_{\min}^2}{\mathrm{rank}(K_z) \cdot \mathrm{rank}(K_x) \log\left(\frac{\mathrm{Trace}\left(I + \xi_*^{-1} K_{\lambda^*}\right)}{\mathrm{rank}(K_{\lambda^*})}\right) \log(\omega(\tilde{\mathcal{X}}))}\right) \cdot n^2 V \log(\omega(\tilde{\mathcal{X}}))\right)$.

Remark 6. Corollaries 2(a), 2(b) bound the error probability by maximum information gain and effective dimension, respectively, which capture essential structures of tasks and arm features and only depend on the effective dimension of the feature space of kernel representation. Furthermore, we exhibit how task similarities influence the error probability performance in Corollary 2(c). For example, in the fully-collaborative case where $\text{rank}(K_z) = 1$, the error probability enjoys an exponential decay factor of V compared to conventional single-agent results [23] (achieves a V-speedup). Conversely, when the tasks are totally different with $\text{rank}(K_z) = V$, the error probability degenerates to conventional single-agent results [23], since in this case information sharing brings no benefit.

5.3 Lower Bound for Fixed-Budget Setting

In this subsection, we establish a round-speedup lower bound for the FB setting.

Theorem 5 (Fixed-Budget Round-Speedup Lower Bound). *There exists a fully-collaborative instance of the fixed-budget CoPE-KB problem with multi-armed and linear reward structures, for which given any $\beta \in [\frac{V}{\log(\omega(\tilde{\mathcal{X}}))}, V]$, a β-speedup distributed algorithm $\mathcal{A}$ must utilize*

$$\Omega\left(\frac{\log(\omega(\tilde{\mathcal{X}}))}{\log(\frac{V}{\beta}) + \log\log(\omega(\tilde{\mathcal{X}}))}\right)$$

communication rounds in expectation. In particular, when $\beta = V$, $\mathcal{A}$ must use $\Omega(\frac{\log(\omega(\tilde{\mathcal{X}}))}{\log\log(\omega(\tilde{\mathcal{X}}))})$ communication rounds in expectation.

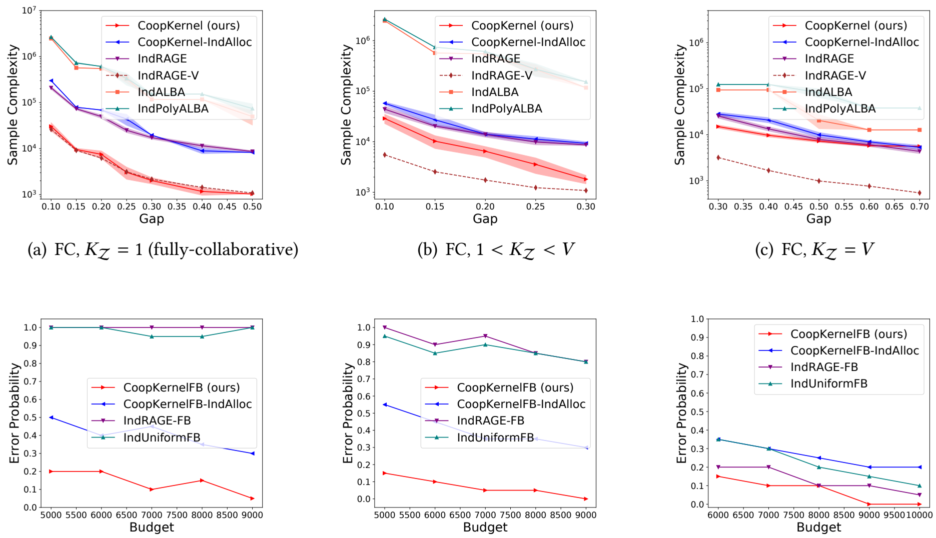
Remark 7. Theorem 5 shows that under the FB setting, to achieve the full speedup, agents require at least logarithmic communication rounds with respect to the principle dimension $\omega(\tilde{\mathcal{X}})$, which validates the communication optimality of CoopKernelFB. In the degenerated case when all agents solve the same non-structured pure exploration problem, same as in prior classic MAB setting [20, 38], both our upper (Theorem 4) and lower (Theorem 5) bounds match the state-of-the-art results in [38].

Novel Analysis for Fixed-Budget Round-Speedup Lower Bound. Different from the FC setting, here we borrow the proof idea of prior limited adaptivity work [1] to establish a non-trivial lower bound analysis under Bayesian environments, and perform instance transformation by changing data dimension instead of tuning reward gaps. In our analysis, we employ novel techniques to calculate the information entropy and support size of posterior reward distributions in order to build induction among different rounds and derive the required communication rounds.

6 EXPERIMENTS

In this section, we conduct experiments to validate the empirical performance of our algorithms. In our experiments, we set $V = 5$, $d = 4$, $n = 6$, $\delta = 0.005$ and $\phi(\tilde{x}) = I\tilde{x}$ for any $\tilde{x} \in \tilde{\mathcal{X}}$. The entries of θ^* form an arithmetic sequence that starts from 0.1 and has the common difference $\Delta_{\min}$, i.e., $\theta^* = [0.1, 0.1 + \Delta_{\min}, \dots, 0.1 + (d-1)\Delta_{\min}]^\top$. For the FC setting, we vary the gap $\Delta_{\min} \in [0.1, 0.8]$ to generate different instances (points), and run 50 independent simulations to plot the average sample complexity with 95% confidence intervals. For the FB setting, we change the budget $T \in [7000, 300000]$ to obtain different instances, and perform 100 independent runs to show the error probability across runs. The specific values of gap $\Delta_{\min}$ and budget T can be seen in X-axis of the figures.

Fixed-Confidence. In the FC setting (Figures 2(a)-2(c)), we compare CoopKernelFC with five baselines: CoopKernel-IndAlloc is an ablation variant of CoopKernelFC which individually calculates sample allocations for different agents. IndRAGE [19], IndALBA [34] and IndPolyALBA [15] are single-agent algorithms, which use V copies of single-agent RAGE [19],



(a) FC, $K_{\mathcal{Z}} = 1$ (fully-collaborative)

(b) FC, $1 < K_{\mathcal{Z}} < V$

(c) FC, $K_{\mathcal{Z}} = V$

(d) FB, $K_{\mathcal{Z}} = 1$ (fully-collaborative)

(e) FB, $1 < K_{\mathcal{Z}} < V$

(f) FB, $K_{\mathcal{Z}} = V$

Fig. 2. Experimental results for FC and FB settings.

ALBA [34] and PolyALBA [15] policies to solve the V tasks independently. IndRAGE/V is a V-speedup baseline, which divides the sample complexity of the best single-agent algorithm IndRAGE by V. One can see that CoopKernelFC achieves the best sample complexity in Figures 2(a),2(b), which demonstrates the effectiveness of our sample allocation and cooperation scheme. Moreover, the empirical results also reflect the impacts of task similarities on learning speedup, and keep consistent with our theoretical analysis. Specifically, in the fully-collaborative case (Figure 2(a)), CoopKernelFC matches IndRAGE-V since it attains the V speedup; in the intermediate ($1 < K_{\mathcal{Z}} < V$) case (Figure 2(b)), the curve of CoopKernelFC lies between IndRAGE/V and IndRAGE, since it only achieves smaller than V speedup due to the decrease of task similarity; in the totally-different-task ($K_{\mathcal{Z}} = V$) case (Figure 2(c)), CoopKernelFC performs similar to the single-agent algorithm IndRAGE, since information sharing among agents brings no advantage in this case.

Fixed-Budget. In the FB setting (Figures 2(d)-2(f)), we compare CoopKernelFB with three baselines: CoopKernelFB-IndAlloc is an ablation variant of CoopKernelFB where agents calculate and use different sample allocations. IndPeaceFB [23] and IndUniformFB solve the V tasks independently by calling V copies of single-agent PeaceFB [23] and uniform sampling policies, respectively. As shown in Figures 2(d),2(e), our CoopKernelFB enjoys a lower error probability than all other algorithms. In addition, these empirical results also validate the influences of task similarities on learning performance, and match our theoretical analysis. Specifically, as the task similarity decreases in Figures 2(d) to 2(f), the error probability of CoopKernelFB gets closer to that of single-agent IndRAGE-FB, due to the slow-down of its learning speedup.

7 CONCLUSION

In this paper, we propose a novel Collaborative Pure Exploration in Kernel Bandit (CoPE-KB) problem with Fixed-Confidence (FC) and Fixed-Budget (FB) settings. CoPE-KB aims to coordinate multiple agents to identify best arms with general reward functions. We design two computation and communication efficient algorithms CoopKernelFC and CoopKernelFB based on novel kernelized estimators. Matching upper and lower bounds are established to demonstrate the statistical and communication optimality of our algorithms. Our theoretical results explicitly characterize the impacts of task similarities on learning speedup and avoid heavy dependency on the high dimension of the kernelized feature space. In our analysis, we also develop novel analytical techniques, including data dimension decomposition, linear structured instance transformation and (communication) round-speedup induction, which are applicable to other bandit problems and can be of independent interests.

REFERENCES

[1] Arpit Agarwal, Shivani Agarwal, Sepehr Assadi, and Sanjeev Khanna. 2017. Learning with limited rounds of adaptivity: Coin tossing, multi-armed bandits, and ranking from pairwise comparisons. In *Conference on Learning Theory*. PMLR, 39–75.

[2] Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. 2021. Near-optimal discrete optimization for experimental design: A regret minimization approach. *Mathematical Programming* 186, 1 (2021), 439–478.

[3] Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. 2010. Best arm identification in multi-armed bandits.. In *Conference on Learning Theory*. Citeseer, 41–53.

[4] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. 2002. Finite-time analysis of the multiarmed bandit problem. *Machine Learning* 47, 2-3 (2002), 235–256.

[5] Ilai Bistritz and Amir Leshem. 2018. Distributed Multi-Player Bandits-a Game of Thrones Approach.. In *Advances in Neural Information Processing Systems*. 7222–7232.

[6] Sébastien Bubeck and Thomas Budzinski. 2020. Coordination without communication: optimal regret in two players multi-armed bandits. In *Conference on Learning Theory*. PMLR, 916–939.

[7] Sébastien Bubeck, Thomas Budzinski, and Mark Sellke. 2021. Cooperative and stochastic multi-player multi-armed bandit: Optimal regret with neither communication nor collisions. In *Conference on Learning Theory*. PMLR, 821–822.

[8] Sébastien Bubeck and Nicolo Cesa-Bianchi. 2012. Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. *Machine Learning* 5, 1 (2012), 1–122.

[9] Séebastian Bubeck, Tengyao Wang, and Nitin Viswanathan. 2013. Multiple identifications in multi-armed bandits. In *International Conference on Machine Learning*. PMLR, 258–265.

[10] Romain Camilleri, Kevin Jamieson, and Julian Katz-Samuels. 2021. High-Dimensional Experimental Design and Kernel Bandits. In *International Conference on Machine Learning*. PMLR, 1227–1237.

[11] Mithun Chakraborty, Kai Yee Phoebe Chua, Sanmay Das, and Brendan Juba. 2017. Coordinated Versus Decentralized Exploration In Multi-Agent Multi-Armed Bandits.. In *Proceedings of the International Joint Conference on Artifical Intelligence*. 164–170.

[12] Lijie Chen, Jian Li, and Mingda Qiao. 2017. Towards instance optimal bounds for best arm identification. In *Conference on Learning Theory*. PMLR, 535–592.

[13] Sayak Ray Chowdhury and Aditya Gopalan. 2017. On kernelized multi-armed bandits. In *International Conference on Machine Learning*. PMLR, 844–853.

[14] Aniket Anand Deshmukh, Urun Dogan, and Clayton Scott. 2017. Multi-task learning for contextual bandits. In *Advances in Neural Information Processing Systems*. 4851–4859.

[15] Yihan Du, Yuko Kuroki, and Wei Chen. 2021. Combinatorial pure exploration with full-bandit or partial linear feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 7262–7270.

[16] Abhimanyu Dubey et al. 2020. Kernel methods for cooperative multi-agent contextual bandits. In *International Conference on Machine Learning*. 2740–2750.

[17] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. 2019. Neural architecture search: A survey. *The Journal of Machine Learning Research* 20, 1 (2019), 1997–2017.

[18] Eyal Even-Dar, Shie Mannor, Yishay Mansour, and Sridhar Mahadevan. 2006. Action Elimination and Stopping Conditions for the Multi-Armed Bandit and Reinforcement Learning Problems. *Journal of Machine Learning Research* 7, 6 (2006).

[19] Tanner Fiez, Lalit Jain, Kevin G Jamieson, and Lillian Ratliff. 2019. Sequential experimental design for transductive linear bandits. *Advances in neural information processing systems* 32 (2019), 10667–10677.

[20] Eshcar Hillel, Zohar S Karnin, Tomer Koren, Ronny Lempel, and Oren Somekh. 2013. Distributed Exploration in Multi-Armed Bandits. In *Advances in Neural Information Processing Systems*, Vol. 26. 854–862.

[21] Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. 2012. PAC subset selection in stochastic multi-armed bandits. In *International Conference on Machine Learning*, Vol. 12. 655–662.

[22] Nikolai Karpov, Qin Zhang, and Yuan Zhou. 2020. Collaborative top distribution identifications with limited interaction. In *IEEE 61st Annual Symposium on Foundations of Computer Science*. 160–171.

[23] Julian Katz-Samuels, Lalit Jain, Kevin G Jamieson, et al. 2020. An Empirical Process Approach to the Union Bound: Practical Algorithms for Combinatorial and Linear Bandits. In *Advances in Neural Information Processing Systems*, Vol. 33.

[24] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. 2016. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research* 17, 1 (2016), 1–42.

[25] Nathan Korda, Balazs Szorenyi, and Shuai Li. 2016. Distributed clustering of linear bandits in peer to peer networks. In *International Conference on Machine Learning*. PMLR, 1301–1309.

[26] Andreas Krause and Cheng Soon Ong. 2011. Contextual Gaussian Process Bandit Optimization.. In *Advances in Neural Information Processing Systems*. 2447–2455.

[27] Tze Leung Lai and Herbert Robbins. 1985. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics* 6, 1 (1985), 4–22.

[28] Tor Lattimore and Csaba Szepesvári. 2020. *Bandit algorithms*. Cambridge University Press.

[29] Keqin Liu and Qing Zhao. 2010. Distributed learning in multi-armed bandit with multiple players. *IEEE Transactions on Signal Processing* 58, 11 (2010), 5667–5681.

[30] Zhenhua Liu, Minghong Lin, Adam Wierman, Steven H Low, and Lachlan LH Andrew. 2011. Geographical load balancing with renewables. 39, 3 (2011), 62–66.

[31] Trung T. Nguyen. 2021. On the Edge and Cloud: Recommendation Systems with Distributed Machine Learning. In *2021 International Conference on Information Technology (ICIT)*. 929–934. https://doi.org/10.1109/ICIT52682.2021.9491121

[32] Jonathan Rosenski, Ohad Shamir, and Liran Szlak. 2016. Multi-player bandits–a musical chairs approach. In *International Conference on Machine Learning*. PMLR, 155–163.

[33] Bernhard Schölkopf, Alexander J Smola, Francis Bach, et al. 2002. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

[34] Marta Soare, Alessandro Lazaric, and Rémi Munos. 2014. Best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, Vol. 27. 828–836.

[35] Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. 2010. Gaussian process optimization in the bandit setting: No regret and experimental design. In *International Conference on Machine Learning*.

[36] Balazs Szorenyi, Róbert Busa-Fekete, István Hegedus, Róbert Ormándi, Márk Jelasity, and Balázs Kégl. 2013. Gossip-based distributed stochastic bandit algorithms. In *International Conference on Machine Learning*. PMLR, 19–27.

[37] Liang Tang, Romer Rosales, Ajit Singh, and Deepak Agarwal. 2013. Automatic ad format selection via contextual bandits. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*. 1587–1594.

[38] Chao Tao, Qin Zhang, and Yuan Zhou. 2019. Collaborative learning with limited interaction: Tight bounds for distributed exploration in multi-armed bandits. In *IEEE 60th Annual Symposium on Foundations of Computer Science*. 126–146.

[39] William R Thompson. 1933. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25, 3/4 (1933), 285–294.

[40] Michal Valko, Nathan Korda, Rémi Munos, Ilias Flaounas, and Nello Cristianini. 2013. Finite-Time Analysis of Kernelised Contextual Bandits. In *Uncertainty in Artificial Intelligence*.

[41] Sofía S Villar, Jack Bowden, and James Wason. 2015. Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 30, 2 (2015), 199.

[42] Grace Wahba. 1990. *Spline models for observational data*. SIAM.

[43] Yinglun Zhu, Dongruo Zhou, Ruoxi Jiang, Quanquan Gu, Rebecca Willett, and Robert Nowak. 2021. Pure Exploration in Kernel and Neural Bandits. *arXiv preprint arXiv:2106.12034* (2021).

[44] Ling Zhuo, Cho-Li Wang, and Francis C.M. Lau. 2003. Document replication and distribution in extensible geographically distributed web servers. *J. Parallel and Distrib. Comput.* 63, 10 (2003), 927–944.

APPENDIX

A PROOFS FOR THE FIXED-CONFIDENCE SETTING

A.1 Kernelized Computations in Algorithm CoopKernelFC

Kernelized Condition for Regularization Parameter ξ_t. We first introduce how to compute the condition Eq. (3) for regularization parameter ξ_t in Line 4 of Algorithm 1.

