Model-Based Best Arm Identification for Decreasing Bandits

Sho Takemori[1] **Yuhei Umeda**[1] **Aditya Gopalan**[2]
[1]Fujitsu Limited
Kawasaki, Japan
{takemori.sho, umeda.yuhei}@fujitsu.com
[2]Indian Institute of Science
Bengaluru 560012, India
aditya@iisc.ac.in

Abstract

We study the problem of reliably identifying the best (lowest loss) arm in a stochastic multi-armed bandit when the expected loss of each arm is monotone decreasing as a function of its pull count. This models, for instance, scenarios where each arm itself represents an optimization algorithm for finding the minimizer of a common function, and there is a limited time available to test the algorithms before committing to one of them. We assume that the decreasing expected loss of each arm depends on the number of its pulls as a (inverse) polynomial with unknown coefficients. We propose two fixed-budget best arm identification algorithms – one for the case of sparse polynomial decay models and the other for general polynomial models – along with bounds on the identification error probability. We also derive algorithm-independent lower bounds on the error probability. These bounds are seen to be factored into the product of the usual problem complexity and the model complexity that only depends on the parameters of the model. This indicates that our methods can identify the best arm even when the budget is smaller. We conduct empirical studies of our algorithms to complement our theoretical findings.

1 INTRODUCTION

The multi-armed bandit model for sequential decision-making (Thompson, 1933) has proven to be popular for learning over actions (arms) via iterative trial and error. Most of the standard treatment of stochastic bandits is based on the assumption that each arm's payoff is stationary, i.e., its mean loss (or reward) does not change with time or how often it is played, see e.g., (Lattimore and Szepesvári, 2020, Chap. 4). This assumption is quite restrictive in many scenarios where bandit algorithms can be used as selection strategies over a pool of 'base algorithms' to achieve a task, especially when each base algorithm is a *learning* agent. This process of learning confers an inherent degree of nonstationarity to the base algorithms, in that they 'improve' themselves over time. More formally, in this setting, we aim to optimize (minimize) a black-box function $F : \mathcal{X} \to \mathbb{R}$ with a limited number of function queries, where a set of learning algorithms $\mathcal{A}_1, \dots, \mathcal{A}_K$ (e.g., bandits, online convex optimization, or Bayesian optimization algorithms) for optimizing F is given. In this nonstationarity bandit problem, for each round $1 \le t \le T$, an agent selects an algorithm $\mathcal{A}_{i_t}$, the algorithm $\mathcal{A}_{i_t}$ selects $\xi_{i_t,\tau} \in \mathcal{X}$ and the agent observes the loss $F(\xi_{i_t,\tau})$, where $\tau = \tau(i_t, t)$ is the number of times the algorithm $\mathcal{A}_{i_t}$ has been selected up to time step t. If learning algorithms work properly, for each arm i, the expected loss decreases as τ increases.

Motivated by this setting, we consider the problem of best arm identification (BAI) in bandits with a decreasing loss profile for the arms. We term this problem setting *decreasing bandits*, which is equivalent to rising bandits (Li et al., 2020) in the reward formulation and a special case of rested bandits (Tekin and Liu, 2012). More precisely, we model the (expected) loss of an arm i after τ pulls as:

$$\mathbb{E}\left[y_{i,\tau}\right] = \sum_{m=1}^{d} \frac{\theta_m(i)}{\tau^{\rho_m}}, \tag{1}$$

where $\theta_m(i) \in \mathbb{R}$ is unknown to the learner, and the sequence of powers $\boldsymbol{\rho} = (\rho_m)_{1 \le m \le d} \in \mathbb{R}^d_{\ge 0}$ is known. We note that if $\theta_m(i) \ge 0$ for all m with $\rho_m > 0$, then $\mathbb{E}\left[y_{i,\tau}\right]$ is a non-increasing function of τ.

The decreasing bandits BAI problem with a fixed bud-

Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238.

get has been studied under several settings. Under a general assumption that for each arm i, the expected loss $\mathbb{E}[y_{i,\tau}]$ is non-increasing with respect to the number τ of arm pulls, the expected loss $\mathbb{E}[y_{i,\tau}]$ is "convex" with respect to τ, and the decay of the loss decrement is given as $\mathbb{E}[y_{i,\tau}] - \mathbb{E}[y_{i,\tau+1}] = O(\tau^{-1-\rho})$ with $\rho > 0$, Mussi et al. (2023) shows the following positive and negative results. (i) If the budget T is sufficiently large, then the probability of the error of identifying the best arm i^* is given as $\exp(-\Omega(T/H))$, where $i^* = \operatorname{argmin}_i \mathbb{E}[y_{i,T}]$, $H = \sum_{i \neq i^*} \Delta_i(T)^{-2}$, and $\Delta_i(T) = \mathbb{E}[y_{i,T}] - \mathbb{E}[y_{i^*,T}]$. (ii) However, unless the budget T satisfies the following inequality, then no algorithm can identify the best arm (Mussi et al., 2023):

$$T \geq H^{(\rho)}(T), \text{ where } H^{(\rho)}(T) = \sum_{i \neq i^*} \frac{1}{\Delta_i(T)^{1/\rho}}. \quad (2)$$

These results are natural, since if τ is sufficiently large, then the expected loss $\mathbb{E}[y_{i,\tau}]$ becomes approximately stationary and otherwise, the assumption is too general to identify the best arm. More precisely, if the rate ρ of the decay is not large (e.g., $\rho < 1/2$), then $H^{(\rho)}(T)$ is large compared to the problem complexity H and unlike the stationary setting (Audibert et al., 2010), even if the upper bound $\exp(-\Omega(T/H))$ is less than 1, there is no theoretical guarantee unless the budget is large enough. This negative result motivates study of the decreasing bandits BAI problem under a more specific assumption such as the loss model (1).

We remark that our structured model (1) is well-supported by empirical evidence, e.g., for the case of the loss profile of the TuRBO algorithm (Eriksson et al., 2019) (a Bayesian optimization algorithm) applied to minimization of the test function Levy(10) (Laguna and Marti, 2005). Figure 1 shows the loss trajectories of the learning algorithm over 100 independent experiments, along with the average loss. We also show the loss predicted by the model (1) with $d = 10$ fitted using the loss trajectories. The mean absolute error on the data is 0.09 (we detail how we select $(\rho_m)_m$ in Sec. 7).

Contributions. Under the structured loss model (1), we can prove more fine-grained and positive results. We provide two algorithms – one for the general loss model and one for a model with sparse coefficients $\theta_m(i)$, along with upper bounds on their probabilities of the error of the form $\exp\left(-\Omega\left(\frac{T}{\gamma H_2 \log_2 K}\right)\right)$ under mild assumptions on the budget T (Sections 4, 5). Here, H_2 is a measure of problem complexity, defined in Section 3, and γ is a constant independent of H_2 and T. In this paper, we term a constant that is independent of the problem complexity, and mainly depends on the parameter $\boldsymbol{\rho}$ of the loss model a *model complexity*. We then provide universal lower bounds on the probability of the error (Section 6), which shows that the probability error involves a model complexity other than the the problem complexity H. These results indicate that our methods can identity the best arm with a smaller budget compared to existing methods (Mussi et al., 2023). We emphasize that even though one can regard the loss model (1) as a linear model with a feature vector $(\tau^{-\rho_m})_{1 \leq m \leq d}$, these results are highly non-trivial since the feature vector is time dependent. Finally, we empirically verify our theoretical findings in synthetic environments (Section 7). All the omitted proofs can be found in Section E.

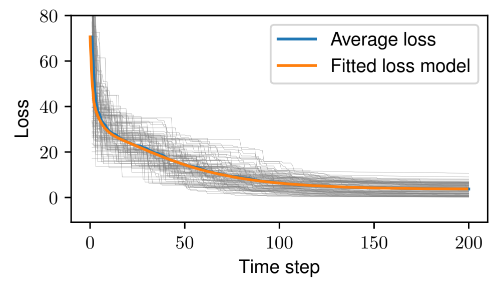

Figure 1: Loss trajectories of the TuRBO-1 optimization algorithm of Eriksson et al. (2019). Each independently-sampled trajectory plots the best achieved loss so far for the trust region vs. time steps. The orange curve shows a polynomial model fit to the ensemble of loss trajectories.

2 RELATED WORK

This paper studies a non-stationary BAI problem in the fixed budget setting. The BAI problem has been extensively studied especially for the stationary case (Audibert et al., 2010; Mannor and Tsitsiklis, 2004; Garivier and Kaufmann, 2016). Abbasi-Yadkori et al. (2018) proposed a BAI algorithm in the best of both worlds setting. While they define the best arm as the arm with the largest cumulative rewards, we define the best arm using a specific time step $\widetilde{T}$.

The BAI problem has been extended to the case of linear bandits (Yang and Tan, 2022; Yavas and Tan, 2023). In their problem setting, the set of feature vectors associated to the arm set is assumed to be stationary. Although our loss model (1) is linear with respect to the feature vectors $x(\tau_1), \ldots, x(\tau_K)$, the set of feature vectors is non-stationary, where $\tau_i = \tau(i,t)$ is the pull count of arm i up to time step t, and $x(\tau) = (\tau^{-\rho_1}, \ldots, \tau^{-\rho_d})$. Thus, we cannot apply existing linear BAI algorithms to our problem.

The problem setting where each arm is an online learning algorithm is known as the bandit model selection problem (Agarwal et al., 2017; Pacchiano et al., 2020), however, most papers on the model selection focus on the cumulative regret minimization problem.

The decreasing bandits problem is a special instance of the rested bandit problem (Tekin and Liu, 2012), where the expected loss is a function of the number of arm pulls, and equivalent to rising bandits (Li et al., 2020) in the rewards formulation. Rising bandits were studied by (Heidari et al., 2016; Li et al., 2020) in the case of noise-free rewards setting for regret minimization, and were studied by (Metelli et al., 2022) for cumulative regret minimization. The assumption in (Metelli et al., 2022) is more general than ours, i.e., the expected rewards are assumed to be non-decreasing and concave with respect to the number of arm pulls. Mussi et al. (2023) studied BAI problem for rising bandits in the same setting as (Metelli et al., 2022), however, as mentioned in the introduction, they proved that no algorithm can identify the best arm unless the budget T is sufficiently large.

Cella et al. (2021) considered a pseudo (simple) regret minimization problem for decreasing bandits assuming that the loss is modeled as $\alpha/\tau^{\rho} + \beta$, where α, β are unknown coefficients. However, they assume that ρ is the same for all arms. In practice, the rate of the decay is unknown and it depends on each arm. The assumption of the model in this paper can be regarded as a generalization of Cella et al. (2021). Moreover, since their objective is the pseudo regret minimization, Cella et al. (2021) have not provided analysis on the probability of the error. Here, the pseudo regret is defined as defined as $R(T) = \mathbb{E}\left[y_{\widehat{i},\tau}\right] - \min_i \mathbb{E}\left[y_{i,T}\right]$, where $\widehat{i}$ is the selected arm by the BAI algorithm, τ is the pull counts of the arm $\widehat{i}$ up to the time step T. If we can construct an algorithm for the probability of the error minimization with $\widetilde{T} = T$, then we can also optimize the pseudo-regret by the Explore-Then-Commit framework (Lattimore and Szepesvári, 2020). We discuss application of our methods to pseudo regret minimization in Sec. B.