Let $\Phi_\lambda = [\sqrt{\lambda_1}\phi(\tilde{x}_1)^\top; \dots; \sqrt{\lambda_{nV}}\phi(\tilde{x}_{nV})^\top]$ and $K_\lambda = \Phi_\lambda \Phi_\lambda^\top = [\sqrt{\lambda_i \lambda_j} K(\tilde{x}_i, \tilde{x}_j)]_{i,j \in [nV]}$ for any $\lambda \in \triangle_{\tilde{\mathcal{X}}}$. Let $k_\lambda(\tilde{x}) = \Phi_\lambda \phi(\tilde{x}) = [\sqrt{\lambda_1} K(\tilde{x}, \tilde{x}_1), \dots, \sqrt{\lambda_{nV}} K(\tilde{x}, \tilde{x}_{nV})]^\top$ for any $\lambda \in \triangle_{\tilde{\mathcal{X}}}$, $\tilde{x} \in \tilde{\mathcal{X}}$. Since $\left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda\right) \phi(\tilde{x}) = \xi_t \phi(\tilde{x}) + \Phi_\lambda^\top k_\lambda(\tilde{x})$ for any $\tilde{x} \in \tilde{\mathcal{X}}$, we have

$$
\begin{aligned}
\phi(\tilde{x}) =& \xi_t \left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1} \phi(\tilde{x}) + \left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1} \Phi_\lambda^\top k_\lambda(\tilde{x}) \\
=& \xi_t \left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1} \phi(\tilde{x}) + \Phi_\lambda^\top \left(\xi_t I + K_\lambda\right)^{-1} k_\lambda(\tilde{x})
\end{aligned}
$$

Thus,

$$
\phi(\tilde{x}_i) - \phi(\tilde{x}_j) = \xi_t \left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1} \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right) + \Phi_\lambda^\top \left(\xi_t I + K_\lambda\right)^{-1} \left(k_\lambda(\tilde{x}_i) - k_\lambda(\tilde{x}_j)\right)
$$

Multiplying $\left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top$ on both sides, we have

$$
\begin{aligned}
& \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right) \\
=& \xi_t \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1} \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right) + \left(k_\lambda(\tilde{x}_i) - k_\lambda(\tilde{x}_j)\right)^\top \left(\xi_t I + K_\lambda\right)^{-1} \left(k_\lambda(\tilde{x}_i) - k_\lambda(\tilde{x}_j)\right)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{\left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}} \\
=& \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1} \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right) \\
=& \xi_t^{-1} \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right) - \xi_t^{-1} \left(k_\lambda(\tilde{x}_i) - k_\lambda(\tilde{x}_j)\right)^\top \left(\xi_t I + K_\lambda\right)^{-1} \left(k_\lambda(\tilde{x}_i) - k_\lambda(\tilde{x}_j)\right) \\
=& \xi_t^{-1} \left(K(\tilde{x}_i, \tilde{x}_i) + K(\tilde{x}_j, \tilde{x}_j) - 2K(\tilde{x}_i, \tilde{x}_j)\right) - \xi_t^{-1} \left(k_\lambda(\tilde{x}_i) - k_\lambda(\tilde{x}_j)\right)^\top \left(\xi_t I + K_\lambda\right)^{-1} \left(k_\lambda(\tilde{x}_i) - k_\lambda(\tilde{x}_j)\right) \\
=& \xi_t^{-1} \left(K(\tilde{x}_i, \tilde{x}_i) + K(\tilde{x}_j, \tilde{x}_j) - 2K(\tilde{x}_i, \tilde{x}_j)\right) - \xi_t^{-1} \|k_\lambda(\tilde{x}_i) - k_\lambda(\tilde{x}_j)\|^2_{\left(\xi_t I + K_\lambda\right)^{-1}}
\end{aligned}
$$

Let $\lambda_u = \frac{1}{nV}\mathbf{1}$ be the uniform distribution on $\tilde{\mathcal{X}}$. Then, the condition Eq. (3) for regularization parameter ξ_t

$$
\sqrt{\xi_t} \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{X}}_v, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{\left(\xi_t I + \sum_{x \in \mathcal{X}} \frac{1}{nV} \phi(x)\phi(x)^\top\right)^{-1}} \leq \frac{1}{(1+\varepsilon) B \cdot 2^{t+1}}
$$

is equivalent to the following efficient kernelized statement

$$
\max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{X}}_v, v \in [V]} \sqrt{\left(K(\tilde{x}_i, \tilde{x}_i) + K(\tilde{x}_j, \tilde{x}_j) - 2K(\tilde{x}_i, \tilde{x}_j)\right) - \|k_{\lambda_u}(\tilde{x}_i) - k_{\lambda_u}(\tilde{x}_j)\|^2_{\left(\xi_t I + K_{\lambda_u}\right)^{-1}}} \leq \frac{1}{(1+\varepsilon) B \cdot 2^{t+1}}.
$$

Kernelized Optimization Solver. Now we present the efficient kernelized optimization solver for the following convex optimization in Line 4 of Algorithm 1:

$$
\min_{\lambda \in \triangle_{\tilde{\mathcal{X}}}} \max_{\tilde{x}_i, \tilde{x}_j \in \mathcal{B}_v^{(t)}, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{A(\xi_t, \lambda)^{-1}}, \tag{10}
$$

where $A(\xi, \lambda) = \xi I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda_{\tilde{x}} \phi(\tilde{x}) \phi(\tilde{x})^\top$ for any $\xi > 0, \lambda \in \triangle_{\tilde{\mathcal{X}}}$.

Define function $h(\lambda) = \max_{\tilde{x}_i, \tilde{x}_j \in \mathcal{B}_v^{(t)}, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{A(\xi_t, \lambda)^{-1}}$, and define $\tilde{x}_i^*(\lambda)$, $\tilde{x}_j^*(\lambda)$ as the optimal solution of $h(\lambda)$. Then, the gradient of $h(\lambda)$ with respect to λ is

$$[\nabla_\lambda h(\lambda)]_{\tilde{x}} = -\left(\left(\phi(\tilde{x}_i^*(\lambda)) - \phi(\tilde{x}_j^*(\lambda)) \right)^\top A(\xi_t, \lambda)^{-1} \phi(\tilde{x}) \right)^2, \forall \tilde{x} \in \tilde{\mathcal{X}}. \tag{11}$$

Next, we show how to efficiently compute gradient $[\nabla_\lambda h(\lambda)]_{\tilde{x}}$ with kernel function $K(\cdot, \cdot)$.

Since $\left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda \right) \phi(\tilde{x}) = \xi_t \phi(\tilde{x}) + \Phi_\lambda^\top k_\lambda(\tilde{x})$ for any $\tilde{x} \in \tilde{\mathcal{X}}$, we have

$$\begin{aligned} \phi(\tilde{x}) =& \xi_t \left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda \right)^{-1} \phi(\tilde{x}) + \left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda \right)^{-1} \Phi_\lambda^\top k_\lambda(\tilde{x}) \\ =& \xi_t \left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda \right)^{-1} \phi(\tilde{x}) + \Phi_\lambda^\top \left(\xi_t I + K_\lambda \right)^{-1} k_\lambda(\tilde{x}) \end{aligned}$$

Multiplying $\left(\phi(\tilde{x}_i^*(\lambda)) - \phi(\tilde{x}_j^*(\lambda)) \right)^\top$ on both sides, we have

$$\begin{aligned} & \left(\phi(\tilde{x}_i^*(\lambda)) - \phi(\tilde{x}_j^*(\lambda)) \right)^\top \phi(\tilde{x}) \\ =& \xi_t \left(\phi(\tilde{x}_i^*(\lambda)) - \phi(\tilde{x}_j^*(\lambda)) \right)^\top \left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda \right)^{-1} \phi(\tilde{x}) + \left(k_\lambda(\tilde{x}_i^*(\lambda)) - k_\lambda(\tilde{x}_j^*(\lambda)) \right)^\top \left(\xi_t I + K_\lambda \right)^{-1} k_\lambda(\tilde{x}) \end{aligned}$$

Then,

$$\begin{aligned} & \left(\phi(\tilde{x}_i^*(\lambda)) - \phi(\tilde{x}_j^*(\lambda)) \right)^\top \left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda \right)^{-1} \phi(\tilde{x}) \\ =& \xi_t^{-1} \left(\phi(\tilde{x}_i^*(\lambda)) - \phi(\tilde{x}_j^*(\lambda)) \right)^\top \phi(\tilde{x}) - \xi_t^{-1} \left(k_\lambda(\tilde{x}_i^*(\lambda)) - k_\lambda(\tilde{x}_j^*(\lambda)) \right)^\top \left(\xi_t I + K_\lambda \right)^{-1} k_\lambda(\tilde{x}) \\ =& \xi_t^{-1} \left(K(\tilde{x}_i^*(\lambda), \tilde{x}) - K(\tilde{x}_j^*(\lambda), \tilde{x}) - \left(k_\lambda(\tilde{x}_i^*(\lambda)) - k_\lambda(\tilde{x}_j^*(\lambda)) \right)^\top \left(\xi_t I + K_\lambda \right)^{-1} k_\lambda(\tilde{x}) \right) \end{aligned} \tag{12}$$

Therefore, we can compute gradient $\nabla_\lambda h(\lambda)$ (Eq. (11)) using the equivalent kernelized expression Eq. (12), and then the optimization (Eq. (10)) can be efficiently solved by projected gradient descent.

Innovative Kernelized Estimator. Finally, we explicate the innovative kernelized estimator of reward gaps in Line 13 of Algorithm 1, which plays an important role in boosting the computation and communication efficiency.

Let $\hat{\theta}_t$ denote the minimizer of the following regularized least square loss function:

$$\mathcal{L}(\theta) = N^{(t)} \xi_t \|\theta\|^2 + \sum_{j=1}^{N^{(t)}} (y_j - \phi(\tilde{s}_j)^\top \theta)^2.$$

Letting the derivative of $\mathcal{L}(\theta)$ equal to zero, we have

$$N^{(t)} \xi_t \hat{\theta}_t + \sum_{j=1}^{N^{(t)}} \phi(x_j) \phi(x_j)^\top \hat{\theta}_t = \sum_{j=1}^{N^{(t)}} \phi(x_j) y_j.$$

Rearranging the summation, we can obtain

$$N^{(t)} \xi_t \hat{\theta}_t + \left(\sum_{i=1}^{nV} N_i^{(t)} \phi(\tilde{x}_i) \phi(\tilde{x}_i)^\top \right) \hat{\theta}_t = \sum_{i=1}^{nV} N_i^{(t)} \phi(\tilde{x}_i) \bar{y}_i^{(t)}, \tag{13}$$

where $N_i^{(t)}$ is the number of samples and $\bar{y}_i^{(t)}$ is the average observation on arm $\tilde{x}_i$ for any $i \in [nV]$. Let $\Phi_t = [\sqrt{N_1^{(t)}}\phi(\tilde{x}_1)^\top; \dots; \sqrt{N_{nV}^{(t)}}\phi(\tilde{x}_{nV})^\top]$, $K^{(t)} = \Phi_t \Phi_t^\top = [\sqrt{N_i^{(t)} N_j^{(t)}} K(\tilde{x}_i, \tilde{x}_j)]_{i,j \in [nV]}$ and $\bar{y}^{(t)} = [\sqrt{N_1^{(t)}}\bar{y}_1^{(t)}, \dots, \sqrt{N_{nV}^{(t)}}\bar{y}_{nV}^{(t)}]^\top$. Then, we can write Eq. (13) as

$$\left(N^{(t)}\xi_t I + \Phi_t^\top \Phi_t\right)\hat{\theta}_t = \Phi_t^\top \bar{y}^{(t)}.$$

Since $\left(N^{(t)}\xi_t I + \Phi_t^\top \Phi_t\right) \succ 0$ and $\left(N^{(t)}\xi_t I + \Phi_t \Phi_t^\top\right) \succ 0$,

$$\begin{aligned}
\hat{\theta}_t &= \left(N^{(t)}\xi_t I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \bar{y}^{(t)} \\
&= \Phi_t^\top \left(N^{(t)}\xi_t I + \Phi_t \Phi_t^\top\right)^{-1} \bar{y}^{(t)} \\
&= \Phi_t^\top \left(N^{(t)}\xi_t I + K^{(t)}\right)^{-1} \bar{y}^{(t)}.
\end{aligned}$$

Let $k_t(\tilde{x}) = \Phi_t \phi(\tilde{x}) = [\sqrt{N_1^{(t)}}K(\tilde{x}, \tilde{x}_1), \dots, \sqrt{N_{nV}^{(t)}}K(\tilde{x}, \tilde{x}_{nV})]^\top$ for any $\tilde{x} \in \tilde{\mathcal{X}}$. Then, we obtain the efficient kernelized estimators of $f(\tilde{x}_i)$ and $f(\tilde{x}_i) - f(\tilde{x}_j)$ as

$$\begin{aligned}
\hat{f}(\tilde{x}_i) &= \phi(\tilde{x}_i)^\top \hat{\theta}_t = k_t(\tilde{x}_i)^\top \left(N^{(t)}\xi_t I + K^{(t)}\right)^{-1} \bar{y}^{(t)}, \\
\hat{\Delta}(\tilde{x}_i, \tilde{x}_j) &= \left(k_t(\tilde{x}_i) - k_t(\tilde{x}_j)\right)^\top \left(N^{(t)}\xi_t I + K^{(t)}\right)^{-1} \bar{y}^{(t)}.
\end{aligned}$$

A.2 Proof of Theorem 1

Our proof of Theorem 1 adapts the analysis procedure of [19, 23] to the multi-agent setting.

For any $\lambda \in \triangle_{\tilde{\mathcal{X}}}$, let $\Phi_\lambda = [\sqrt{\lambda_1}\phi(x_v^*)^\top; \dots; \sqrt{\lambda_{nV}}\phi(\tilde{x}_{nV})^\top]$ and $\Phi_\lambda^\top \Phi_\lambda = \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda_{\tilde{x}} \phi(\tilde{x})\phi(\tilde{x})^\top$. In order to prove Theorem 1, we first introduce the following lemmas.

Lemma 1 (Concentration). *Defining event*

$$\begin{aligned}
\mathcal{G} = \Bigg\{ &\left|\left(\hat{f}_t(\tilde{x}_i) - \hat{f}_t(\tilde{x}_j)\right) - \left(f(\tilde{x}_i) - f(\tilde{x}_j)\right)\right| < (1+\varepsilon) \cdot \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_t I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}} \cdot \\
&\left(\sqrt{\frac{2\log\left(2n^2V/\delta_t\right)}{N_t}} + \sqrt{\xi_t} \cdot \|\theta^*\|_2\right) \le 2^{-t},\ \forall \tilde{x}_i, \tilde{x}_j \in \mathcal{B}_v^{(t)},\ \forall v \in [V],\ \forall t \ge 1 \Bigg\},
\end{aligned}$$

we have

$$\Pr[\mathcal{G}] \ge 1 - \delta.$$

Proof of Lemma 1. Let $\hat{\theta}_t$ be the regularized least square estimator of θ^* with samples $x_v^*, \dots, \tilde{x}_{N_t}$ and $\gamma_t = N_t \xi_t$.

Recall that $\Phi_t = [\sqrt{N_1^{(t)}}\phi(x_v^*)^\top; \dots; \sqrt{N_{nV}^{(t)}}\phi(\tilde{x}_{nV})^\top]$ and $\Phi_t^\top \Phi_t = \sum_{i=1}^{nV} N_i^{(t)} \phi(\tilde{x}_i)\phi(\tilde{x}_i)^\top$.

In addition, $\bar{y}^{(t)} = [\sqrt{N_1^{(t)}}\bar{y}_1^{(t)}, \dots, \sqrt{N_{nV}^{(t)}}\bar{y}_{nV}^{(t)}]^\top$ and $\hat{\theta}_t = \left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \bar{y}^{(t)}$.

Let $\bar{\eta}^{(t)} = [\sqrt{N_1^{(t)}}\bar{\eta}_1^{(t)}, \dots, \sqrt{N_{nV}^{(t)}}\bar{\eta}_{nV}^{(t)}]^\top$, where $\bar{\eta}_i^{(t)} = \bar{y}_i^{(t)} - \phi(\tilde{x}_i)^\top \theta^*$ denote the average noise of the $N_i^{(t)}$ pulls on arm $\tilde{x}_i$ for any $i \in [nV]$.