3 PRELIMINARIES

3.1 Problem Setting

Let K be the number of arms and T be the number of decision rounds. We consider a non-stationary best arm identification problem, where the expected observed loss depends only on the number of times that the arm i has been pulled up to time step t. For each time step t, a learner selects an arm $i_t \in [K]$, and observes loss $y_{i_t,\tau(i,t)}$, where $\{y_{i,\tau}\}_{i\in[K],\tau\in[T]}$ are random variables representing loss and $\tau(i,t) = \sum_{s=1}^{t} \mathbb{1}\{i_s = i\}$ denotes the number of times that the arm i has been pulled up to time step t. We assume the following model for the random variables $\{y_{i,\tau}\}$: $y_{i,\tau} = \sum_{m=1}^{d} \frac{\theta_m(i)}{\tau^{\rho_m}} + \varepsilon_{i,\tau}$, where for each i, $\theta(i) \in \mathbb{R}^d$ is a parameter unknown to the learner a priori, $\rho_1, \ldots, \rho_d \geq 0$ are known parameters, and $\varepsilon_{i,t}$ is a noise random variable. The random variables $\{\varepsilon_{i,t}\}_{1\leq i\leq K, 1\leq t\leq T}$ are taken to be independent and σ_0-subgaussian, i.e., $\mathbb{E}\left[\exp(\xi\varepsilon_{i,\tau})\right] \leq \exp(\xi^2\sigma_0^2/2)$, $\forall \xi \in \mathbb{R}$.

Given a target pull count $\widetilde{T}$, the best arm $i^* \equiv i^*(\widetilde{T})$ is understood to be the arm with the minimum expected loss after being pulled $\widetilde{T}$ times, i.e., $i^* = \operatorname{argmin}_{i\in[K]} \overline{y}_{i,\widetilde{T}}$, where $\overline{y}_{i,\widetilde{T}} = \mathbb{E}\left[y_{i,\widetilde{T}}\right]$, and where we assume that the minimum is uniquely achieved. Note that $\widetilde{T}$ may be different from T, the number of decision rounds in the learning process. After the final round T, the learner outputs an arm $\widehat{i} \in [K]$ as a guess for i^*. The objective of the learner is to keep the error probability $P(\widehat{i} \neq i^*)$ as small as possible. We note that existing works (Mussi et al., 2023; Cella et al., 2021) conduct analysis for the case $\widetilde{T} = T$. In Sections 4 & 5, we derive upper bounds in more relaxed conditions, i.e., $\widetilde{T} \geq \max_i \tau(i, T)$.

3.2 Assumptions

We assume that the expected losses $\overline{y}_{i,\tau}$ are normalized so that $0 \leq \overline{y}_{i,t} \leq 1$ for $1 \leq i \leq K, 1 \leq t \leq T$. For $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_d)$, let $\widetilde{\Sigma} = \widetilde{\Sigma}(\boldsymbol{\rho}) \in \mathbb{R}^{d\times d}$ denote the symmetric matrix with entries $(\widetilde{\Sigma})_{ij} = 1/(1-\rho_i-\rho_j)$. We assume that $\det \widetilde{\Sigma} \neq 0$. Moreover, we assume that the degrees of the monomials in the model satisfy $0 \leq \rho_m < 1/2$ for $1 \leq m \leq d$. This assumption can be justified as follows. If we only want to estimate $\overline{y}_{i,\widetilde{T}}$ for an arm i using observed samples $y_{i,1}, \ldots, y_{i,\tau}$, even in the stationary case (i.e., $d = 1$ and $\rho_1 = 0$), the width of the $(1-\delta)$-confidence interval of the estimator is about $O(\sqrt{\log(1/\delta)}/\sqrt{\tau}) = O(1/\tau^{1/2})$. Intuitively, one cannot hope for a better estimation error in the non-stationary case. If $\rho_m > 1/2$, then the term $\tau^{-\rho_m}$ in the loss is too small compared to the width of the confidence interval. Therefore, it is natural to assume $0 \leq \rho_m < 1/2$ for all $1 \leq m \leq d$.

3.3 Problem Complexity

Since our problem formulation includes the conventional stationary BAI problem, the conventional problem complexities such as H and H_2 (Audibert et al., 2010; Mannor and Tsitsiklis, 2004) play an important role in our analysis. We provide definitions of these notions in our setup. The problem complexity $H = H(\widetilde{T})$ is defined as $\sum_{i\in[K], i\neq i^*} \Delta_i(\widetilde{T})^{-2}$ Here, $\Delta_i(\widetilde{T}) = \overline{y}_{i,\widetilde{T}} - \overline{y}_{i^*,\widetilde{T}}$ is the optimality gap after $\widetilde{T}$ arm pulls. To define $H_2(\widetilde{T})$, here, we assume that arms $1, 2, \dots, K$ are sorted so that $\overline{y}_{1,\widetilde{T}} < \overline{y}_{2,\widetilde{T}} \dots \leq \overline{y}_{K,\widetilde{T}}$. and $H_2 = H_2(\widetilde{T})$ is defined as $\max_{2 \leq i \leq K} i\Delta_i(\widetilde{T})^{-2}$.

3.4 Notation

For a vector $x \in \mathbb{R}^d$ and $1 \leq p \leq \infty$, $\|x\|_p$ denotes the ℓ^p-norm. If $p = 2$, we simply denote $\|x\|_2$ by $\|x\|$. For an integer $d \geq 1$, we denote by $1_d \in \mathbb{R}^{d \times d}$ the identity matrix. For $n \in \mathbb{Z}_{\geq 1}$, $[n]$ denotes the set $\{1, \dots, n\}$. For a symmetric matrix $A \in \mathbb{R}^{d \times d}$, we denote by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ the minimum and maximum eigenvalue of A, respectively. For $i, j \in \mathbb{Z}$, δ_{ij} denotes Kronecker's delta. We put $\rho_{\max} = \max_i \rho_i$, and $\rho_{\min} = \min_i \rho_i$. We provide a list of notations in Section A.