Then,

$$\left(\hat{f}_t(\tilde{x}_i) - \hat{f}_t(\tilde{x}_j)\right) - \left(f(\tilde{x}_i) - f(\tilde{x}_j)\right)$$

$$
\begin{aligned}
&= \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\hat{\theta}_t - \theta^*\right) \\
&= \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \bar{y}^{(t)} - \theta^*\right) \\
&= \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \left(\Phi_t \theta^* + \bar{\eta}^{(t)}\right) - \theta^*\right) \\
&= \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \Phi_t \theta^* + \left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \bar{\eta}^{(t)} - \theta^*\right) \\
&= \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \Big(\left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \left(\Phi_t^\top \Phi_t + \gamma_t I\right) \theta^* + \left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \bar{\eta}^{(t)} \qquad (14) \\
&\quad - \theta^* - \gamma_t \left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \theta^* \Big) \\
&= \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \bar{\eta}^{(t)} - \gamma_t \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \theta^* \qquad (15)
\end{aligned}
$$

Since the mean of the first term is zero and its variance is bounded by

$$
\begin{aligned}
&\left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \Phi_t \left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right) \\
\leq &\left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \left(\gamma_t I + \Phi_t^\top \Phi_t\right) \left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right) \\
= &\left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right) \\
= &\|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{(\gamma_t I + \Phi_t^\top \Phi_t)^{-1}},
\end{aligned}
$$

using the Hoeffding inequality, we have that with probability at least $1 - \frac{\delta_t}{n^2 V}$,

$$
\left|\left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\gamma_t I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \eta_v^{(t)}\right| < \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\gamma_t I + \Phi_t^\top \Phi_t)^{-1}} \sqrt{2 \log\left(\frac{2n^2 V}{\delta_t}\right)}
$$

Thus, with probability at least $1 - \frac{\delta_t}{n^2 V}$,

$$
\begin{aligned}
&\left|\left(\hat{f}_t(\tilde{x}_i) - \hat{f}_t(\tilde{x}_j)\right) - \left(f(\tilde{x}_i) - f(\tilde{x}_j)\right)\right| \\
<&\|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\gamma_t I + \Phi_t^\top \Phi_t)^{-1}} \sqrt{2 \log\left(\frac{2n^2 V}{\delta_t}\right)} + \gamma_t \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\gamma_t I + \Phi_t^\top \Phi_t)^{-1}} \|\theta^*\|_{(\gamma_t I + \Phi_t^\top \Phi_t)^{-1}} \\
\leq&\|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\gamma_t I + \Phi_t^\top \Phi_t)^{-1}} \sqrt{2 \log\left(\frac{2n^2 V}{\delta_t}\right)} + \sqrt{\gamma_t} \cdot \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\gamma_t I + \Phi_t^\top \Phi_t)^{-1}} \|\theta^*\|_2 \\
\overset{\text{(a)}}{\leq}& \frac{(1+\varepsilon) \cdot \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_t I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}}}{\sqrt{N_t}} \sqrt{2 \log\left(\frac{2n^2 V}{\delta_t}\right)} \\
&+ \sqrt{\xi_t N_t} \cdot \frac{(1+\varepsilon) \cdot \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_t I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}}}{\sqrt{N_t}} \cdot \|\theta^*\|_2 \\
=&(1+\varepsilon) \cdot \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_t I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}} \sqrt{\frac{2 \log\left(2n^2 V/\delta_t\right)}{N_t}} \\
&+ (1+\varepsilon) \sqrt{\xi_t} \cdot \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_t I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}} \|\theta^*\|_2 \\
\leq&(1+\varepsilon) \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{B}}_v^{(t)}, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_t I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}} \sqrt{\frac{2 \log\left(2n^2 V/\delta_t\right)}{N_t}}
\end{aligned}
$$

$$+ (1+\varepsilon)\sqrt{\xi_t} \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{B}}_v^{(t)}, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_t I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}} \|\theta^*\|_2,$$

where (a) is due to the rounding procedure.

According to the choice of ξ_t, it holds that

$$\begin{aligned}
&(1+\varepsilon)\sqrt{\xi_t} \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{B}}_v^{(t)}, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{\left(\xi_t I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*}\right)^{-1}} \|\theta^*\|_2 \\
\le &(1+\varepsilon)\sqrt{\xi_t} \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{B}}_v^{(t)}, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{\left(\xi_t I + \Phi_{\lambda_u}^\top \Phi_{\lambda_u}\right)^{-1}} \|\theta^*\|_2 \\
\le &(1+\varepsilon)\sqrt{\xi_t} \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{X}}_v, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{\left(\xi_t I + \Phi_{\lambda_u}^\top \Phi_{\lambda_u}\right)^{-1}} \cdot B \\
\le &\frac{1}{2^{t+1}}.
\end{aligned}$$

Thus, with probability at least $1 - \frac{\delta_t}{n^2 V}$,

$$\begin{aligned}
&\left|\left(\hat{f}_t(\tilde{x}_i) - \hat{f}_t(\tilde{x}_j)\right) - \left(f(\tilde{x}_i) - f(\tilde{x}_j)\right)\right| \\
<&(1+\varepsilon) \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{B}}_v^{(t)}, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_t I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}} \sqrt{\frac{2\log\left(2n^2 V/\delta_t\right)}{N_t}} + \frac{1}{2^{t+1}} \\
=&\sqrt{\frac{2(1+\varepsilon)^2 \rho_t^* \log\left(2n^2 V/\delta_t\right)}{N_t}} + \frac{1}{2^{t+1}} \\
\le&\frac{1}{2^{t+1}} + \frac{1}{2^{t+1}} \\
=&\frac{1}{2^t}
\end{aligned}$$

By a union bound over arms $\tilde{x}_i, \tilde{x}_j$, agent v and phase t, we have that

$$\Pr\left[\mathcal{G}\right] \ge 1 - \delta.$$

□

For any $t > 1$ and $v \in [V]$, let $\mathcal{S}_v^{(t)} = \{\tilde{x} \in \tilde{\mathcal{X}}_v : f(\tilde{x}_v^*) - f(\tilde{x}) \le 2^{-t+2}\}$.

Lemma 2. *Assume that event $\mathcal{G}$ occurs. Then, for any phase $t > 1$ and agent $v \in [V]$, we have that $\tilde{x}_v^* \in \mathcal{B}_v^{(t)}$ and $\mathcal{B}_v^{(t)} \subseteq \mathcal{S}_v^{(t)}$.*

Proof of Lemma 2. We prove the first statement by induction.

To begin, for any $v \in [V]$, $\tilde{x}_v^* \in \mathcal{B}_v^{(1)}$ trivially holds.

Suppose that $\tilde{x}_v^* \in \mathcal{B}_v^{(t)}$ holds for any $v \in [V]$, and there exists some $v' \in [V]$ such that $\tilde{x}_{v'}^* \notin \mathcal{B}_{v'}^{(t+1)}$. According to the elimination rule of algorithm CoopKernelFC, we have that these exists some $\tilde{x}' \in \mathcal{B}_v^{(t)}$ such that

$$\hat{f}_t(\tilde{x}') - \hat{f}_t(\tilde{x}_{v'}^*) \ge 2^{-t}.$$

Using Lemma 1, we have

$$f(\tilde{x}') - f(\tilde{x}_{v'}^*) > \hat{f}_t(\tilde{x}') - \hat{f}_t(\tilde{x}_{v'}^*) - 2^{-t} \ge 0,$$

which contradicts the definition of $\tilde{x}^*_{v'}$. Thus, we have that for any $v \in [V]$, $\tilde{x}^*_v \in \mathcal{B}_v^{(t+1)}$, which completes the proof of the first statement.

Now, we prove the second statement also by induction.

To begin, we prove that for any $v \in [V]$, $\mathcal{B}_v^{(2)} \subseteq \mathcal{S}_v^{(2)}$. Suppose that there exists some $v' \in [V]$ such that $\mathcal{B}_{v'}^{(2)} \subsetneq \mathcal{S}_{v'}^{(2)}$. Then, there exists some $\tilde{x}' \in \mathcal{B}_{v'}^{(2)}$ such that $f(\tilde{x}^*_{v'}) - f(\tilde{x}') > 2^{-2+2} = 1$. Using Lemma 1, we have that at the phase $t = 1$,

$$\hat{f}_t(\tilde{x}^*_{v'}) - \hat{f}_t(\tilde{x}') \geq f(\tilde{x}^*_{v'}) - f(\tilde{x}') - 2^{-1} > 1 - 2^{-1} = 2^{-1},$$

which implies that $\tilde{x}'$ should have been eliminated in phase $t = 1$ and gives a contradiction.

Suppose that $\mathcal{B}_v^{(t)} \subseteq \mathcal{S}_v^{(t)}$ $(t > 1)$ holds for any $v \in [V]$, and there exists some $v' \in [V]$ such that $\mathcal{B}_{v'}^{(t+1)} \subsetneq \mathcal{S}_{v'}^{(t+1)}$. Then, there exists some $\tilde{x}' \in \mathcal{B}_{v'}^{(t+1)}$ such that $f(\tilde{x}^*_{v'}) - f(\tilde{x}') > 2^{-(t+1)+2} = 2 \cdot 2^{-t}$. Using Lemma 1, we have that at the phase t,

$$\hat{f}_t(\tilde{x}^*_{v'}) - \hat{f}_t(\tilde{x}') \geq f(\tilde{x}^*_{v'}) - f(\tilde{x}') - 2^{-t} > 2 \cdot 2^{-t} - 2^{-t} = 2^{-t},$$

which implies that $\tilde{x}'$ should have been eliminated in phase t and gives a contradiction. Thus, we complete the proof of Lemma 2. □

Now we prove Theorem 1.

Proof of Theorem 1. We first prove the correctness.

Let $t^* = \lceil \log_2 \Delta_{\min}^{-1} \rceil + 1$ be the index of the last phase of algorithm CoopKernelFC. According to Lemma 2, when $t = t^*$, $\mathcal{B}_v^{(t)} = \{\tilde{x}^*_v\}$ holds for any $v \in [V]$, and thus algorithm CoopKernelFC returns the correct answer $\tilde{x}^*_v$ for all $v \in [V]$.

Next, we prove the sample complexity.

In algorithm CoopKernelFC, the computation of λ^*_t, ρ^*_t and N_t is the same for all agents, and each agent v just generates partial samples that belong to his arm set $\tilde{\mathcal{X}}_v$ from the total N_t samples. Hence, to bound the overall sample complexity, it suffices to bound $\sum_{t=1}^{t^*} N_t$, and then we can obtain the per-agent sample complexity by dividing V. Let $\varepsilon = 0.1$. We have

$$\begin{aligned}
&\sum_{t=1}^{t^*} N_t \\
=&\sum_{t=1}^{t^*} \left(8(2^t)^2(1+\varepsilon)^2 \rho^*_t \log\left(\frac{2n^2V}{\delta_t}\right) + 1 \right) \\
=&\sum_{t=2}^{t^*} 8(2^t)^2 \left(2^{-t+2}\right)^2 (1+\varepsilon)^2 \frac{\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\tilde{x}_i, \tilde{x}_j \in \mathcal{B}_v^{(t)}, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{\left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}}}{\left(2^{-t+2}\right)^2} \log\left(\frac{4Vn^2t^2}{\delta}\right) \\
&+ N_1 + t^* \\
\leq&\sum_{t=2}^{t^*} \left(128(1+\varepsilon)^2 \frac{\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\tilde{x} \in \mathcal{B}_v^{(t)}, v \in [V]} \|\phi(\tilde{x}^*_v) - \phi(\tilde{x})\|^2_{\left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}}}{\left(2^{-t+2}\right)^2} \log\left(\frac{4Vn^2(t^*)^2}{\delta}\right) \right) + N_1 + t^*
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{t=2}^{t^*} \left(128(1+\varepsilon)^2 \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\tilde{x} \in \mathcal{B}_v^{(t)}, v \in [V]} \frac{\|\phi(\tilde{x}_v^*) - \phi(\tilde{x})\|^2_{\left(\xi_t I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}}}{(f(\tilde{x}_v^*) - f(\tilde{x}))^2} \log\left(\frac{4Vn^2(t^*)^2}{\delta}\right)\right) + N_1 + t^* \\
&\leq \sum_{t=2}^{t^*} \left(128(1+\varepsilon)^2 \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\tilde{x} \in \tilde{\mathcal{X}}_v, v \in [V]} \frac{\|\phi(\tilde{x}_v^*) - \phi(\tilde{x})\|^2_{\left(\xi_* I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}}}{(f(\tilde{x}_v^*) - f(\tilde{x}))^2} \log\left(\frac{4Vn^2(t^*)^2}{\delta}\right)\right) + N_1 + t^* \\
&\leq t^* \cdot \left(128(1+\varepsilon)^2 \rho^* \log\left(\frac{4Vn^2(t^*)^2}{\delta}\right)\right) + N_1 + t^* \\
&= O\left(\rho^* \cdot \log \Delta_{\min}^{-1} \left(\log\left(\frac{Vn}{\delta}\right) + \log\log \Delta_{\min}^{-1}\right)\right)
\end{aligned}
$$

Thus, the per-agent sample complexity is bounded by

$$
O\left(\frac{\rho^*}{V} \cdot \log \Delta_{\min}^{-1} \left(\log\left(\frac{Vn}{\delta}\right) + \log\log \Delta_{\min}^{-1}\right)\right).
$$

Since algorithm CoopKernelFC has at most $t^* = \lceil \log_2 \Delta_{\min}^{-1} \rceil + 1$ phases, the number of communication rounds is bounded by $O(\log \Delta_{\min}^{-1})$. □

A.3 Proof of Corollary 1

Proof of Corollary 1. Recall that $K_\lambda = \Phi_\lambda \Phi_\lambda^\top$ and $\lambda^* = \operatorname{argmax}_{\lambda \in \triangle_{\tilde{\mathcal{X}}}} \log\det\left(I + {\xi_*}^{-1} K_\lambda\right)$. We have that $\log\det\left(I + {\xi_*}^{-1} K_\lambda\right) = \log\det\left(I + {\xi_*}^{-1} \Phi_\lambda^\top \Phi_\lambda\right) = \log\det\left(I + {\xi_*}^{-1} \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda_{\tilde{x}'} \phi(\tilde{x}') \phi(\tilde{x}')^\top\right)$. Then,

$$
\begin{aligned}
\rho^* &= \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\tilde{x} \in \tilde{\mathcal{X}}_v, v \in [V]} \frac{\|\phi(\tilde{x}_v^*) - \phi(\tilde{x})\|^2_{\left(\xi_* I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}}}{(f(\tilde{x}_v^*) - f(\tilde{x}))^2} \\
&\leq \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\tilde{x} \in \tilde{\mathcal{X}}_v, v \in [V]} \frac{\|\phi(\tilde{x}_v^*) - \phi(\tilde{x})\|^2_{\left(\xi_* I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}}}{\Delta_{\min}^2} \\
&= \frac{1}{\Delta_{\min}^2} \cdot \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\tilde{x} \in \tilde{\mathcal{X}}_v, v \in [V]} \|\phi(\tilde{x}_v^*) - \phi(\tilde{x})\|^2_{\left(\xi_* I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}} \\
&\leq \frac{1}{\Delta_{\min}^2} \cdot \min_{\lambda \in \triangle_{\mathcal{X}}} \left(2 \max_{\tilde{x} \in \tilde{\mathcal{X}}} \|\phi(\tilde{x})\|_{\left(\xi_* I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}}\right)^2 \\
&= \frac{4}{\Delta_{\min}^2} \cdot \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\tilde{x} \in \tilde{\mathcal{X}}} \|\phi(\tilde{x})\|^2_{\left(\xi_* I + \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda_{\tilde{x}'} \phi(\tilde{x}') \phi(\tilde{x}')^\top\right)^{-1}} \\
&= \frac{4}{\Delta_{\min}^2} \cdot \max_{\tilde{x} \in \tilde{\mathcal{X}}} \|\phi(\tilde{x})\|^2_{\left(\xi_* I + \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}'} \phi(\tilde{x}') \phi(\tilde{x}')^\top\right)^{-1}} \\
&\overset{(b)}{=} \frac{4}{\Delta_{\min}^2} \cdot \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}} \|\phi(\tilde{x})\|^2_{\left(\xi_* I + \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}'} \phi(\tilde{x}') \phi(\tilde{x}')^\top\right)^{-1}},
\end{aligned}
$$

where (b) is due to Lemma 9.

Since $\lambda^*_{\tilde{x}} \|\phi(\tilde{x})\|^2_{\left(\xi_* I + \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}'} \phi(\tilde{x}')\phi(\tilde{x}')^\top\right)^{-1}} \leq 1$ for any $\tilde{x} \in \mathcal{X}$,

$$
\begin{aligned}
&\sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}} \|\phi(\tilde{x})\|^2_{\left(\xi_* I + \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}'} \phi(\tilde{x}')\phi(\tilde{x}')^\top\right)^{-1}} \\
\leq& 2 \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \log \left(1 + \lambda^*_{\tilde{x}} \|\phi(\tilde{x})\|^2_{\left(\xi_* I + \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}'} \phi(\tilde{x}')\phi(\tilde{x}')^\top\right)^{-1}} \right) \\
\overset{(c)}{\leq}& 2 \log \frac{\det \left(\xi_* I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}} \phi(\tilde{x})\phi(\tilde{x})^\top \right)}{\det \left(\xi_* I \right)} \\
=& 2 \log \det \left(I + \xi_*^{-1} \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}} \phi(\tilde{x})\phi(\tilde{x})^\top \right) \\
=& 2 \log \det \left(I + \xi_*^{-1} K_{\lambda^*} \right),
\end{aligned}
$$

where (c) comes from Lemma 10.

Thus, we have

$$
\begin{aligned}
\rho^* =& \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\tilde{x} \in \tilde{\mathcal{X}}_v, v \in [V]} \frac{\|\phi(\tilde{x}^*_v) - \phi(\tilde{x})\|^2_{\left(\xi_* I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}}}{\left(f(\tilde{x}^*_v) - f(\tilde{x})\right)^2} \\
\leq& \frac{4}{\Delta^2_{\min}} \cdot \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}} \|\phi(\tilde{x})\|^2_{\left(\xi_* I + \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}'} \phi(\tilde{x}')\phi(\tilde{x}')^\top\right)^{-1}} \\
\leq& \frac{8}{\Delta^2_{\min}} \cdot \log \det \left(I + \xi_*^{-1} K_{\lambda^*} \right)
\end{aligned}
\tag{16}
$$

In the following, we interpret the term $\log \det \left(I + \xi_*^{-1} K_{\lambda^*} \right)$ using two standard expressive tools, i.e., maximum information gain and effective dimension, respectively.

Maximum Information Gain. Recall that the maximum information gain over all sample allocation $\lambda \in \triangle_{\tilde{\mathcal{X}}}$ is defined as

$$
\Upsilon = \max_{\lambda \in \triangle_{\tilde{\mathcal{X}}}} \log \det \left(I + {\xi_*}^{-1} K_\lambda \right).
$$

Then, using Eq. (16) and the definitions of λ_*, the per-agent sample complexity is bounded by

$$
\begin{aligned}
& O \left(\frac{\rho^*}{V} \cdot \log \Delta^{-1}_{\min} \left(\log \left(\frac{Vn}{\delta} \right) + \log \log \Delta^{-1}_{\min} \right) \right) \\
=& O \left(\frac{\log \det \left(I + {\xi_*}^{-1} K_{\lambda_*} \right)}{\Delta^2_{\min} V} \cdot \log \Delta^{-1}_{\min} \left(\log \left(\frac{Vn}{\delta} \right) + \log \log \Delta^{-1}_{\min} \right) \right) \\
=& O \left(\frac{\Upsilon}{\Delta^2_{\min} V} \cdot \log \Delta^{-1}_{\min} \left(\log \left(\frac{Vn}{\delta} \right) + \log \log \Delta^{-1}_{\min} \right) \right)
\end{aligned}
$$

Effective Dimension. Recall that $\alpha_1 \geq \cdots \geq \alpha_{nV}$ denote the eigenvalues of K_{λ^*} in decreasing order. The effective dimension of K_{λ^*} is defined as

$$d_{\text{eff}} = \min \left\{ j : j\xi_* \log(nV) \geq \sum_{i=j+1}^{nV} \alpha_i \right\},$$

and it holds that $d_{\text{eff}}\xi_* \log(nV) \geq \sum_{i=d_{\text{eff}}+1}^{nV} \alpha_i$.

Let $\varepsilon = d_{\text{eff}}\xi_* \log(nV) - \sum_{i=d_{\text{eff}}+1}^{nV} \alpha_i$, and thus $\varepsilon \leq d_{\text{eff}}\xi_* \log(nV)$. Then, we have $\sum_{i=1}^{d_{\text{eff}}} \alpha_i = \text{Trace}(K_{\lambda^*}) - \sum_{i=d_{\text{eff}}+1}^{nV} \alpha_i = \text{Trace}(K_{\lambda^*}) - d_{\text{eff}}\xi_* \log(nV) + \varepsilon$ and $\sum_{i=d_{\text{eff}}+1}^{nV} \alpha_i = d_{\text{eff}}\xi_* \log(nV) - \varepsilon$.

$$\begin{aligned}
&\log\det\left(I + \xi_*^{-1} K_{\lambda^*}\right) \\
=&\log\left(\Pi_{i=1}^{nV}\left(1 + \xi_*^{-1}\alpha_i\right)\right) \\
=&\log\left(\Pi_{i=1}^{d_{\text{eff}}}\left(1 + \xi_*^{-1}\alpha_i\right) \cdot \Pi_{i=d_{\text{eff}}+1}^{nV}\left(1 + \xi_*^{-1}\alpha_i\right)\right) \\
\leq&\log\left(\left(1 + \xi_*^{-1} \cdot \frac{\text{Trace}(K_{\lambda^*}) - d_{\text{eff}}\xi_* \log(nV) + \varepsilon}{d_{\text{eff}}}\right)^{d_{\text{eff}}} \left(1 + \xi_*^{-1} \cdot \frac{d_{\text{eff}}\xi_* \log(nV) - \varepsilon}{nV - d_{\text{eff}}}\right)^{nV - d_{\text{eff}}}\right) \\
\leq& d_{\text{eff}} \log\left(1 + \xi_*^{-1} \cdot \frac{\text{Trace}(K_{\lambda^*}) - d_{\text{eff}}\xi_* \log(nV) + \varepsilon}{d_{\text{eff}}}\right) + \log\left(1 + \frac{d_{\text{eff}} \log(nV)}{nV - d_{\text{eff}}}\right)^{nV - d_{\text{eff}}} \\
=& d_{\text{eff}} \log\left(1 + \xi_*^{-1} \cdot \frac{\text{Trace}(K_{\lambda^*}) - d_{\text{eff}}\xi_* \log(nV) + \varepsilon}{d_{\text{eff}}}\right) + \log\left(1 + \frac{d_{\text{eff}} \log(nV - d_{\text{eff}} + d_{\text{eff}})}{nV - d_{\text{eff}}}\right)^{nV - d_{\text{eff}}} \\
\overset{(d)}{\leq}& d_{\text{eff}} \log\left(1 + \xi_*^{-1} \cdot \frac{\text{Trace}(K_{\lambda^*}) - d_{\text{eff}}\xi_* \log(nV) + \varepsilon}{d_{\text{eff}}}\right) + \log\left(1 + \frac{d_{\text{eff}} \log(nV + d_{\text{eff}})}{nV}\right)^{nV} \\
=& d_{\text{eff}} \log\left(1 + \xi_*^{-1} \cdot \frac{\text{Trace}(K_{\lambda^*}) - d_{\text{eff}}\xi_* \log(nV) + \varepsilon}{d_{\text{eff}}}\right) + nV \log\left(1 + \frac{d_{\text{eff}} \log(nV + d_{\text{eff}})}{nV}\right) \\
\leq& d_{\text{eff}} \log\left(1 + \frac{\text{Trace}(K_{\lambda^*})}{\xi_* d_{\text{eff}}}\right) + d_{\text{eff}} \log(nV + d_{\text{eff}}) \\
\leq& d_{\text{eff}} \log\left(2nV \cdot \left(1 + \frac{\text{Trace}(K_{\lambda^*})}{\xi_* d_{\text{eff}}}\right)\right),
\end{aligned}$$

where inequality (d) is due to that $\left(1 + \frac{d_{\text{eff}} \log(x + d_{\text{eff}})}{x}\right)^x$ is monotonically increasing with respect to $x \geq 1$.

Then, using Eq. (16), the per-agent sample complexity is bounded by

$$\begin{aligned}
&O\left(\frac{\rho^*}{V} \cdot \log \Delta_{\min}^{-1}\left(\log\left(\frac{Vn}{\delta}\right) + \log\log \Delta_{\min}^{-1}\right)\right) \\
=&O\left(\frac{\log\det\left(I + {\xi_*}^{-1} K_{\lambda_*}\right)}{\Delta_{\min}^2 V} \cdot \log \Delta_{\min}^{-1}\left(\log\left(\frac{Vn}{\delta}\right) + \log\log \Delta_{\min}^{-1}\right)\right) \\
=&O\left(\frac{d_{\text{eff}}}{\Delta_{\min}^2 V} \cdot \log\left(nV \cdot \left(1 + \frac{\text{Trace}(K_{\lambda^*})}{\xi_* d_{\text{eff}}}\right)\right) \cdot \log \Delta_{\min}^{-1}\left(\log\left(\frac{Vn}{\delta}\right) + \log\log \Delta_{\min}^{-1}\right)\right)
\end{aligned}$$

Decomposition. Let $K = [K(\tilde{x}_i, \tilde{x}_j)]_{i,j \in [nV]}$, $K_z = [K_z(z_v, z_{v'})]_{v,v' \in [V]}$ and $K_x = [K_x(x_i, x_j)]_{i,j \in [nV]}$. Since kernel function $K(\dot{;})$ is a Hadamard composition of $K_z(\dot{;})$ and $K_x(\dot{;})$, it holds that $\text{rank}(K_{\lambda^*}) = \text{rank}(K) \leq \text{rank}(K_z) \cdot \text{rank}(K_x)$.

$$
\begin{aligned}
&\log\det\left(I+\xi_*^{-1}K_{\lambda^*}\right)\\
=&\log\left(\Pi_{i=1}^{nV}\left(1+\xi_*^{-1}\alpha_i\right)\right)\\
=&\log\left(\Pi_{i=1}^{\text{rank}(K_{\lambda^*})}\left(1+\xi_*^{-1}\alpha_i\right)\right)\\
\leq&\log\left(\frac{\sum_{i=1}^{\text{rank}(K_{\lambda^*})}\left(1+\xi_*^{-1}\alpha_i\right)}{\text{rank}(K_{\lambda^*})}\right)^{\text{rank}(K_{\lambda^*})}\\
=&\text{rank}(K_{\lambda^*})\log\left(\frac{\sum_{i=1}^{\text{rank}(K_{\lambda^*})}\left(1+\xi_*^{-1}\alpha_i\right)}{\text{rank}(K_{\lambda^*})}\right)\\
\leq&\text{rank}(K_z)\cdot\text{rank}(K_x)\log\left(\frac{\text{Trace}\left(I+\xi_*^{-1}K_{\lambda^*}\right)}{\text{rank}(K_{\lambda^*})}\right)
\end{aligned}
$$

Then, using Eq. (16), the per-agent sample complexity is bounded by

$$
\begin{aligned}
&O\left(\frac{\rho^*}{V}\cdot\log\Delta_{\min}^{-1}\left(\log\left(\frac{Vn}{\delta}\right)+\log\log\Delta_{\min}^{-1}\right)\right)\\
=&O\left(\frac{\log\det\left(I+\xi_*{}^{-1}K_{\lambda_*}\right)}{\Delta_{\min}^2 V}\cdot\log\Delta_{\min}^{-1}\left(\log\left(\frac{Vn}{\delta}\right)+\log\log\Delta_{\min}^{-1}\right)\right)\\
=&O\left(\frac{\text{rank}(K_z)\cdot\text{rank}(K_x)}{\Delta_{\min}^2 V}\cdot\log\left(\frac{\text{Trace}\left(I+\xi_*^{-1}K_{\lambda^*}\right)}{\text{rank}(K_{\lambda^*})}\right)\cdot\log\Delta_{\min}^{-1}\left(\log\left(\frac{Vn}{\delta}\right)+\log\log\Delta_{\min}^{-1}\right)\right)
\end{aligned}
$$

□

A.4 Proof of Theorem 2

Proof of Theorem 2. Our proof of Theorem 2 adapts the analysis procedure of [19] to the multi-agent setting.

Suppose that $\mathcal{A}$ is a δ-correct algorithm for CoPE-KB. For any $i\in[nV]$, let $\nu_{\theta^*,i}=\mathcal{N}(\phi(\tilde{x}_i)^\top\theta^*,1)$ denote the reward distribution of arm $\tilde{x}_i$, and T_i denote the number of times arm $\tilde{x}_i$ is pulled by algorithm $\mathcal{A}$. Let $\Theta=\{\theta\in\mathcal{H}_K:\exists v,\ \exists\tilde{x}\in\tilde{\mathcal{X}}_v\setminus\{\tilde{x}_v^*\},\ \left(\phi(\tilde{x}_v^*)-\phi(\tilde{x})\right)^\top\theta<0\}$.

Since $\mathcal{A}$ is δ-correct, according to the "Change of Distribution" lemma (Lemma A.3) in [24], we have that for any $\theta\in\Theta$,

$$\sum_{i=1}^{nV}\mathbb{E}[T_i]\cdot\text{KL}(\nu_{\theta^*,i},\nu_{\theta,i})\geq\log(1/2.4\delta).$$

Thus, we have

$$\min_{\theta\in\Theta}\sum_{i=1}^{nV}\mathbb{E}[T_i]\cdot\text{KL}(\nu_{\theta^*,i},\nu_{\theta,i})\geq\log(1/2.4\delta).$$

Let $\boldsymbol{t}^*$ be the optimal solution of the following optimization problem

$$\min\sum_{i=1}^{nV}t_i$$

$$s.t. \min_{\theta \in \mathcal{H}_K} \sum_{i=1}^{nV} t_i \cdot \mathrm{KL}(\nu_{\theta^*,i}, \nu_{\theta,i}) \geq \log(1/2.4\delta).$$

Then, we have

$$\min_{\theta \in \mathcal{H}_K} \sum_{i=1}^{nV} \frac{t_i^*}{\sum_{j=1}^{nV} t_j^*} \cdot \mathrm{KL}(\nu_{\theta^*,i}, \nu_{\theta,i}) \geq \frac{\log(1/2.4\delta)}{\sum_{j=1}^{nV} t_j^*} \geq \frac{\log(1/2.4\delta)}{\sum_{j=1}^{nV} \mathbb{E}[T_j]}.$$

Since $\sum_{i=1}^{nV} \frac{t_i^*}{\sum_{j=1}^{nV} t_j^*} = 1$,

$$\max_{\lambda \in \triangle_{\mathcal{X}}} \min_{\theta \in \Theta} \sum_{i=1}^{nV} \lambda_i \cdot \mathrm{KL}(\nu_{\theta^*,i}, \nu_{\theta,i}) \geq \frac{\log(1/2.4\delta)}{\sum_{i=1}^{nV} \mathbb{E}[T_i]}.$$

Thus, we have

$$\begin{aligned}
\sum_{i=1}^{nV} \mathbb{E}[T_i] \geq & \frac{\log(1/2.4\delta)}{\max_{\lambda \in \triangle_{\mathcal{X}}} \min_{\theta \in \Theta} \sum_{i=1}^{nV} \lambda_i \cdot \mathrm{KL}(\nu_{\theta^*,i}, \nu_{\theta,i})} \\
= & \log(1/2.4\delta) \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\theta \in \Theta} \frac{1}{\sum_{i=1}^{nV} \lambda_i \cdot \mathrm{KL}(\nu_{\theta^*,i}, \nu_{\theta,i})}
\end{aligned} \tag{17}$$

For $\lambda \in \triangle_{\mathcal{X}}$, let $A(\xi_*, \lambda) = \xi_* I + \sum_{i=1}^{nV} \lambda_i \phi(\tilde{x}_i)\phi(\tilde{x}_i)^\top$. For any $\lambda \in \triangle_{\mathcal{X}}, v \in [V], j \in [nV]$ such that $\tilde{x}_j \in \tilde{\mathcal{X}}_v \setminus \{x_v^*\}$, define

$$\theta_j(\lambda) = \theta^* - \frac{(2(\phi(x_v^*) - \phi(\tilde{x}_j))^\top \theta^*) \cdot A(\xi_*, \lambda)^{-1}(\phi(x_v^*) - \phi(\tilde{x}_j))}{(\phi(x_v^*) - \phi(\tilde{x}_j))^\top A(\xi_*, \lambda)^{-1}(\phi(x_v^*) - \phi(\tilde{x}_j))}$$

Here $(\phi(x_v^*) - \phi(\tilde{x}_j))^\top \theta_j(\lambda) = -(\phi(x_v^*) - \phi(\tilde{x}_j))^\top \theta^* < 0$, and thus $\theta_j(\lambda) \in \Theta$.

The KL-divergence between $\nu_{\theta^*,i}$ and $\nu_{\theta_j(\lambda),i}$ is

$$\begin{aligned}
& \mathrm{KL}(\nu_{\theta^*,i}, \nu_{\theta_j(\lambda),i}) \\
= & \frac{1}{2}\left(\phi(\tilde{x}_i)^\top(\theta^* - \theta_j(\lambda))\right)^2 \\
= & \frac{2(\phi(x_v^*) - \phi(\tilde{x}_j))^\top \theta^*)^2 (\phi(x_v^*) - \phi(\tilde{x}_j))^\top \cdot A(\xi_*, \lambda)^{-1} \phi(\tilde{x}_i)\phi(\tilde{x}_i)^\top A(\xi_*, \lambda)^{-1}(\phi(x_v^*) - \phi(\tilde{x}_j))}{\left((\phi(x_v^*) - \phi(\tilde{x}_j))^\top A(\xi_*, \lambda)^{-1}(\phi(x_v^*) - \phi(\tilde{x}_j))\right)^2},
\end{aligned}$$

and thus,

$$\begin{aligned}
& \sum_{i=1}^{nV} \lambda_i \cdot \mathrm{KL}(\nu_{\theta^*,i}, \nu_{\theta_j(\lambda),i}) \\
= & \frac{2(\phi(x_v^*) - \phi(\tilde{x}_j))^\top \theta^*)^2 (\phi(x_v^*) - \phi(\tilde{x}_j))^\top \cdot A(\xi_*, \lambda)^{-1} (\sum_{i=1}^{nV} \lambda_i \phi(\tilde{x}_i)\phi(\tilde{x}_i)^\top) A(\xi_*, \lambda)^{-1}(\phi(x_v^*) - \phi(\tilde{x}_j))}{\left((\phi(x_v^*) - \phi(\tilde{x}_j))^\top A(\xi_*, \lambda)^{-1}(\phi(x_v^*) - \phi(\tilde{x}_j))\right)^2} \\
\leq & \frac{2(\phi(x_v^*) - \phi(\tilde{x}_j))^\top \theta^*)^2}{(\phi(x_v^*) - \phi(\tilde{x}_j))^\top A(\xi_*, \lambda)^{-1}(\phi(x_v^*) - \phi(\tilde{x}_j))} \\
= & \frac{2(f(x_v^*) - f(\tilde{x}_j))^2}{(\phi(x_v^*) - \phi(\tilde{x}_j))^\top A(\xi_*, \lambda)^{-1}(\phi(x_v^*) - \phi(\tilde{x}_j))}
\end{aligned} \tag{18}$$

Let $\mathcal{J} = \{j \in [nV] : \tilde{x}_j \neq x^*_v, \forall v \in [V]\}$ denote the set of indices of all sub-optimal arms. Then, plugging the above Eq. (18) into Eq. (17), we have

$$
\begin{aligned}
\sum_{i=1}^{nV} \mathbb{E}[T_i] \geq & \log(1/2.4\delta) \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\theta \in \Theta} \frac{1}{\sum_{i=1}^{nV} \lambda_i \cdot \mathrm{KL}(\nu_{\theta^*,i}, \nu_{\theta,i})} \\
\geq & \log(1/2.4\delta) \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{j \in \mathcal{J}} \frac{1}{\sum_{i=1}^{nV} \lambda_i \cdot \mathrm{KL}(\nu_{\theta^*,i}, \nu_{\theta_j(\lambda),i})} \\
\geq & \log(1/2.4\delta) \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\tilde{x}_j \in \tilde{\mathcal{X}}_v \setminus \{x^*_v\}, v \in [V]} \frac{(\phi(x^*_v) - \phi(\tilde{x}_j))^\top A(\xi_*, \lambda)^{-1} (\phi(x^*_v) - \phi(\tilde{x}_j))}{2(f(x^*_v) - f(\tilde{x}_j))^2} \\
= & \frac{1}{2} \log(1/2.4\delta) \rho^*,
\end{aligned}
$$

which completes the proof of Theorem 2.