4 GENERAL LOSS MODEL

In this section, we present an algorithm (Algorithm 1) that can be regarded as a non-stationary, model-based extension of Sequential Halving (abbreviated as SH) (Karnin et al., 2013), and provide upper bounds on the probability of its best arm identification error.

4.1 Estimators and the Model Complexity

Since the model (1) is a linear model with the time varying feature vector $x(\tau) = (\tau^{-\rho_1}, \dots, \tau^{-\rho_d})^\top \in \mathbb{R}^d$, it is natural to consider an estimator of the expected loss $\overline{y}_{i,\widetilde{T}}$ using an online least squares (LS) estimator. In this section, we assume that there exists $B > 0$ such that $\|\theta(i)\|_2 \leq B$ for $1 \leq i \leq K$. We fix an arm $1 \leq i \leq K$ and define an estimator $\overline{y}_{i,\widetilde{T}}$ as follows. Let $y_{i,\tau}, \dots, y_{i,\tau}$ be observed losses when the arm i has been selected τ times up to a time step t. Let $\lambda \geq 0$, be a regularizer, and we define an estimator $\widehat{y}^{(\tau)}_{i,\widetilde{T}}$ as:

$$\widehat{y}^{(\tau)}_{i,\widetilde{T}} = x^\top(\widetilde{T}) V_\tau^{-1} \sum_{s=1}^{\tau} y_{i,s} x(s), \tag{3}$$

where $V_\tau = \lambda 1_d + \sum_{s=1}^{\tau} x(s)x^\top(s) \in \mathbb{R}^{d \times d}$. One can consider an online least squares estimator $\widehat{y}^{(\tau)}_{i,\widetilde{T}}$ of $\overline{y}_{i,\widetilde{T}}$ for a linear model with any feature vectors $x(\tau)$. However, as we will see soon (Proposition 4.3 in the case of $\lambda = 0$), a non-trivial fact here is that $\widehat{y}^{(\tau)}_{i,\widetilde{T}} - \overline{y}_{i,\widetilde{T}}$ is $O(\frac{1}{\sqrt{\tau}})$-subgaussian, despite the non-stationary nature of the feature vector $x(\tau)$, where τ is the number of samples and the big-O notation hides constants depending on $\boldsymbol{\rho}, \sigma_0$.

To provide a more precise statement, we introduce a complexity measure for the general loss model. We note that the (i, j)-entry of the gram matrix $V_\tau = \lambda 1_d + \sum_{s=1}^{\tau} x(s)x^\top(s)$ is given as $\lambda\delta_{ij} + H(\tau, \rho_i + \rho_j)$. Here, for $\tau \geq 1, \rho > 0$, $H(\tau, \rho)$ is the generalized harmonic number defined as $H(\tau, \rho) := \sum_{s=1}^{\tau} s^{-\rho}$. The positive results in this paper are due to the special property of the generalized harmonic numbers explained as follows. It is easy to see that $H(\tau, \rho) \approx \tau^{1-\rho}/(1-\rho)$, but a more precise inequality can be derived using the Euler-Maclaurin formula. The following lemma is well-known in the field of analytic number theory, but for the sake of completeness, we provide a proof in Appendix.

Algorithm 1 SH with the LS estimator

- **Input**: time interval T, target pull count $\widetilde{T}$, number of arms K, regularizer $\lambda > 0$
- Initialize $A_0 \leftarrow [K]$
- **for** $r = 0, \dots, \lceil \log_2 K \rceil - 1$ **do**
  - Sample each arm $i \in A_r$ for $t_r = \left\lfloor \frac{T}{|A_r| \lceil \log_2 K \rceil} \right\rfloor$ times.
  - For $1 \leq i \leq K$, let $\widehat{y}^{(\tau_r)}_{i,\widetilde{T}}$ be the estimation (3) using the $\tau_r = \sum_{s=0}^{r} t_s$ observed losses.
  - Let A_{r+1} be the set of arms in A_r with the cardinality $\lceil |A_r|/2 \rceil$ with the smallest estimations.
- **end for**
- Output arm in $S_{\lceil \log_2 K \rceil}$

Lemma 4.1. *For $\tau \in \mathbb{Z}_{\geq 1}$ and $0 \leq \rho < 1$, we have*

$$\left| \tau^{\rho-1} H(\tau, \rho) - \frac{1}{1-\rho} \right| \leq \frac{2-\rho}{1-\rho} \tau^{\rho-1}.$$

Motivated by Lemma 4.1, we define a normalized Gram matrix $\widetilde{\Sigma}_{\tau,\lambda}$ and the model complexity as follows.

Proposition 4.2 (The model complexity in the general loss case)**.** *We assume vectors $x(1), \dots, x(d)$ are linearly independent. We define the model complexity γ as*

$$\gamma \equiv \gamma_{\boldsymbol{\rho}} = d \sup_{\tau \geq d} \left(\lambda_{\min}(\widetilde{\Sigma}_\tau) \right)^{-1}.$$

Here, $\widetilde{\Sigma}_{\tau,\lambda} \in \mathbb{R}^{d \times d}$ is defined as $(\widetilde{\Sigma}_\tau)_{ij} = \tau^{\rho_i+\rho_j-1} H(\tau, \rho_i + \rho_j)$. Then, we have $0 < \gamma < \infty$ and γ depends only on d, $\boldsymbol{\rho}$.

As previously mentioned, the following proposition states that $\widehat{y}^{(\tau)}_{i,\widetilde{T}} - \overline{y}_{i,\widetilde{T}}$ is $O(1/\sqrt{\tau})$-subgaussian (in the case of $\lambda = 0$). This unusual property of the estimator comes from the special property of the feature vector $x(\tau)$ (Lemma 4.1).