□

A.5 Proof of Theorem 3

Our proof of Theorem 3 generalizes the 2-armed lower bound analysis in [38] to the multi-armed case with linear reward structures. We first introduce some notations and definitions.

Consider a fully-collaborative instance $\mathcal{I}(\mathcal{X}, \theta^*)$ of the CoPE-KB problem, where $\tilde{\mathcal{X}} = \mathcal{X} = \mathcal{X}_v, f = f_v$ for all $v \in [V]$ and $\phi(x_i)^\top \theta^*$ is equal for all $x_i \neq x_*$. Let $\Delta = \Delta_{\min} = (\phi(x_*) - \phi(x_i))^\top \theta^*$ and $c = \frac{\phi(x_i)^\top \theta^*}{\Delta}$ for any $x_i \neq x_*$, where $c > 0$ is a constant. Then, we have that $\frac{\phi(x_*)^\top \theta^*}{\Delta} = \frac{\phi(x_i)^\top \theta^* + \Delta}{\Delta} = 1 + c$.

For any integer $\alpha \geq 0$, let $\mathcal{E}(\alpha, T)$ be the event that $\mathcal{A}$ uses at least α communication rounds and at most T samples before the end of the α-th round, and let $\mathcal{E}^{+1}(\alpha, T)$ be the event that $\mathcal{A}$ uses at least $\alpha + 1$ communication rounds and at most T samples before the end of the α-th round. Let $T_{\mathcal{A}}$ and $T_{\mathcal{A},x_i}$ denote the expected number of samples used by $\mathcal{A}$, and the expected number of samples used on arm x_i by $\mathcal{A}$, respectively. Let λ be the sample allocation of $\mathcal{A}$, i.e., $\lambda_i = \frac{T_{\mathcal{A},x_i}}{T_{\mathcal{A}}}$. Let $\rho(\mathcal{X}, \theta^*) = \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \in \mathcal{X} \setminus \{x_*\}} \frac{\|\phi(x_*) - \phi(x)\|^2_{(\xi_* I + \sum_{x \in \mathcal{X}} \lambda_x \phi(x) \phi(x)^\top)^{-1}}}{(f(x_*) - f(x))^2}$ and $d(\mathcal{X}) = \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{x \in \mathcal{X} \setminus \{x_*\}} \|\phi(x_*) - \phi(x)\|^2_{(\xi_* I + \sum_{x \in \mathcal{X}} \lambda_x \phi(x) \phi(x)^\top)^{-1}}$. Then, we have $\rho(\mathcal{X}, \theta^*) = \frac{d(\mathcal{X})}{\Delta^2}$.

In order to prove Theorem 3, we first prove the following lemmas.

Lemma 3 (Linear Structured Progress Lemma). *For any integer $\alpha \geq 0$ and any $q \geq 1$, we have*

$$
\Pr_{\mathcal{I}(\mathcal{X}, \theta^*)} \left[\mathcal{E}^{+1} \left(\alpha, \frac{\rho(\mathcal{X}, \theta^*)}{Vq} \right) \right] \geq \Pr_{\mathcal{I}(\mathcal{X}, \theta^*)} \left[\mathcal{E} \left(\alpha, \frac{\rho(\mathcal{X}, \theta^*)}{Vq} \right) \right] - 2\delta - \frac{1}{\sqrt{q}}.
$$

Proof of Lemma 3. Let $\mathcal{F}$ be the event that $\mathcal{A}$ uses exactly α communication rounds and at most $\frac{\rho(\mathcal{X}, \theta^*)}{Vq}$ samples before the end of the α-th round. Then, we have

$$
\Pr_{\mathcal{I}(\mathcal{X}, \theta^*)} \left[\mathcal{E}^{+1} \left(\alpha, \frac{\rho(\mathcal{X}, \theta^*)}{Vq} \right) \right] \geq \Pr_{\mathcal{I}(\mathcal{X}, \theta^*)} \left[\mathcal{E} \left(\alpha, \frac{\rho(\mathcal{X}, \theta^*)}{Vq} \right) \right] - \Pr_{\mathcal{I}(\mathcal{X}, \theta^*)} [\mathcal{F}] .
$$

Thus, to prove Lemma 3, it suffices to prove

$$
\Pr_{\mathcal{I}(\mathcal{X}, \theta^*)} [\mathcal{F}] \leq 2\delta + \frac{1}{\sqrt{q}}. \tag{19}
$$

We can decompose $\mathcal{F}$ as

$$
\Pr_{\mathcal{I}(\mathcal{X}, \theta^*)} [\mathcal{F}] = \Pr_{\mathcal{I}(\mathcal{X}, \theta^*)} [\mathcal{F}, \mathcal{A} \text{ returns } x_*] + \Pr_{\mathcal{I}(\mathcal{X}, \theta^*)} [\mathcal{F}, \mathcal{A} \text{ does not return } x_*]
$$

$$= \Pr_{\mathcal{I}(\mathcal{X},\theta^*)} [\mathcal{F}, \mathcal{A} \text{ returns } x_*] + \delta \tag{20}$$

Let $y_i = \phi(x_*) - \phi(x_i)$ for any $i \in [n]$. Let $\theta(\xi_*, \lambda) = \theta^* - \frac{2(y_j^\top \theta^*) A(\xi_*,\lambda)^{-1} y_j}{y_j^\top A(\xi_*,\lambda)^{-1} y_j}$, where $A(\xi_*, \lambda) = \xi_* I + \sum_{i=1}^n \lambda_i \phi(x_i)\phi(x_i)^\top$ and $j = \operatorname{argmax}_{i\in[n]} \frac{y_i^\top A(\xi_*,\lambda)^{-1} y_i}{(y_i^\top \theta^*)^2}$. Let $\mathcal{I}(\mathcal{X}, \theta(\lambda))$ denote the instance where the underlying parameter is $\theta(\lambda)$. Under $\mathcal{I}(\mathcal{X}, \theta(\lambda))$, it holds that $y_j^\top \theta(\lambda) = -y_j^\top \theta^* < 0$, and thus x_* is sub-optimal. Let $\mathcal{D}_{\mathcal{I}}$ denote the product distribution of instance $\mathcal{I}$ with at most $\frac{\rho(\mathcal{X},\theta^*)}{q}$ samples over all agents.

Using the Pinsker's inequality (Lemma 11) and Gaussian KL-divergence computation, we have

$$\begin{aligned}
&\|\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)} - \mathcal{D}_{\mathcal{I}(\mathcal{X},\theta(\lambda))}\|_{\mathrm{TV}} \\
\le&\sqrt{\frac{1}{2}\mathrm{KL}\left(\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)} \| \mathcal{D}_{\mathcal{I}(\mathcal{X},\theta(\lambda))}\right)} \\
\le&\sqrt{\frac{1}{4}\sum_{i\in[n]} (\phi(x_i)^\top(\theta^* - \theta(\lambda)))^2 \cdot \lambda_i T_{\mathcal{A}}} \\
=&\sqrt{\frac{1}{4}\sum_{i\in[n]} \frac{4(y_j^\top\theta^*)^2 \cdot y_j^\top A(\xi_*,\lambda)^{-1}\phi(x_i)\phi(x_i)^\top A(\xi_*,\lambda)^{-1} y_j}{(y_j^\top A(\xi_*,\lambda)^{-1} y_j)^2} \cdot \lambda_i T_{\mathcal{A}}} \\
=&\sqrt{T_{\mathcal{A}} \frac{(y_j^\top\theta^*)^2 \cdot y_j^\top A(\xi_*,\lambda)^{-1}(\sum_{i\in[n]}\lambda_i\phi(x_i)\phi(x_i)^\top) A(\xi_*,\lambda)^{-1} y_j}{(y_j^\top A(\xi_*,\lambda)^{-1} y_j)^2}} \\
\le&\sqrt{T_{\mathcal{A}} \frac{(y_j^\top\theta^*)^2 \cdot y_j^\top A(\xi_*,\lambda)^{-1}(\xi_* I + \sum_{i\in[n]}\lambda_i\phi(x_i)\phi(x_i)^\top) A(\xi_*,\lambda)^{-1} y_j}{(y_j^\top A(\xi_*,\lambda)^{-1} y_j)^2}} \\
=&\sqrt{T_{\mathcal{A}} \cdot \frac{(y_j^\top\theta^*)^2}{y_j^\top A(\xi_*,\lambda)^{-1} y_j}} \\
\le&\sqrt{\frac{\rho(\mathcal{X},\theta^*)}{q} \cdot \frac{1}{\frac{y_j^\top A(\xi_*,\lambda)^{-1} y_j}{(y_j^\top\theta^*)^2}}} \\
\le&\sqrt{\frac{\rho(\mathcal{X},\theta^*)}{q} \cdot \frac{1}{\min_{\lambda\in\triangle_{\mathcal{X}}} \frac{y_j^\top A(\xi_*,\lambda)^{-1} y_j}{(y_j^\top\theta^*)^2}}} \\
=&\frac{1}{\sqrt{q}}
\end{aligned} \tag{21}$$

Since x_* is sub-optimal under $\mathcal{I}(\mathcal{X}, \theta(\lambda))$, using the measure change technique, we have

$$\begin{aligned}
\Pr_{\mathcal{I}(\mathcal{X},\theta^*)} [\mathcal{F}, \mathcal{A} \text{ returns } x_*] \le& \Pr_{\mathcal{I}(\mathcal{X},\theta(\lambda))} [\mathcal{F}, \mathcal{A} \text{ returns } x_*] \\
&+ \|\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)} - \mathcal{D}_{\mathcal{I}(\mathcal{X},\theta(\lambda))}\|_{\mathrm{TV}} \\
\le&\delta + \frac{1}{\sqrt{q}}
\end{aligned}$$

Plugging the above equality into Eq. (20), we have

$$\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\mathcal{F}] \leq 2\delta + \frac{1}{\sqrt{q}},$$

which completes the proof of Lemma 3. □

Let $\mathcal{I}(\mathcal{X},\theta/\kappa)$ denote the instance where the underlying parameter is θ/κ. Under $\mathcal{I}(\mathcal{X},\theta/\kappa)$, the reward gap is $(\phi(x_*)-\phi(x_i))^\top\theta/\kappa = \frac{1}{\kappa}\Delta$ and the sample complexity is $\rho(\mathcal{X},\theta/\kappa) = \frac{\kappa^2 d(\mathcal{X})}{\Delta^2}$. Let $\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}$ and $\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}$ denote the product distributions of instances $\mathcal{I}(\mathcal{X},\theta^*)$ and $\mathcal{I}(\mathcal{X},\theta/\kappa)$ with $T_{\mathcal{A}}$ samples, respectively.

Lemma 4 (Multi-armed Measure Transformation Lemma). *Suppose that algorithm $\mathcal{A}$ uses $T_{\mathcal{A}} = \frac{\rho(\mathcal{X},\theta^*)}{\zeta}$ samples over all agents on instance $\mathcal{I}(\mathcal{X},\theta^*)$, where $\zeta \geq 100$. Then, for any event E on $\mathcal{I}(\mathcal{X},\theta^*)$ and any $Q \geq \zeta$, we have*

$$\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}}[E] \leq \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E] \cdot \exp\left((1+c)^2 d(\mathcal{X}) n \sqrt{\frac{\log(nQ)}{\zeta}}\right) + \frac{1}{nQ^2}.$$

Proof. For any $i \in [n]$, let $Z_{i,1},\dots,Z_{i,T_{\mathcal{A},x_i}}$ denote the observed $T_{\mathcal{A},x_i}$ samples on arm x_i, and define event

$$L_i = \left\{ \sum_{t=1}^{T_{\mathcal{A},x_i}} Z_{i,t} \geq T_{\mathcal{A},x_i} \cdot \phi(x_i)^\top\theta/\kappa + \frac{z}{\phi(x_i)^\top\theta/\kappa - \phi(x_i)^\top\theta^*} \right\},$$

where $z \geq 0$ is a parameter specified later. We also define event $L = \cap_{i\in[n]} L_i$. Then,

$$\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}}[E] \leq \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}}[E, L] + \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}}[\neg L]$$

Using the measure change technique, we bound the term $\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}}[E, L]$ as

$$\begin{aligned}
&\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}}[E, L] \\
=&\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E, L] \cdot \exp\left(-\frac{1}{2}\sum_{i\in[n]}\sum_{t=1}^{T_{\mathcal{A},x_i}}\left((Z_{i,t}-\phi(x_i)^\top\theta/\kappa)^2 - (Z_{i,t}-\phi(x_i)^\top\theta^*)^2\right)\right) \\
=&\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E, L] \cdot \exp\left(-\frac{1}{2}\sum_{i\in[n]}\sum_{t=1}^{T_{\mathcal{A},x_i}}\left((\phi(x_i)^\top\theta/\kappa)^2 - (\phi(x_i)^\top\theta^*)^2 - 2Z_{i,t}(\phi(x_i)^\top\theta/\kappa - \phi(x_i)^\top\theta^*)\right)\right) \\
=&\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E, L] \cdot \exp\Bigg(-\frac{1}{2}\sum_{i\in[n]}\Bigg(\left((\phi(x_i)^\top\theta/\kappa)^2 - (\phi(x_i)^\top\theta^*)^2\right)\cdot T_{\mathcal{A},x_i} \\
&-2\left(\phi(x_i)^\top\theta/\kappa - \phi(x_i)^\top\theta^*\right)\cdot\sum_{t=1}^{T_{\mathcal{A},x_i}} Z_{i,t}\Bigg)\Bigg) \\
\leq&\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E, L] \cdot \exp\Bigg(-\frac{1}{2}\sum_{i\in[n]}\Bigg(\left((\phi(x_i)^\top\theta/\kappa)^2 - (\phi(x_i)^\top\theta^*)^2\right)\cdot T_{\mathcal{A},x_i} \\
&-2\left(\phi(x_i)^\top\theta/\kappa - \phi(x_i)^\top\theta^*\right)\cdot\left(T_{\mathcal{A},x_i}\cdot\phi(x_i)^\top\theta/\kappa + \frac{z}{\phi(x_i)^\top\theta/\kappa - \phi(x_i)^\top\theta^*}\right)\Bigg)\Bigg) \\
=&\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E, L] \cdot \exp\Bigg(-\frac{1}{2}\sum_{i\in[n]}\Bigg(\left((\phi(x_i)^\top\theta/\kappa)^2 - (\phi(x_i)^\top\theta^*)^2\right)\cdot T_{\mathcal{A},x_i}
\end{aligned}$$

$$
\begin{aligned}
& - 2\left(\phi(x_i)^\top\theta/\kappa - \phi(x_i)^\top\theta^*\right)\cdot T_{\mathcal{A},x_i}\cdot\phi(x_i)^\top\theta/\kappa - 2z\Bigg)\Bigg) \\
= & \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E,L]\cdot\exp\left(-\frac{1}{2}\sum_{i\in[n]}\left(\left(\phi(x_i)^\top\theta/\kappa - \phi(x_i)^\top\theta^*\right)\left(\phi(x_i)^\top\theta/\kappa\cdot T_{\mathcal{A},x_i} + \phi(x_i)^\top\theta^*\cdot T_{\mathcal{A},x_i}\right.\right.\right. \\
& \left.\left.\left. - 2\cdot\phi(x_i)^\top\theta/\kappa\cdot T_{\mathcal{A},x_i}\right) - 2z\right)\right) \\
= & \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E,L]\cdot\exp\left(-\frac{1}{2}\sum_{i\in[n]}\left(-\left(\phi(x_i)^\top\theta/\kappa - \phi(x_i)^\top\theta^*\right)^2\cdot T_{\mathcal{A},x_i} - 2z\right)\right) \\
= & \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E,L]\cdot\exp\left(\sum_{i\in[n]}\left(\frac{1}{2}\left(\phi(x_i)^\top\theta/\kappa - \phi(x_i)^\top\theta^*\right)^2\cdot T_{\mathcal{A},x_i} + z\right)\right) \\
= & \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E,L]\cdot\exp\left(\sum_{i\in[n]}\left(\frac{1}{2}(1-\frac{1}{\kappa})^2(\phi(x_i)^\top\theta^*)^2\cdot\frac{d(\mathcal{X})}{\Delta^2\zeta}\cdot\lambda_i + z\right)\right) \\
\le & \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E,L]\cdot\exp\left(\frac{(1+c)^2}{2}\cdot\frac{d(\mathcal{X})}{\zeta} + nz\right)
\end{aligned}
$$

Next, using the Chernoff-Hoeffding inequality, we bound the second term as

$$
\begin{aligned}
\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}}[\neg L] \le & \sum_{i\in[n]}\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}}[\neg L_i] \\
\le & \sum_{i\in[n]}\exp\left(-2\frac{z^2}{(\phi(x_i)^\top\theta/\kappa - \phi(x_i)^\top\theta^*)^2}\cdot\frac{\zeta}{\lambda_i\rho(\mathcal{X},\theta^*)}\right) \\
= & \sum_{i\in[n]}\exp\left(-2\frac{z^2}{(1-\frac{1}{\kappa})^2(\phi(x_i)^\top\theta^*)^2}\cdot\frac{\Delta^2\zeta}{\lambda_i d(\mathcal{X})}\right) \\
\le & \sum_{i\in[n]}\exp\left(-\frac{2}{(1+c)^2}\cdot\frac{z^2\zeta}{d(\mathcal{X})}\right) \\
= & n\cdot\exp\left(-\frac{2}{(1+c)^2}\cdot\frac{z^2\zeta}{d(\mathcal{X})}\right)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}}[E] \le & \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}}[E,L] + \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}}[\neg L] \\
\le & \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E,L]\cdot\exp\left(\frac{(1+c)^2}{2}\cdot\frac{d(\mathcal{X})}{\zeta} + nz\right) + n\cdot\exp\left(-\frac{2}{(1+c)^2}\cdot\frac{z^2\zeta}{d(\mathcal{X})}\right)
\end{aligned}
$$