Proposition 4.3. *Suppose that $x(1), \dots, x(d)$ are linearly independent and $d \leq \tau \leq \widetilde{T}$. (i) If $\lambda = 0$, then $\widehat{y}^{(\tau)}_{i,\widetilde{T}} - \overline{y}_{i,\widetilde{T}}$ is $\sigma_0\sqrt{\gamma_{\boldsymbol{\rho}}}/\sqrt{\tau}$-subgaussian. (ii) Suppose $\lambda > 0$. Then, for each τ, with probability at least*

$1-\delta$, we have

$$|\widehat{y}^{(\tau)}_{i,\widetilde{T}} - \overline{y}_{i,\widetilde{T}}| \leq \left(\sigma_0\sqrt{2\log 2/\delta} + \sqrt{\lambda}\|\theta(i)\|\right)\frac{\sqrt{\gamma_{\boldsymbol{\rho}}}}{\sqrt{\tau}}.$$

4.2 Proposed Method and Upper Bounds

Proposition 4.3 shows that we can construct a estimator $\widehat{y}^{(\tau)}_{i,\widetilde{T}}$ of $\overline{y}_{i,\widetilde{T}}$ such that $\widehat{y}^{(\tau)}_{i,\widetilde{T}} - \overline{y}_{i,\widetilde{T}}$ is $R/\sqrt{\tau}$-subgaussian, where $R > 0$ is a constant. Here, for simplicity we consider the case of $\lambda = 0$. Since the $R/\sqrt{\tau}$-subgaussian property is the only property required of the loss (or reward) estimator for the performance analysis of SH (Karnin et al., 2013), it is reasonable to consider a variant (displayed in Algorithm 1) of SH. Similar to SH, in each phase $r = 0, \dots, \lceil \log_2 K \rceil - 1$, Algorithm 1 maintains a set of good arms A_r, then pulls each arm in A_r for t_r times, and A_{r+1} is the top half arms in A_r in terms of the estimations. Here, unlike SH, we use estimators provided by Proposition 4.3 when halving A_r and we use $\tau_r = \sum_{s=0}^{r} t_r$ samples in the estimation.

By Proposition 4.3 and the proof of (Karnin et al., 2013, Theorem 4.1), we have the following theorem:

Theorem 4.4. *Suppose $\|\theta_i\| \leq B$ for any $1 \leq i \leq K$ with $B > 0$ and that $\widetilde{T} \geq \tau_r \approx \left\lfloor \frac{T}{\lceil \log_2 K \rceil} \right\rfloor$, where $r = \lceil \log_2 K \rceil - 1$. Moreover, we assume that vectors $x(1), \dots, x(d)$ are linearly independent, and $t_0 \geq d$, where $t_0 = \left\lfloor \frac{T}{K\lceil \log_2 K \rceil} \right\rfloor$. Then, we have the following statements: (i) Suppose $\lambda = 0$. Then,*

$$P(\widehat{i} \neq i^*) \leq 4\lceil \log_2 K \rceil \exp\left(-\frac{T}{2^4\sigma_0^2\gamma_{\boldsymbol{\rho}} H_2 \lceil \log_2 K \rceil}\right),$$

(ii) Suppose $\lambda > 0$ and we put

$$p = 2K\lceil \log_2 K \rceil \exp\left(-\frac{T}{2^9\sigma_0^2\gamma_{\boldsymbol{\rho}} H_2 \lceil \log_2(K) \rceil}\right).$$

For $\delta \in (0,1)$, we define

$$\lambda(\delta) = 2\sigma_0\frac{\sqrt{2\log(2\lceil \log_2 K \rceil K/\delta)}}{B}.$$

We assume that $p < 1$. Then, for any $\varepsilon > 0$ satisfying $(1+\varepsilon)p \in (0,1)$, the probability of the error of Algorithm 1 with $\lambda = \lambda(\delta)$, $\delta = (1+\varepsilon)p$ is upper bounded as $P(\widehat{i} \neq i^*) \leq (1+\varepsilon)p$. Here $H_2 \equiv H_2(\widetilde{T})$.

Note that the theorem treats the case $\widetilde{T} \geq \tau_{\lceil \log_2 K \rceil - 1}$, i.e., $\widetilde{T} \geq \max_{1 \leq i \leq T} T_i$, where $T_i = \tau(i, T)$ the number of times the arm i has been selected. As mentioned in Section 3.1, this is a more relaxed condition compared to existing work. Moreover, we shall discuss the generalization to the case of general $\widetilde{T}$ in Section F. As discussed in the introduction, although the existing result requires the condition that the budget T is larger than $H^{(\rho)}(T)$ to identify the best arm, where ρ corresponds to $\min\{\rho_i : \rho_i > 0\}$ in our setting, Theorem 4.4 indicates that our algorithm can identify the best arm if $T \gtrsim \max(H_2(\widetilde{T}), dK)\log_2 K$, which can be smaller than $H^{(\rho)}(T)$. Here, the notation $\gtrsim$ hides constants such as the model complexity $\gamma_{\boldsymbol{\rho}}$, which is determined by the parameters of the model and algorithm.

Since the model complexity $\gamma_{\boldsymbol{\rho}}$ only depends on the parameter $\boldsymbol{\rho}$ of the model, it can be controlled by the algorithm, however, it could be large. To mitigate this issue, in the experiment section (Section 7), for a given d, we consider an optimization problem to select $\boldsymbol{\rho}$.

5 SPARSE LOSS MODEL

In the previous section, we defined the model complexity and provided upper bounds of the probability of the error under the general loss model setting. In this section, to reduce the model complexity, assuming the sparsity of the model (1), we define another model complexity γ_{s}, which can be smaller than $\gamma_{\boldsymbol{\rho},\lambda}$ under the sparsity assumption, and provide upper bounds of the probability of the error.

5.1 Sparsity of the Model and Estimator

Let $S \subseteq [d]$ be a subset. We assume that $\{m \in [d] : \theta_m(i) \neq 0\} \subseteq S$ for all arms $1 \leq i \leq K$. By definition the set S represents the sparsity of our loss model.

Next, we construct estimators. We fix $i \in [K]$ and $1 \leq t \leq T$. We assume that an algorithm has selected the i-th arm τ times, and losses $\{y_{i,s}\}_{1 \leq s \leq \tau}$ are observed up to time step t. We modify the feature vector $x(s) = (s^{-\rho_1}, \dots, s^{-\rho_d})^\top \in \mathbb{R}^d$ and the vector $\theta(i)$ as follows:

$$x'(s) = \mathrm{diag}(\tau^{\rho_1}, \dots, \tau^{\rho_d})x(s) = x(s/\tau).$$

and $\theta'(i) = \mathrm{diag}(\tau^{-\rho_1}, \dots, \tau^{-\rho_d})\theta(i)$. For $\lambda > 0$, we consider the optimization objective of LASSO with the modified feature vectors $\{x'(s)\}_{1 \leq s \leq \tau}$ as follows:

$$\widehat{\theta}_{\lambda,\tau}(i) = \mathrm{argmin}_{\alpha \in \mathbb{R}^d} \|Y_{i,\tau} - X'_\tau\alpha\|_2^2/\tau + \lambda\|\alpha\|_1. \quad (4)$$

Here $Y_{i,\tau} = (y_{i,s})_{1 \leq s \leq \tau} \in \mathbb{R}^\tau$ and $X'_\tau = (x'(1), \dots, x'(\tau))^\top \in \mathbb{R}^{\tau \times d}$. Then the empirical gram matrix $(X'_\tau)^\top X'_\tau/\tau$ is equal to $\widetilde{\Sigma}_\tau \in \mathbb{R}^{d \times d}$ defined as follows.

$$(\widetilde{\Sigma}_\tau)_{i,j} = \tau^{\rho_i + \rho_j - 1} H(\tau, \rho_i + \rho_j), \quad (5)$$

for $1 \leq i, j \leq K$.

5.2 Compatibility Constant

Following Bühlmann and Van De Geer (2011), we introduce the compatibility constant. As we will see shortly, the compatibility constant is a similar notion to the minimum eigenvalue of $\widetilde{\Sigma}_\tau$, however, its definition depends on the sparsity S, and can be larger than the minimum eigenvalue.

Definition 5.1 (Bühlmann and Van De Geer (2011))**.** Let $A \in \mathbb{R}^{d\times d}$ be a positive semi-definite matrix. For $S \subseteq [d], L > 1$, we define the set $\mathcal{R}(S, L) = \left\{\alpha \in \mathbb{R}^d : \|\alpha_{S^c}\|_1 \le L\|\alpha_S\|_1 \ne 0\right\}$. Here, for $\alpha \in \mathbb{R}^d$, $\alpha_S \in \mathbb{R}^d$ is defined as $(\alpha_S)_i = \alpha_i$ if $i \in S$ and $(\alpha_S)_i = 0$ otherwise. The *compatibility constant* $\phi^2(S, A, L) \ge 0$ is defined as follows:

$$\phi^2(S, A, L) = \min\left\{\frac{|S|\,\alpha^\top A\alpha}{\|\alpha_S\|_1^2} : \alpha \in \mathcal{R}(S, L)\right\}.$$

The definition of the compatibility constant is similar to that of the minimum eigenvalue. However, the minimization in the definition is restricted to the set $\mathcal{R}(S, L)$, it can be larger than the minimum eigenvalue if the sparsity if high. By definition, we have the following.

Proposition 5.2 (Bühlmann and Van De Geer (2011), Lemma 6.20, Lemma 6.23)**.** *For a positive semi-definite matrix $A \in \mathbb{R}^{d\times d}$, we have $\lambda_{\min}(A) \le \phi^2(S, A, L)$.*

We also remark that in the literature, L is assumed to be a fixed constant ($L = 3$) (Bühlmann and Van De Geer, 2011). We refer to (Bühlmann and Van De Geer, 2011, Chapter 6.13) for further lower and upper bounds of compatibility constants.

5.3 Concentration Inequality

The following proposition provides a confidence interval of an estimator of $\overline{y}_{i,\widetilde{T}}$ using (4). This proposition can be proved in the same way to (Bühlmann and Van De Geer, 2011, Theorem 6.1). However, due to some assumptions specific to this setting (e.g., $\theta'(i)$ depends on the number of samples τ, the upper bound is provided by a value specific to $\widetilde{\Sigma}_\tau$), we provide a proof for the sake of completeness in Section E.3.

Proposition 5.3. *Let $1 \le i \le K$ be an arm. For $\delta \in (0, 1)$, let $\lambda = \frac{(L+1)\beta}{L-1}\sqrt{\frac{2\log(2d/\delta)}{\tau}}$ be a regularizer, and $\widehat{\theta}_{\lambda,\tau}$ be a solution of the LASSO problem Eq. (4), where $\beta = 2\sqrt{(3 - 2\rho_{\max})/(1 - 2\rho_{\max})}$. We define an estimator $\widetilde{y}_{i,\tau}(\widetilde{T})$ of $\overline{y}_{i,\widetilde{T}}$ as*

$$\widetilde{y}^{(\tau)}_{i,\widetilde{T}} = \widehat{\theta}_{\lambda,\tau}(i) \cdot x'(\widetilde{T}). \tag{6}$$

We assume that $\tau \le \widetilde{T}$. Then, with probability at least $1-\delta$, we have $\left|\widetilde{y}^{(\tau)}_{i,\widetilde{T}} - \overline{y}_{i,\widetilde{T}}\right| \le C\sqrt{\frac{\log(2d/\delta)}{\tau}}$, where $C = \frac{2\sqrt{2}\beta L|S|}{(L-1)\phi^2(S,\widetilde{\Sigma}_\tau,L)}$.