Let $z = \sqrt{\frac{(1+c)^2 d(\mathcal{X})\log(nQ)}{\zeta}}$. Then, we have

$$
\begin{aligned}
\Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)}}[E] \le & \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E,L]\cdot\exp\left(\frac{(1+c)^2}{2}\cdot\frac{d(\mathcal{X})}{\zeta} + n\sqrt{\frac{(1+c)^2 d(\mathcal{X})\log(nQ)}{\zeta}}\right) \\
& + n\cdot\exp\left(-2\log(nQ)\right)
\end{aligned}
$$

$$\leq \Pr_{\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)}}[E] \cdot \exp\left((1+c)^2 d(\mathcal{X}) n \sqrt{\frac{\log(nQ)}{\zeta}}\right) + \frac{1}{nQ^2}$$

□

Lemma 5 (Linear Structured Instance Transformation Lemma). *For any integer $\alpha \geq 0$, $q \geq 100$ and $\kappa \geq 1$, we have*

$$\Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}\left[\mathcal{E}\left(\alpha+1, \frac{\rho(\mathcal{X},\theta^*)}{Vq} + \frac{\rho(\mathcal{X},\theta^*)}{\beta}\right)\right] \geq \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}\left[\mathcal{E}^{+1}\left(\alpha, \frac{\rho(\mathcal{X},\theta^*)}{Vq}\right)\right] - \delta - \sqrt{\frac{(1+c)^2}{4} \cdot \frac{d(\mathcal{X})}{q}}$$
$$- \left(\exp\left((1+c)^2 d(\mathcal{X}) n \sqrt{\frac{\log(nV)}{\beta}}\right) - 1\right) - \frac{1}{nV}$$

Proof. Let $\ell = \{Z_{i,1}, \dots, Z_{i,T_{\mathcal{A},x_i}}\}_{i \in [n]}$ denote the $T_{\mathcal{A}}$ samples of algorithm $\mathcal{A}$ on instance $\mathcal{I}(\mathcal{X}, \theta^*)$. Let $\mathcal{S}$ denote the set of all possible ℓ, conditioned on which $\mathcal{E}^{+1}(\alpha, \frac{\rho(\mathcal{X},\theta^*)}{Vq})$ holds. Then, we have

$$\sum_{s \in \mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\ell = s] = \sum_{s \in \mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}\left[\mathcal{E}^{+1}\left(\alpha, \frac{\rho(\mathcal{X},\theta^*)}{Vq}\right)\right]$$

For any agent $v \in [V]$, let $\mathcal{K}_v$ be the event that agent v uses more than $\frac{\rho(\mathcal{X},\theta^*)}{\beta}$ samples during the $(\alpha+1)$-st round. Conditioned on $s \in \mathcal{S}$, $\mathcal{K}_v$ only depends on the samples of agent v during the $(\alpha+1)$-st round, and is independent of other agents.

Using the facts that $\mathcal{A}$ is δ-correct and β-speedup and all agents have the same performance on fully-collaborative instances, we have

$$\begin{aligned}
\delta &\geq \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}\left[\mathcal{A} \text{ uses more than } \frac{\rho(\mathcal{X},\theta^*)}{\beta}\right] \\
&\geq \sum_{s \in \mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\ell = s] \cdot \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\mathcal{K}_1 \vee \cdots \vee \mathcal{K}_V | \ell = s] \\
&= \sum_{s \in \mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\ell = s] \cdot \left(1 - \prod_{v \in [V]}\left(1 - \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\mathcal{K}_i | \ell = s]\right)\right) \\
&= \mathcal{E}^{+1}\left(\alpha, \frac{\rho(\mathcal{X},\theta^*)}{Vq}\right) - \sum_{s \in \mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\ell = s] \cdot \prod_{v \in [V]}\left(1 - \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\mathcal{K}_i | \ell = s]\right)
\end{aligned}$$

Rearranging the above inequality, we have

$$\sum_{s \in \mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\ell = s] \cdot \prod_{v \in [V]}\left(1 - \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\mathcal{K}_i | \ell = s]\right) \geq \mathcal{E}^{+1}\left(\alpha, \frac{\rho(\mathcal{X},\theta^*)}{Vq}\right) - \delta \tag{22}$$

Using Lemma 4 with $\zeta = \beta$ and $Q = V$, we have

$$\begin{aligned}
&\Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}\left[\mathcal{E}\left(\alpha+1, \frac{\rho(\mathcal{X},\theta^*)}{Vq} + \frac{\rho(\mathcal{X},\theta^*)}{\beta}\right)\right] \\
&\geq \sum_{s \in \mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}[\ell = s] \cdot \Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}[\neg\mathcal{K}_1 \wedge \cdots \wedge \neg\mathcal{K}_V | \ell = s] \\
&= \sum_{s \in \mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}[\ell = s] \cdot \prod_{v \in [V]}\left(1 - \Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}[\mathcal{K}_i | \ell = s]\right)
\end{aligned}$$

$$
\begin{aligned}
&\geq \sum_{s\in\mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}[\ell=s]\cdot \\
&\quad \prod_{v\in[V]} \max\left\{1-\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\mathcal{K}_i|\ell=s]\cdot\exp\left((1+c)^2 d(\mathcal{X})n\sqrt{\frac{\log(nV)}{\beta}}\right)-\frac{1}{nV^2},0\right\} \\
&\geq \sum_{s\in\mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}[\ell=s]\cdot \\
&\quad \left(\prod_{v\in[V]} \max\left\{1-\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\mathcal{K}_i|\ell=s]\cdot\exp\left((1+c)^2 d(\mathcal{X})n\sqrt{\frac{\log(nV)}{\beta}}\right),0\right\}-\frac{1}{nV}\right) \\
&= \sum_{s\in\mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}[\ell=s]\cdot\left(\prod_{v\in[V]} \max\left(1-\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\mathcal{K}_i|\ell=s]\right.\right. \\
&\quad \left.\left.-\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\mathcal{K}_i|\ell=s]\cdot\left(\exp\left((1+c)^2 d(\mathcal{X})n\sqrt{\frac{\log(nV)}{\beta}}\right)-1\right),0\right)-\frac{1}{nV}\right) \\
&\overset{(e)}{\geq} \sum_{s\in\mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}[\ell=s]\cdot \\
&\quad \left(\prod_{v\in[V]} \left(1-\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\mathcal{K}_i|\ell=s]\right)-\left(\exp\left((1+c)^2 d(\mathcal{X})n\sqrt{\frac{\log(nV)}{\beta}}\right)-1\right)-\frac{1}{nV}\right) \\
&\geq \sum_{s\in\mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}[\ell=s]\cdot \\
&\quad \prod_{v\in[V]} \left(1-\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\mathcal{K}_i|\ell=s]\right)-\left(\exp\left((1+c)^2 d(\mathcal{X})n\sqrt{\frac{\log(nV)}{\beta}}\right)-1\right)-\frac{1}{nV} \\
&\geq \sum_{s\in\mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\ell=s]\cdot\prod_{v\in[V]} \left(1-\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\mathcal{K}_i|\ell=s]\right) \\
&\quad -\sum_{s\in\mathcal{S}}\left|\Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}[\ell=s]-\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\ell=s]\right|-\left(\exp\left((1+c)^2 d(\mathcal{X})n\sqrt{\frac{\log(nV)}{\beta}}\right)-1\right)-\frac{1}{nV}
\end{aligned}
$$

where (e) comes from Lemma 12.

Using the Pinsker's inequality (Lemma 11), we have

$$
\begin{aligned}
&\sum_{s\in\mathcal{S}}\left|\Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)}[\ell=s]-\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}[\ell=s]\right| \\
\leq&\sqrt{\frac{1}{2}\mathrm{KL}(\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta/\kappa)},\mathcal{D}_{\mathcal{I}(\mathcal{X},\theta^*)})} \\
\leq&\sqrt{\frac{1}{4}\sum_{i\in[n]}(\phi(x_i)^\top\theta/\kappa-\phi(x_i)^\top\theta^*)^2\cdot\lambda_i T_{\mathcal{A}}} \\
=&\sqrt{\frac{1}{4}\sum_{i\in[n]}\left((1-\frac{1}{\kappa})^2(\phi(x_i)^\top\theta^*)^2\cdot\frac{d(\mathcal{X})}{\Delta^2 q}\lambda_i\right)}
\end{aligned}
\tag{23}
$$

$$\leq \sqrt{\frac{(1+c)^2}{4} \cdot \frac{d(\mathcal{X})}{q}} \tag{24}$$

Thus, using Eqs. (22),(24), we have

$$\begin{aligned}
&\Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)} \left[\mathcal{E}\left(\alpha+1, \frac{\rho(\mathcal{X},\theta^*)}{Vq} + \frac{\rho(\mathcal{X},\theta^*)}{\beta}\right) \right] \\
\geq & \sum_{s \in \mathcal{S}} \Pr_{\mathcal{I}(\mathcal{X},\theta^*)} [\ell = s] \cdot \prod_{v \in [V]} \left(1 - \Pr_{\mathcal{I}(\mathcal{X},\theta^*)} [\mathcal{K}_i | \ell = s]\right) - \sqrt{\frac{(1+c)^2}{4} \cdot \frac{d(\mathcal{X})}{q}} \\
& - \left(\exp\left((1+c)^2 d(\mathcal{X}) n \sqrt{\frac{\log(nV)}{\beta}}\right) - 1\right) - \frac{1}{nV} \\
\geq & \Pr_{\mathcal{I}(\mathcal{X},\theta^*)} \left[\mathcal{E}^{+1}\left(\alpha, \frac{\rho(\mathcal{X},\theta^*)}{Vq}\right) \right] - \delta - \sqrt{\frac{(1+c)^2}{4} \cdot \frac{d(\mathcal{X})}{q}} \\
& - \left(\exp\left((1+c)^2 d(\mathcal{X}) n \sqrt{\frac{\log(nV)}{\beta}}\right) - 1\right) - \frac{1}{nV}
\end{aligned}$$

□

Now we prove Theorem 3.

Proof of Theorem 3. Combining Lemmas 3,5, we have

$$\begin{aligned}
&\Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)} \left[\mathcal{E}\left(\alpha+1, \frac{\rho(\mathcal{X},\theta^*)}{Vq} + \frac{\rho(\mathcal{X},\theta^*)}{\beta}\right) \right] \\
\geq & \Pr_{\mathcal{I}(\mathcal{X},\theta^*)} \left[\mathcal{E}^{+1}\left(\alpha, \frac{\rho(\mathcal{X},\theta^*)}{Vq}\right) \right] - \delta - \sqrt{\frac{(1+c)^2}{4} \cdot \frac{d(\mathcal{X})}{q}} \\
& - \left(\exp\left((1+c)^2 d(\mathcal{X}) n \sqrt{\frac{\log(nV)}{\beta}}\right) - 1\right) - \frac{1}{nV} \\
\geq & \Pr_{\mathcal{I}(\mathcal{X},\theta^*)} \left[\mathcal{E}\left(\alpha, \frac{\rho(\mathcal{X},\theta^*)}{Vq}\right) \right] - 3\delta - \sqrt{\frac{(1+c)^2 d(\mathcal{X})}{q}} \\
& - \left(\exp\left((1+c)^2 d(\mathcal{X}) n \sqrt{\frac{\log(nV)}{\beta}}\right) - 1\right) - \frac{1}{nV}
\end{aligned}$$

Let $\kappa = \sqrt{1 + \frac{Vq}{\beta}}$. Then, we have

$$\begin{aligned}
&\Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)} \left[\mathcal{E}\left(\alpha+1, \frac{\rho(\mathcal{X},\theta/\kappa)}{Vq}\right) \right] \\
= & \Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)} \left[\mathcal{E}\left(\alpha+1, \frac{\kappa^2 \cdot \rho(\mathcal{X},\theta^*)}{Vq}\right) \right] \\
= & \Pr_{\mathcal{I}(\mathcal{X},\theta/\kappa)} \left[\mathcal{E}\left(\alpha+1, \frac{\rho(\mathcal{X},\theta^*)}{Vq} + \frac{\rho(\mathcal{X},\theta^*)}{\beta}\right) \right] \\
\geq & \Pr_{\mathcal{I}(\mathcal{X},\theta^*)} \left[\mathcal{E}\left(\alpha, \frac{\rho(\mathcal{X},\theta^*)}{Vq}\right) \right] - 3\delta - \sqrt{\frac{(1+c)^2 d(\mathcal{X})}{q}}
\end{aligned}$$

$$-\left(\exp\left((1+c)^2 d(\mathcal{X}) n \sqrt{\frac{\log(nV)}{\beta}}\right)-1\right)-\frac{1}{nV} \tag{25}$$

Let $\mathcal{I}(\mathcal{X},\theta_0)$ be the basic instance of induction, where the reward gap is $\Delta_0=(\phi(x_*)-\phi(x_i))^\top\theta_0=1$ for any $i\in[n]$. Let t_0 be the largest integer such that

$$\Delta\cdot\left(1+\frac{Vq}{\beta}\right)^{\frac{t_0}{2}}\leq 1,$$

where $q=1000t_0^2$. Then, we have

$$t_0=\Omega\left(\frac{\log(\frac{1}{\Delta})}{\log(1+\frac{V}{\beta})+\log\log(\frac{1}{\Delta})}\right)$$

Starting from $\mathcal{I}(\mathcal{X},\theta_0)$, we repeatedly apply Eq. (25) for t_0 times to switch to $\mathcal{I}(\mathcal{X},\theta^*)$ where the reward gap is Δ. Since $\Pr_{\mathcal{I}(\mathcal{X},\theta_0)}\left[\mathcal{E}\left(0,\frac{\rho(\mathcal{X},\theta_0)}{Vq}\right)\right]=1$, by induction, we have

$$\begin{aligned}
&\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}\left[\mathcal{E}\left(t_0,\frac{\rho(\mathcal{X},\theta^*)}{Vq}\right)\right]\\
\geq&1-\left(3\delta+\sqrt{\frac{(1+c)^2 d(\mathcal{X})}{q}}+\left(\exp\left((1+c)^2 d(\mathcal{X}) n\sqrt{\frac{\log(nV)}{\beta}}\right)-1\right)+\frac{1}{nV}\right)\cdot t_0
\end{aligned}$$

When $\delta=O(\frac{1}{nV})$ and q,V,β are large enough, we can have

$$\left(3\delta+\sqrt{\frac{(1+c)^2 d(\mathcal{X})}{q}}+\left(\exp\left((1+c)^2 d(\mathcal{X}) n\sqrt{\frac{\log(nV)}{\beta}}\right)-1\right)+\frac{1}{nV}\right)\cdot t_0\leq\frac{1}{2}$$

and then

$$\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}\left[\mathcal{E}\left(t_0,\frac{\rho(\mathcal{X},\theta^*)}{Vq}\right)\right]\geq\frac{1}{2}.$$

Thus,

$$\begin{aligned}
&\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}\left[\mathcal{A}\text{ uses }\Omega\left(\frac{\log(\frac{1}{\Delta})}{\log(1+\frac{V}{\beta})+\log\log(\frac{1}{\Delta})}\right)\text{communication rounds}\right]\\
\geq&\Pr_{\mathcal{I}(\mathcal{X},\theta^*)}\left[\mathcal{E}\left(t_0,\frac{\rho(\mathcal{X},\theta^*)}{Vq}\right)\right]\\
\geq&\frac{1}{2},
\end{aligned}$$

which completes the proof of Theorem 3. □

B PROOFS FOR THE FIXED-BUDGET SETTING

B.1 Proof of Theorem 4

Proof of Theorem 4. Our proof of Theorem 4 adapts the error probability analysis in [23] to the multi-agent setting.

Since the number of samples used over all agents in each phase is $N=\lfloor TV/R\rfloor$, the total number of samples used by algorithm CoopKernelFB is at most TV and the total number of samples used per agent is at most T.

Now we prove the error probability upper bound.

Recall that for any $\lambda\in\triangle_{\tilde{\mathcal{X}}}$, $\Phi_\lambda=[\sqrt{\lambda_1}\phi(x_v^*)^\top;\dots;\sqrt{\lambda_{nV}}\phi(\tilde{x}_{nV})^\top]$ and $\Phi_\lambda^\top\Phi_\lambda=\sum_{\tilde{x}\in\tilde{\mathcal{X}}}\lambda_{\tilde{x}}\phi(\tilde{x})\phi(\tilde{x})^\top$. Let $\gamma_*=\xi_*N$.

For any $\tilde{x}_i, \tilde{x}_j \in \mathcal{B}_v^{(t)}, v \in [V], t \in [R]$, define

$$\Delta_{t,\tilde{x}_i,\tilde{x}_j} = \inf_{\Delta>0} \left\{ \frac{\|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{(\xi_* I + \Phi^\top_{\lambda^*_t} \Phi_{\lambda^*_t})^{-1}}}{\Delta^2} \leq 2\rho^* \right\}$$

and event

$$\mathcal{J}_{t,\tilde{x}_i,\tilde{x}_j} = \left\{ \left| \left(\hat{f}_t(\tilde{x}_i) - \hat{f}_t(\tilde{x}_j) \right) - \left(f(\tilde{x}_i) - f(\tilde{x}_j) \right) \right| < \Delta_{t,\tilde{x}_i,\tilde{x}_j} \right\}.$$

In the following, we prove $\Pr\left[\neg \mathcal{J}_{t,\tilde{x}_i,\tilde{x}_j}\right] \leq 2\exp\left(-\frac{N}{2(1+\varepsilon)\rho^*}\right)$.