5.4 The Model Complexity

We provide definition of the model complexity for the sparse case and we also define a time step τ_0 that will be used in the assumption of the main result of this section (Theorem 5.6).

Definition 5.4 (The complexity of the model in the sparse case)**.** We define $\tau_0 = \tau_0(S, L)$ as the smallest integer satisfying $\phi^2(S, \widetilde{\Sigma}_\tau, L) > 0$, for any $\tau \ge \tau_0$. Then, we define the complexity of the model $\gamma_{\mathrm{s}} = \gamma_{\mathrm{s}}(\boldsymbol{\rho}, S, L)$ by

$$\gamma_{\mathrm{s}} = \sup_{\tau \ge \tau_0(L,S)}\left(\frac{|S|}{\phi^2(S, \widetilde{\Sigma}_\tau, L)}\right)^2. \tag{7}$$

Compare to the definition of the model complexity provided in Proposition 4.2, although we take a square in Eq. (7), since it involves $|S|$ and the compatibility constants instead of the dimension and the minimum eigenvalues, γ_{s} can be smaller if the sparsity of the model is high by the remarks provided in Section 5.2. The following proposition states that similarly to the non-sparse case (Proposition 4.2), $\gamma_{\mathrm{s}}(\boldsymbol{\rho}, L, S)$ is a well-defined (finite) constant depending only on $\boldsymbol{\rho}, L, S$. We also provide an upper bound for the integer τ_0 under a mild condition.

Proposition 5.5. *(i) The integer $\tau_0 = \tau_0(S, L)$ satisfying the condition in Definition 5.4 exists. Moreover, if we assume that vectors $x(1), \dots, x(d) \in \mathbb{R}^d$ are linearly independent, then we have $\tau_0 \le d$.*

(ii) The model complexity $\gamma_{\mathrm{s}}(\boldsymbol{\rho}, L, S)$ depends only on $\boldsymbol{\rho}, L, S$ and is finite.

5.5 Upper Bounds

We propose another variant (Algorithm 2) of SH (Karnin et al., 2013). Similar to Algorithm 1, in each phase r, we construct estimators of $\overline{y}_{i,\widetilde{T}}$. For each phase r and arm $i \in A_r$, we consider estimators defined by (6) using $\sum_{s=0}^{r} t_s$ samples with the regularizer $\lambda = l/\sqrt{\tau_r}$, where $l > 0$ is an input of the algorithm.

Algorithm 2 SH with LASSO estimators

Input: T, $\widetilde{T}$, K, $l > 0$
The same procedure as Algorithm 1 except that we use estimators $\widetilde{y}^{(\tau_r)}_{i,\widetilde{T}}$ defined by (6) with the regularizer $\lambda = l/\sqrt{\tau_r}$ at line 5, where $\tau_r = \sum_{s=0}^{r} t_s$.

The next theorem is the main result in this section and provides upper bounds of the probability of the error of Algorithm 2.

Theorem 5.6. *Let $\gamma_{\mathrm{s}} = \gamma_{\mathrm{s}}(\boldsymbol{\rho}, S, L)$ be the complexity of the model and $\tau_0(S, L) \geq 1$ be the integer in Definition 5.4. We put*

$$p = 2dK\lceil \log_2 K \rceil \exp\left(-\frac{T}{c\beta^2\gamma_{\mathrm{s}} H_2 \lceil \log_2(K) \rceil}\right). \quad (8)$$

where $c = 2^9(1 - L^{-1})^2$, $H_2 = H_2(\widetilde{T})$ and β is a constant depending only on $\rho_{\max}$ given in Proposition 5.3. We define $l(\delta)$ as $\frac{(L+1)\beta}{L-1}\sqrt{2\log(2dK\lceil \log_2 K\rceil/\delta)}$. We assume that $p < 1$, $\widetilde{T} \geq \tau_r \approx \left\lfloor \frac{T}{\lceil \log_2 K \rceil} \right\rfloor$ and $t_0 \geq \tau_0(S, L)$, where $r = \lceil \log_2 K \rceil - 1$, $t_0 = \left\lfloor \frac{T}{K\lceil \log_2 K\rceil} \right\rfloor$. Then, for any $\varepsilon > 0$ satisfying $(1 + \varepsilon)p \in (0, 1)$, the probability of the error of Algorithm 2 with $l = l(\delta)$, $\delta = (1 + \varepsilon)p$ is upper bounded as $P(\widehat{i} \neq i^*) \leq (1 + \varepsilon)p$.

This theorem states that if the parameter l is appropriately selected, then the probability of the error is given as $O(p)$. As we discussed in Section 5.4, the model complexity γ_{s} can be smaller than $\gamma_{\boldsymbol{\rho},\lambda}$ if the sparsity is high. Therefore, Algorithm 2 can identify the best arm with a smaller budget than Algorithm 1 under the assumption. In Theorem 5.6, we assume that $\widetilde{T} \geq \tau_{\lceil \log_2 K\rceil - 1} \approx \left\lfloor \frac{T}{\lceil \log_2 K \rceil} \right\rfloor$ as in Theorem 4.4. In addition, we assume that $t_0 \geq \tau_0$, which is roughly equivalent to $T \geq \tau_0 K \log_2 K$. By Proposition 5.5, this assumption is satisfied if $T \geq dK \log_2 K$ under a mild assumption on $\boldsymbol{\rho}$.