Similar to the analysis procedure of Lemma 1, we have

$$\begin{aligned}
&\left(\hat{f}_t(\tilde{x}_i) - \hat{f}_t(\tilde{x}_j) \right) - \left(f(\tilde{x}_i) - f(\tilde{x}_j) \right) \\
=& \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\gamma_* I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \bar{\eta}^{(t)} - \gamma_* \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\gamma_* I + \Phi_t^\top \Phi_t\right)^{-1} \theta^*,
\end{aligned}$$

where the mean of the first term is zero and its variance is bounded by

$$\|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|^2_{(\gamma_* I + \Phi_t^\top \Phi_t)^{-1}} \leq \frac{(1+\varepsilon) \cdot \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_* I + \Phi^\top_{\lambda^*_t} \Phi_{\lambda^*_t})^{-1}}}{N}.$$

Using the Hoeffding inequality, we have

$$\begin{aligned}
&\Pr\left[\left| \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\gamma_* I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \eta_v^{(t)} \right| \geq \frac{1}{2} \Delta_{t,\tilde{x}_i,\tilde{x}_j} \right] \\
\leq& 2\exp\left(-2 \frac{\frac{1}{4}\Delta^2_{t,\tilde{x}_i,\tilde{x}_j}}{\frac{(1+\varepsilon)\cdot \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_* I + \Phi^\top_{\lambda^*_t} \Phi_{\lambda^*_t})^{-1}}}{N}} \right) \\
\leq& 2\exp\left(-\frac{1}{2} \frac{N}{\frac{(1+\varepsilon)\cdot \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_* I + \Phi^\top_{\lambda^*_t} \Phi_{\lambda^*_t})^{-1}}}{\Delta^2_{t,\tilde{x}_i,\tilde{x}_j}}} \right) \\
\leq& 2\exp\left(-\frac{N}{2(1+\varepsilon)\rho^*} \right)
\end{aligned}$$

Thus, with probability at least $1 - 2\exp\left(-\frac{N}{2(1+\varepsilon)\rho^*}\right)$, we have

$$\left| \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\gamma_* I + \Phi_t^\top \Phi_t\right)^{-1} \Phi_t^\top \eta_v^{(t)} \right| < \frac{1}{2} \Delta_{t,\tilde{x}_i,\tilde{x}_j}.$$

Recall that ξ_* satisfies $(1+\varepsilon)\sqrt{\xi_*} \max_{\tilde{x}_i,\tilde{x}_j \in \tilde{\mathcal{X}}_v, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{\left(\xi_* I + \Phi^\top_{\lambda_u} \Phi_{\lambda_u}\right)^{-1}} B \leq \frac{1}{2}\Delta_{t,\tilde{x}_i,\tilde{x}_j}$. Then, we bound the regularization (bias) term.

$$\begin{aligned}
&\left| \gamma_* \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\right)^\top \left(\gamma_* I + \Phi_t^\top \Phi_t\right)^{-1} \theta^* \right| \\
\leq& \gamma_* \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{\left(\gamma_* I + \Phi_t^\top \Phi_t\right)^{-1}} \|\theta^*\|_{\left(\gamma_* I + \Phi_t^\top \Phi_t\right)^{-1}} \\
\leq& \sqrt{\gamma_*} \cdot \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{\left(\gamma_* I + \Phi_t^\top \Phi_t\right)^{-1}} \|\theta^*\|_2
\end{aligned}$$

$$
\begin{aligned}
&\le \sqrt{\xi_* N} \cdot \frac{(1+\varepsilon) \cdot \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_* I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}}}{\sqrt{N}} \cdot \|\theta^*\|_2 \\
&\le (1+\varepsilon)\sqrt{\xi_*} \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{B}}_v^{(t)}, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{\left(\xi_* I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*}\right)^{-1}} \|\theta^*\|_2 \\
&\le (1+\varepsilon)\sqrt{\xi_*} \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{B}}_v^{(t)}, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{\left(\xi_* I + \Phi_{\lambda_u}^\top \Phi_{\lambda_u}\right)^{-1}} \|\theta^*\|_2 \\
&\le (1+\varepsilon)\sqrt{\xi_*} \max_{\tilde{x}_i, \tilde{x}_j \in \tilde{\mathcal{X}}_v, v \in [V]} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{\left(\xi_* I + \Phi_{\lambda_u}^\top \Phi_{\lambda_u}\right)^{-1}} \cdot B \\
&\le \frac{1}{2} \Delta_{t, \tilde{x}_i, \tilde{x}_j}
\end{aligned}
$$

Thus, with probability at least $1 - 2\exp\left(-\frac{N}{2(1+\varepsilon)\rho^*}\right)$, we have

$$
\begin{aligned}
&\left| \left(\hat{f}_t(\tilde{x}_i) - \hat{f}_t(\tilde{x}_j) \right) - \left(f(\tilde{x}_i) - f(\tilde{x}_j) \right) \right| \\
\le &\left| \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j) \right)^\top \left(\gamma_* I + \Phi_t^\top \Phi_t \right)^{-1} \Phi_t^\top \eta_v^{(t)} \right| + \left| \gamma_* \left(\phi(\tilde{x}_i) - \phi(\tilde{x}_j) \right)^\top \left(\gamma_* I + \Phi_t^\top \Phi_t \right)^{-1} \theta^* \right| \\
< &\Delta_{t, \tilde{x}_i, \tilde{x}_j},
\end{aligned}
$$

which completes the proof of $\Pr\left[\neg \mathcal{J}_{t, \tilde{x}_i, \tilde{x}_j}\right] \le 2\exp\left(-\frac{N}{2(1+\varepsilon)\rho^*}\right)$.

Define event

$$
\mathcal{J} = \left\{ \left| \left(\hat{f}_t(\tilde{x}_i) - \hat{f}_t(\tilde{x}_j) \right) - \left(f(\tilde{x}_i) - f(\tilde{x}_j) \right) \right| < \Delta_{t, \tilde{x}_i, \tilde{x}_j}, \ \forall \tilde{x}_i, \tilde{x}_j \in \mathcal{B}_v^{(t)}, \forall v \in [V], \forall t \in [R] \right\},
$$

By a union bound over $\tilde{x}_i, \tilde{x}_j \in \mathcal{B}_v^{(t)}$, $v \in [V]$ and $t \in [R]$, we have

$$
\begin{aligned}
\Pr\left[\neg \mathcal{J}\right] \le &2n^2 V \log(\omega(\tilde{\mathcal{X}})) \cdot \exp\left(-\frac{N}{2(1+\varepsilon)\rho^*} \right) \\
= &O\left(n^2 V \log(\omega(\tilde{\mathcal{X}})) \cdot \exp\left(-\frac{TV}{\rho^* \log(\omega(\tilde{\mathcal{X}}))} \right) \right)
\end{aligned}
$$

In order to prove Theorem 4, it suffices to prove that conditioning on $\mathcal{J}$, algorithm CoopKernelFB returns the correct answers $\tilde{x}_v^*$ for all $v \in [V]$.

Suppose that there exist $v \in [V]$ and $t \in [R]$ such that $\tilde{x}_v^*$ is eliminated in phase t. Define

$$
\mathcal{B'}_v^{(t)} = \{ \tilde{x} \in \mathcal{B}_v^{(t)} : \hat{f}_t(\tilde{x}) > \hat{f}_t(\tilde{x}_v^*) \},
$$

which denotes the subset of arms that are ranked before $\tilde{x}_v^*$ by the estimated rewards in $\mathcal{B}_v^{(t)}$. According to the elimination rule, we have

$$
\omega(\mathcal{B'}_v^{(t)} \cup \{\tilde{x}_v^*\}) > \frac{1}{2}\omega(\mathcal{B}_v^{(t)}) = \frac{1}{2} \max_{\tilde{x}_i, \tilde{x}_j \in \mathcal{B}_v^{(t)}} \|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_* I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}} \tag{26}
$$

Define $\tilde{x}_0 = \operatorname{argmax}_{\tilde{x} \in \mathcal{B'}_v^{(t)}} \Delta_{\tilde{x}}$. We have

$$
\frac{1}{2} \frac{\|\phi(\tilde{x}_v^*) - \phi(\tilde{x}_0)\|_{(\xi_* I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}}}{\Delta_{\tilde{x}_0}^2}
$$

$$
\begin{aligned}
&\leq \frac{1}{2} \max_{\tilde{x}_i, \tilde{x}_j \in \mathcal{B}_v^{(t)}} \frac{\|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{(\xi_* I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}}}{\Delta_{\tilde{x}_0}^2} \\
&\overset{\text{(f)}}{<} \frac{\omega(\mathcal{B}'^{(t)}_v \cup \{\tilde{x}_v^*\})}{\Delta_{\tilde{x}_0}^2} \\
&\overset{\text{(g)}}{=} \min_{\lambda \in \triangle \mathcal{X}} \max_{\tilde{x}_i, \tilde{x}_j \in \mathcal{B}'^{(t)}_v \cup \{\tilde{x}_v^*\}} \frac{\|\phi(\tilde{x}_i) - \phi(\tilde{x}_j)\|_{\left(\xi_* I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}}}{\Delta_{\tilde{x}_0}^2} \\
&\overset{\text{(h)}}{\leq} \min_{\lambda \in \triangle \mathcal{X}} \max_{\tilde{x} \in \mathcal{B}'^{(t)}_v} \frac{\|\phi(\tilde{x}_v^*) - \phi(\tilde{x})\|_{\left(\xi_* I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}}}{\Delta_{\tilde{x}}^2} \\
&\leq \min_{\lambda \in \triangle \mathcal{X}} \max_{\tilde{x} \in \mathcal{X}_v \setminus \{\tilde{x}_v^*\}, v \in [V]} \frac{\|\phi(\tilde{x}_v^*) - \phi(\tilde{x})\|_{\left(\xi_* I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}}}{\Delta_{\tilde{x}}^2} \\
&= \rho^*,
\end{aligned}
$$

where (f) can be obtained by dividing Eq. (26) by $\Delta_{\tilde{x}_0}^2$, (g) comes from the definition of $\omega(\cdot)$, and (h) is due to the definition of $\tilde{x}_0$.

According to the definition

$$
\Delta_{t,\tilde{x}_v^*,\tilde{x}_0} = \inf_{\Delta > 0} \left\{ \frac{\|\phi(\tilde{x}_v^*) - \phi(\tilde{x}_0)\|^2_{(\xi_* I + \Phi_{\lambda_t^*}^\top \Phi_{\lambda_t^*})^{-1}}}{\Delta^2} \leq 2\rho^* \right\},
$$

we have $\Delta_{t,\tilde{x}_v^*,\tilde{x}_0} \leq \Delta_{\tilde{x}_0}$.

Conditioning on $\mathcal{J}$, we have $\left| \left(\hat{f}_t(\tilde{x}_v^*) - \hat{f}_t(\tilde{x}_0) \right) - \left(f(\tilde{x}_v^*) - f(\tilde{x}_0) \right) \right| < \Delta_{t,\tilde{x}_v^*,\tilde{x}_0} \leq \Delta_{\tilde{x}_0}$. Then, we have

$$
\hat{f}_t(\tilde{x}_v^*) - \hat{f}_t(\tilde{x}_0) > \left(f(\tilde{x}_v^*) - f(\tilde{x}_0) \right) - \Delta_{\tilde{x}_0} = 0,
$$

which contradicts the definition of $\tilde{x}_0$ that satisfies $\hat{f}_t(\tilde{x}_0) > \hat{f}_t(\tilde{x}_v^*)$. Thus, for any $v \in [V]$, $\tilde{x}_v^*$ will never be eliminated.

Since $\omega(\{\tilde{x}_v^*, \tilde{x}\}) \geq 1$ for any $\tilde{x} \in \mathcal{X}_v \setminus \{\tilde{x}_v^*\}, v \in [V]$ and $R = \lceil \log_2(\omega(\tilde{\mathcal{X}})) \rceil \geq \lceil \log_2(\omega(\mathcal{X}_v)) \rceil$ for any $v \in [V]$, we have that conditioning on $\mathcal{J}$, algorithm CoopKernelFB returns the correct answers $\tilde{x}_v^*$ for all $v \in [V]$.

For communication rounds, since algorithm CoopKernelFB has at most $R = \lceil \log_2(\omega(\tilde{\mathcal{X}})) \rceil$ phases, the number of communication rounds is bounded by $O(\log(\omega(\tilde{\mathcal{X}})))$. □

B.2 Proof of Corollary 2

Proof of Corollary 2. Following the analysis procedure of Corollary 1, we have

$$
\begin{aligned}
\rho^* &= \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{\tilde{x} \in \tilde{\mathcal{X}}_v, v \in [V]} \frac{\|\phi(\tilde{x}_v^*) - \phi(\tilde{x})\|^2_{\left(\xi_* I + \Phi_\lambda^\top \Phi_\lambda\right)^{-1}}}{(f(\tilde{x}_v^*) - f(\tilde{x}))^2} \\
&\leq \frac{8}{\Delta_{\min}^2} \cdot \log\det\left(I + \xi_*^{-1} K_{\lambda^*} \right)
\end{aligned}
$$

Maximum Information Gain. Recall that the maximum information gain over all sample allocation $\lambda \in \triangle_{\tilde{\mathcal{X}}}$ is defined as

$$\Upsilon = \max_{\lambda \in \triangle_{\tilde{\mathcal{X}}}} \log\det\left(I + {\xi_*}^{-1} K_\lambda\right).$$

Then, the error probability is bounded by

$$\begin{aligned} & O\left(n^2 V \log(\omega(\tilde{\mathcal{X}})) \cdot \exp\left(-\frac{TV}{\rho^* \log(\omega(\tilde{\mathcal{X}}))}\right)\right) \\ =& O\left(n^2 V \log(\omega(\tilde{\mathcal{X}})) \cdot \exp\left(-\frac{TV\Delta_{\min}^2}{\Upsilon \log(\omega(\tilde{\mathcal{X}}))}\right)\right) \end{aligned}$$

Effective Dimension. Recall that $\alpha_1 \geq \cdots \geq \alpha_{nV}$ denote the eigenvalues of K_{λ^*} in decreasing order. The effective dimension of K_{λ^*} is defined as

$$d_{\text{eff}} = \min\left\{ j : j\xi_* \log(nV) \geq \sum_{i=j+1}^{nV} \alpha_i \right\},$$

and we have

$$\log\det\left(I + \xi_*^{-1} K_{\lambda^*}\right) \leq d_{\text{eff}} \log\left(2nV \cdot \left(1 + \frac{\text{Trace}(K_{\lambda^*})}{\xi_* d_{\text{eff}}}\right)\right).$$

Then, the error probability is bounded by

$$\begin{aligned} & O\left(n^2 V \log(\omega(\tilde{\mathcal{X}})) \cdot \exp\left(-\frac{TV}{\rho^* \log(\omega(\tilde{\mathcal{X}}))}\right)\right) \\ =& O\left(n^2 V \log(\omega(\tilde{\mathcal{X}})) \cdot \exp\left(-\frac{TV\Delta_{\min}^2}{d_{\text{eff}} \log\left(nV \cdot \left(1 + \frac{\text{Trace}(K_{\lambda^*})}{\xi_* d_{\text{eff}}}\right)\right) \log(\omega(\tilde{\mathcal{X}}))}\right)\right) \end{aligned}$$

Decomposition. Recall that $K = [K(\tilde{x}_i, \tilde{x}_j)]_{i,j \in [nV]}$, $K_z = [K_z(z_v, z_{v'})]_{v,v' \in [V]}$, $K_x = [K_x(x_i, x_j)]_{i,j \in [nV]}$, and $\text{rank}(K_{\lambda^*}) = \text{rank}(K) \leq \text{rank}(K_z) \cdot \text{rank}(K_x)$. We have

$$\log\det\left(I + \xi_*^{-1} K_{\lambda^*}\right) \leq \text{rank}(K_z) \cdot \text{rank}(K_x) \log\left(\frac{\text{Trace}\left(I + \xi_*^{-1} K_{\lambda^*}\right)}{\text{rank}(K_{\lambda^*})}\right)$$

Then, the error probability is bounded by

$$\begin{aligned} & O\left(n^2 V \log(\omega(\tilde{\mathcal{X}})) \cdot \exp\left(-\frac{TV}{\rho^* \log(\omega(\tilde{\mathcal{X}}))}\right)\right) \\ =& O\left(n^2 V \log(\omega(\tilde{\mathcal{X}})) \cdot \exp\left(-\frac{TV\Delta_{\min}^2}{\text{rank}(K_z) \cdot \text{rank}(K_x) \log\left(\frac{\text{Trace}\left(I + \xi_*^{-1} K_{\lambda^*}\right)}{\text{rank}(K_{\lambda^*})}\right) \log(\omega(\tilde{\mathcal{X}}))}\right)\right) \end{aligned}$$

Therefore, we complete the proof of Corollary 2. □

B.3 Proof of Theorem 5

Our proof of Theorem 5 follows the analysis procedure in [1].

We first introduce some definitions in information theory. For any random variable A, let $H(A)$ denote the Shannon entropy of A. If A is uniformly distributed on its support, $H(A) = \log|A|$. For any $p \in (0,1)$, let $H_2(p) = -p\log_2 p - (1-p)\log_2(1-p)$ denote the binary entropy, and $H_2(p) = H(A)$ for $A \sim \text{Bernoulli}(p)$. For any random variables A, B, let $H(A;B) = H(A) - H(A|B) = H(B) - H(B|A)$ denote the mutual information of A and B.

Consider the following fully-collaborative instance $\mathcal{D}_d^{\Delta,p}$: for a uniformly drawn index $i^* \in [d]$, $\theta^*_{i^*} = p + \Delta$ and $\theta^*_j = p$ for all $j \neq i^*$. The arm set $\mathcal{X} = \{x \in \{0,1\}^d : \sum_{i=1}^d x(i) = 1\}$, and the feature mapping $\phi(x) = Ix$ for all $x \in \mathcal{X}$. Under instance $\mathcal{D}_d^{\Delta,p}$, we have $\omega(\tilde{\mathcal{X}}) = \omega(\mathcal{X}) = d$.

Let $\delta_* > 0$ be a small constant that we specify later. There exists a single-agent algorithm $\mathcal{A}_S$ (e.g., [23]) that uses at most T samples and guarantees $O(\log d \cdot \exp(-\frac{\Delta^2 T}{d \log d}))$ error probability on instance $\mathcal{D}_d^{\Delta,p}$ for any $d > 1$. Restricting the error probability to the constant δ_*, we have that for any $d > 1$, $\mathcal{A}_S$ uses at most $T = O(\frac{d}{\Delta^2}\log d \cdot \log\log d)$ samples to guarantee δ_* error probability on instance $\mathcal{D}_d^{\Delta,p}$.

Let $\alpha = \frac{V}{\beta}$ and $1 \leq \alpha \leq \log d$. According to the definition of speedup, a $\frac{V}{\alpha}$-speedup distributed algorithm $\mathcal{A}$ must satisfy that for any $d > 1$, $\mathcal{A}$ uses at most $T = O(\frac{\frac{n}{\Delta^2}}{\frac{V}{\alpha}} \cdot V \log d \cdot \log\log d) = O(\frac{\alpha d}{\Delta^2}\log d \cdot \log\log d)$ samples over all agents and guarantees δ_* error probability on instance $\mathcal{D}_d^{\Delta,p}$.

Main Proof. Now, in order to prove Theorem 5, we first prove the following Lemma 6, which relaxes the sample budget within logarithmic factors.

Lemma 6. *For any $d > 1$, any distributed algorithm $\mathcal{A}$ that can use at most $O(\frac{\alpha d(\log\alpha + \log\log d)^2}{\Delta^2(\log d)^2})$ samples over all agents and guarantee δ_* error probability on instance $\mathcal{D}_d^{\Delta,p}$ needs $\Omega(\frac{\log d}{\log\alpha + \log\log d})$ communication rounds.*

Below we prove Lemma 6 by induction (Lemmas 7,8).

Lemma 7 (Basic Step). *For any $d > 1$ and $1 \leq \alpha \leq \log d$, there is no 1-round algorithm $\mathcal{A}_1$ that can use $O(\frac{\alpha d}{\Delta^2})$ samples and guarantee δ_1 error probability for some constant $\delta_1 \in (0,1)$ on instance $\mathcal{D}_d^{\Delta,p}$.*

Proof of Lemma 7. Let I denote the random variable of index i^*. Since I is uniformly distributed on $[d]$, $H(I) = \log d$. Let S denote the sample profile of $\mathcal{A}_1$ on instance $\mathcal{D}_d^{\Delta,p}$. According to the definitions of $\mathcal{X}$ and θ^*, for instance $\mathcal{D}_d^{\Delta,p}$, identifying the best arm is equivalent to identifying I. Suppose that $\mathcal{A}_1$ returns the best arm with error probability δ_1. Using the Fano's inequality (Lemma 13), we have

$$H(I|S) \leq H_2(\delta_1) + \delta_1 \log d \tag{27}$$

Using Lemma 14, we have

$$\begin{aligned} H(I|S) =& H(I) - H(I;S) \\ =& \log d - O(\frac{\alpha d}{\Delta^2} \cdot \frac{\Delta^2}{d}) \\ =& \log d - O(\alpha) \end{aligned}$$

Then, for some small enough constant $\delta_1 \in (0,1)$, Eq. (27) cannot hold. Thus, for any $d > 1$ and $1 \leq \alpha \leq \log d$, there is no 1-round algorithm $\mathcal{A}_1$ that can use $O(\frac{\alpha d}{\Delta^2})$ samples and guarantee δ_1 error probability on instance $\mathcal{D}_d^{\Delta,p}$. □

Lemma 8 (Induction Step). *Suppose that $1 \leq \alpha \leq \log d$ and $\delta \in (0,1)$. If for any $d > 1$, there is no $(r-1)$-round algorithm $\mathcal{A}_{r-1}$ that can use $O(\frac{\alpha d}{\Delta^2 (r-1)^2})$ samples and guarantee δ error probability on instance $\mathcal{D}_d^{\Delta,p}$, then for any $d > 1$,* *there is no r-round algorithm $\mathcal{A}_r$ that can use $O(\frac{\alpha d}{\Delta^2 r^2})$ samples and guarantee $\delta - O(\frac{1}{r^2})$ error probability on instance $\mathcal{D}_d^{\Delta,p}$.*

Proof of Lemma 8. We prove this lemma by contradiction. Suppose that for some $d > 1$, there exists an r-round algorithm $\mathcal{A}_r$ that can use $O(\frac{\alpha d}{\Delta^2 r^2})$ samples and guarantee δ error probability on instance $\mathcal{D}_d^{\Delta,p}$. In the following, we show how to construct an $(r-1)$-round algorithm $\mathcal{A}_{r-1}$ that can use $O(\frac{\alpha d}{\Delta^2 (r-1)^2})$ samples and guarantee at most $\delta + O(\frac{1}{r^2})$ error probability on instance $\mathcal{D}_{\tilde{d}}^{\Delta,p}$ for some $\tilde{d}$.

Let S_1 denote the sample profile of $\mathcal{A}_r$ in the first round on instance $\mathcal{D}_d^{\Delta,p}$. Since the number of samples of $\mathcal{A}_r$ is bounded by $O(\frac{\alpha d}{\Delta^2 r^2})$, using Lemma 14, we have

$$\begin{aligned} H(I|S_1) =& H(I) - H(I;S_1) \\ =& \log d - O(\frac{\alpha d}{\Delta^2 r^2} \cdot \frac{\Delta^2}{d}) \\ =& \log d - O(\frac{\alpha}{r^2}) \end{aligned}$$

Let $\mathcal{D}_{d|S_1}^{\Delta,p}$ denote the posterior of $\mathcal{D}_d^{\Delta,p}$ after observing the sample profile S_1.

Using Lemma 15 on random variable $I|S_1$ with parameters $\gamma = O(\frac{\alpha}{r^2})$ and $\epsilon = o(\frac{1}{r^2})$, we can write $\mathcal{D}_{d|S_1}^{\Delta,p}$ as a convex combination of distributions $Q_0, Q_1, \ldots, Q_\ell$, i.e., $\mathcal{D}_{d|S_1}^{\Delta,p} = \sum_{j=0}^{\ell} q_j Q_j$ such that $q_0 = o(\frac{1}{r^2})$, and for any $j \geq 1$,

$$\begin{aligned} |\mathrm{supp}(Q_j)| \geq& \exp\left(\log d - \frac{\gamma}{\epsilon}\right) \\ =& \exp\left(\log d - o(\alpha)\right) \\ =& \frac{d}{e^{o(\alpha)}}, \end{aligned}$$

and

$$\|Q_j - \mathcal{U}_j\|_{\mathrm{TV}} = o(\frac{1}{r^2}),$$

where $\mathcal{U}_j$ is the uniform distribution on $\mathrm{supp}(Q_j)$.

Since

$$\Pr[\mathcal{A}_r \text{ has an error}|\mathcal{D}_{d|S_1}^{\Delta,p}] = \sum_{j=0}^{\ell} q_j \Pr[\mathcal{A}_r \text{ has an error}|Q_j] \leq \delta,$$

using an average argument, there exists a distribution Q_j for some $j \geq 1$ such that

$$\Pr[\mathcal{A}_r \text{ has an error}|Q_j] \leq \delta.$$

Let $\tilde{d} = |\mathrm{supp}(Q_j)| \geq \frac{d}{e^{o(\alpha)}}$. Since $\|Q_j - \mathcal{U}_j\|_{\mathrm{TV}} = o(\frac{1}{r^2})$ and $\mathcal{U}_j$ is equivalent to $\mathcal{D}_{\tilde{d}}^{\Delta,p}$, we have

$$\Pr[\mathcal{A}_r \text{ has an error}|\mathcal{D}_{\tilde{d}}^{\Delta,p}] \leq \delta + o(\frac{1}{r^2}).$$

Under instance $\mathcal{D}_{\tilde{d}}^{\Delta,p}$, the sample budget for $(r-1)$-round algorithms is $O(\frac{\alpha \tilde{d}}{\Delta^2 (r-1)^2})$, and it holds that

$$O(\frac{\alpha \tilde{d}}{\Delta^2 (r-1)^2}) \geq O(\frac{\alpha \frac{d}{e^{o(\alpha)}}}{\Delta^2 (r-1)^2}) \geq O(\frac{\alpha d^{1-o(1)}}{\Delta^2 (r-1)^2}) \geq O(\frac{\alpha d}{\Delta^2 r^2})$$

Then, we can construct an $(r-1)$-round algorithm $\mathcal{A}_{r-1}$ using algorithm $\mathcal{A}_r$ from the second round. The constructed $\mathcal{A}_{r-1}$ uses at most $O(\frac{\alpha d}{\Delta^2 r^2}) \leq O(\frac{\alpha \tilde{d}}{\Delta^2 (r-1)^2})$ samples and guarantees $\delta + o(\frac{1}{r^2})$ error probability.

The specific procedure of $\mathcal{A}_{r-1}$ is as follows: if $\mathcal{A}_r$ samples some dimension $i \in \text{supp}(Q_j)$, then $\mathcal{A}_{r-1}$ also samples i; otherwise, if $\mathcal{A}_r$ samples some dimension $i \in [d] \setminus \text{supp}(Q_j)$, then $\mathcal{A}_{r-1}$ samples Bernoulli(p) and feeds the outcome to $\mathcal{A}_r$. Finally, if $\mathcal{A}_r$ returns some dimension $i \in \text{supp}(Q_j)$, then $\mathcal{A}_{r-1}$ also returns i; otherwise, if $\mathcal{A}_r$ returns some dimension $i \in [d] \setminus \text{supp}(Q_j)$, then $\mathcal{A}_{r-1}$ returns an arbitrary dimension in $[d] \setminus \text{supp}(Q_j)$ (the error case). □

Proof of Lemma 6. Let $r_* = \frac{\log d}{\log \alpha + \log\log d}$. Combining Lemmas 7,8, we obtain that there is no r_*-round algorithm $\mathcal{A}$ that can use $O(\frac{\alpha d}{\Delta^2 r_*^2})$ samples and guarantee $\delta_1 - \sum_{r=2}^{r_*} o(\frac{1}{r^2})$ error probability on instance $\mathcal{D}_d^{\Delta,p}$ for any $d > 1$.

Let $\delta_* < \delta_1 - \sum_{r=2}^{r_*} o(\frac{1}{r^2})$. Thus, for any $d > 1$ and $1 \leq \alpha \leq \log d$, any distributed algorithm $\mathcal{A}$ that can use $O(\frac{\alpha d}{\Delta^2 r_*^2})$ samples and guarantee δ_* error probability on instance $\mathcal{D}_d^{\Delta,p}$ must cost $r_* = \Omega\left(\frac{\log d}{\log(\frac{V}{\beta}) + \log\log d}\right)$ communication rounds. □

Therefore, for any $d > 1$ and $\frac{V}{\log d} \leq \beta \leq V$, a β-speedup distributed algorithm $\mathcal{A}$ needs at least $\Omega\left(\frac{\log \omega(\tilde{\mathcal{X}})}{\log(\frac{V}{\beta}) + \log\log \omega(\tilde{\mathcal{X}})}\right)$ communication rounds under instance $\mathcal{D}_d^{\Delta,p}$, which completes the proof of Theorem 5.

C TECHNICAL TOOLS

Lemma 9 (Lemma 15 in [10]). *For $\lambda^* = \text{argmax}_{\lambda \in \triangle_{\tilde{\mathcal{X}}}} \log\det\left(I + {\xi_*}^{-1} \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda_{\tilde{x}'} \phi(\tilde{x}')\phi(\tilde{x}')^\top\right)$, we have*

$$\max_{\tilde{x} \in \tilde{\mathcal{X}}} \|\phi(\tilde{x})\|^2_{\left(\xi_* I + \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}'} \phi(\tilde{x}')\phi(\tilde{x}')^\top\right)^{-1}} = \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}} \|\phi(\tilde{x})\|^2_{\left(\xi_* I + \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}'} \phi(\tilde{x}')\phi(\tilde{x}')^\top\right)^{-1}}$$

Lemma 10. *For $\lambda^* = \text{argmax}_{\lambda \in \triangle_{\tilde{\mathcal{X}}}} \log\det\left(I + {\xi_*}^{-1} \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda_{\tilde{x}'} \phi(\tilde{x}')\phi(\tilde{x}')^\top\right)$, we have*

$$\sum_{\tilde{x} \in \tilde{\mathcal{X}}} \log\left(1 + \lambda^*_{\tilde{x}} \|\phi(\tilde{x})\|^2_{\left(\xi_* I + \sum_{\tilde{x}' \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}'} \phi(\tilde{x}')\phi(\tilde{x}')^\top\right)^{-1}}\right) \leq \log \frac{\det\left(\xi_* I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}} \phi(\tilde{x})\phi(\tilde{x})^\top\right)}{\det(\xi_* I)}$$

Proof of Lemma 10. For any $j \in [nV]$, let $M_j = \det\left(\xi_* I + \sum_{i \in [j]} \lambda^*_i \phi(\tilde{x}_i)\phi(\tilde{x}_i)^\top\right)$.

$$\begin{aligned}
&\det\left(\xi_* I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}} \phi(\tilde{x})\phi(\tilde{x})^\top\right) \\
=& \det\left(\xi_* I + \sum_{i \in [nV-1]} \lambda^*_i \phi(\tilde{x}_i)\phi(\tilde{x}_i)^\top + \lambda^*_{nV} \phi(\tilde{x}_{nV})\phi(\tilde{x}_{nV})^\top\right) \\
=& \det(M_{nV-1}) \det\left(I + \lambda^*_{nV} \cdot M_{nV-1}^{-\frac{1}{2}} \phi(\tilde{x}_{nV}) \left(M_{nV-1}^{-\frac{1}{2}} \phi(\tilde{x}_{nV})\right)^\top\right) \\
=& \det(M_{nV-1}) \det\left(I + \lambda^*_{nV} \cdot \phi(\tilde{x}_{nV})^\top M_{nV-1}^{-1} \phi(\tilde{x}_{nV})\right) \\
=& \det(M_{nV-1}) \left(1 + \lambda^*_{nV} \|\phi(\tilde{x}_{nV})\|^2_{M_{nV-1}^{-1}}\right) \\
=& \det(\xi_* I) \prod_{i=1}^{nV} \left(1 + \lambda^*_i \|\phi(\tilde{x}_i)\|^2_{M_{i-1}^{-1}}\right)
\end{aligned}$$

Thus,

$$\frac{\det\left(\xi_* I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}} \phi(\tilde{x}) \phi(\tilde{x})^\top\right)}{\det(\xi_* I)} = \prod_{i=1}^{nV} \left(1 + \lambda^*_i \|\phi(\tilde{x}_i)\|^2_{M^{-1}_{i-1}}\right)$$

Taking logarithm on both sides, we have

$$\begin{aligned} &\log \frac{\det\left(\xi_* I + \sum_{\tilde{x} \in \tilde{\mathcal{X}}} \lambda^*_{\tilde{x}} \phi(\tilde{x}) \phi(\tilde{x})^\top\right)}{\det(\xi_* I)} \\ &= \sum_{i=1}^{nV} \log \left(1 + \lambda^*_i \|\phi(\tilde{x}_i)\|^2_{M^{-1}_{i-1}}\right) \\ &\geq \sum_{i=1}^{nV} \log \left(1 + \lambda^*_i \|\phi(\tilde{x}_i)\|^2_{M^{-1}_{nV}}\right), \end{aligned}$$

which completes the proof of Lemma 10. □

Lemma 11 (Pinsker's inequality). *If P and Q are two probability distributions on a measurable space (X, Σ), then for any measurable event $A \in \Sigma$, it holds that*

$$|P(A) - Q(A)| \leq \sqrt{\frac{1}{2} \mathrm{KL}(P \| Q)}.$$

Lemma 12 (Lemma 29 in [38]). *For any $\gamma_1, \ldots, \gamma_K \in [0, 1]$ and $x \geq 0$, it holds that*

$$\prod_{i=1}^{K} \max\{1 - \gamma_i - \gamma_i x,\ 0\} \geq \prod_{i=1}^{K} (1 - \gamma_i) - x$$

Lemma 13 (Fano's Inequality). *Let A, B be random variables and f be a function that given A predicts a value for B. If $\Pr(f(A) \neq B) \leq \delta$, then $H(B|A) \leq H_2(\delta) + \delta \cdot \log |B|$.*

Lemma 14. *For the instance $\mathcal{D}^{\Delta,p}_d$ with sample profile $\mathcal{S}$, we have $H(I; S) = O(|S| \cdot \frac{\Delta^2}{d})$.*

The proof of Lemma 14 follows the same analysis procedure of Lemma 7 in [1].

Lemma 15 (Lemma 8 in [1]). *Let $A \sim \mathcal{D}$ be a random variable on $[d]$ with $H(A) \geq \log d - \gamma$ for some $\gamma \geq 1$. For any $\varepsilon > \exp(-\gamma)$, there exists $\ell + 1$ distributions $\psi_0, \psi_1, ..., \psi_\ell$ on $[d]$ along with $\ell + 1$ probabilities $p_0, p_1, \ldots, p_\ell$ ($\sum_{i=0}^{\ell} p_i = 1$) for some $\ell = O(\gamma/\varepsilon^3)$ such that $\mathcal{D} = \sum_{i=1}^{\ell} p_i \psi_i$, $p_0 = O(\varepsilon)$, and for any $i \geq 1$,*

1. *$\log |\mathrm{supp}(\psi_i)| \geq \log d - \gamma/\varepsilon$.*
2. *$\|\psi_i - \mathcal{U}_i\|_{\mathrm{TV}} = O(\varepsilon)$ where $\mathcal{U}_i$ denotes the uniform distribution on $\mathrm{supp}(\psi_i)$.*