6 LOWER BOUNDS

This section provides algorithm-independent lower bounds of the probability of error. Theorem 6.3, to follow, shows a lower bound roughly of the form $\exp\left(-O\left(\frac{T}{H(\widetilde{T})\Gamma_{\boldsymbol{\rho}}}\right)\right)$, where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_d)$ and $\Gamma_{\boldsymbol{\rho}}$ is a model complexity, which a constant depending only on $\boldsymbol{\rho}$. In the following, we provide the definition of the model complexity for the lower bounds, and state lower bounds of the probability of error. We then provide a comparison of our results to those of (Mussi et al., 2023, Theorems 6.1, 6.2), that give lower bounds in a more generalized setting but for possibly larger budgets.

6.1 The Model Complexity

We briefly explain how we construct environments to derive lower bounds and how we derive the model complexity for lower bounds. We provide more details in Section E.4. As explained in Section 3, a problem instance is defined by a set of vectors $\Theta = \{\theta(i)\}_{1 \leq i \leq K}$ and noise random variables $\{\varepsilon_{i,\tau}\}_{1\leq i \leq K, 1 \leq \tau \leq T}$. In this section, we assume the noise random variable $\varepsilon_{i,\tau}$ is given as an independent gaussian noise that follows $\mathcal{N}(0, \sigma_0^2)$. As in the stationary case (Carpentier and Locatelli, 2016), we construct K problem $\Theta(n) = \{\theta(i; n)\}_{1\leq i\leq K}$ for $n = 1, \ldots, K$, where in the n-th problem instance $\Theta(n)$, the best arm is the arm n. We let $f^n_{i,\tau}$ the distribution function of the random variable $y_{i,\tau}$. Similarly to the stationary case, to derive lower bounds, it is important to provide an upper bound of the empirical KL-divergence defined as : $\widehat{\mathrm{KL}}_{i,\tau} = \frac{1}{\tau}\sum_{s=1}^{\tau} \log\left(\frac{f^1_{i,s}(y_{i,s})}{f^i_{i,s}(y_{i,s})}\right)$. By the explicit form of $f^n_{i,\tau}$, we see that the empirical KL-divergence $\widehat{\mathrm{KL}}_{i,\tau}$ involves the gram matrix $\Sigma_\tau \in \mathbb{R}^{d\times d}$, where Σ_τ is defined as $(\Sigma_\tau)_{ij} = H(\tau, \rho_i + \rho_j)$. Similarly to Proposition 4.2 , we define the model complexity as follows:

Proposition 6.1 (The model complexity in the case of lower bound)**.** *We define $\Gamma_{\boldsymbol{\rho}} = \Gamma$ as*

$$\Gamma = \Gamma_{\boldsymbol{\rho}} = \inf_{\tau \in \mathbb{Z}_{\geq 1}} \left(\lambda_{\max}(\widetilde{\Sigma}_\tau)\right)^{-1}, \quad (9)$$

where $\widetilde{\Sigma}_\tau \in \mathbb{R}^{d\times d}$ is defined in (5)*. Then, $0 < \Gamma_{\boldsymbol{\rho}} < \infty$ and depends only on $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_d)$.*

Here, we define the model complexity using the maximum eigenvalues of $\widetilde{\Sigma}_\tau$ instead of the minimum eigenvalues as in Proposition 4.2. This gap is related to the ill-conditionedness of the matrix $\widetilde{\Sigma}$, and to mitigate this issue, we introduced the sparse loss model in Section 5. We refer to Sec. C for more details.

6.2 Assumption of a Class of the Algorithms

To provide an upper bound of the empirical KL-divergence, we require the following mild assumption regarding algorithms.

Assumption 6.2. Let $\mathcal{A}$ be an algorithm and denote $T_i = \tau(i, T)$ the number of times the arm i has been selected. We assume that there exists a positive constant α independent of $\widetilde{T}$ and T such that $\alpha\widetilde{T} \leq T_i \leq \widetilde{T}$, for any $1 \leq i \leq K$.

To explain the assumption, for simplicity, we assume that $\widetilde{T} = T$. Assumption 6.2 requires that the algorithm $\mathcal{A}$ pulls every arm at least a constant multiple of the budget T. This is a natural requirement since our problem setting includes the stationary best arm identification problem as a special case, and existing algorithms such as SH (Karnin et al., 2013) or Successive Rejects (Audibert et al., 2010) satisfy this assumption with $\alpha = \Omega(\frac{1}{K\log K})$ if $T = \widetilde{T}$. We note that our algorithms (Algorithm 1, 2) and an existing method RSR (Mussi et al., 2023) for the non-stationary BAI problem also have the same property. In the stationary case, if the optimality gaps $\{\Delta_i\}_{2 \le i \le K}$ at time step $\widetilde{T}$ are of the same order, a UCB-type algorithm (UCB-E Audibert et al. (2010)) satisfies this assumption and we provide the lower bound using such problem instances.

6.3 Lower Bounds

To state the main result of this section (Theorem 6.3), we introduce some notations as follows. For $\Theta = \{\theta(i)\}_{i=1}^{K} \subset \mathbb{R}^d$ and independent random variables $\mathcal{N} = \{\varepsilon_{i,t}\}_{1 \le i \le K, 1 \le t \le T}$ that are σ_0-subgaussian, we denote by $\mathcal{I} = (\Theta, \mathcal{N})$ the corresponding problem instance. For a problem instance $\mathcal{I}$, we denote by $H_{\mathcal{I}}$ the problem complexity $H(\widetilde{T})$. We define a set Π of problem instances satisfying $\overline{y}_{i,t} \in [0,1]$ for $1 \le i \le K, 1 \le t \le T$. For $a > 0$, we define a set of problem instances $\Pi_{\le a}$ as $\{\mathcal{I} \in \Pi : H_{\mathcal{I}} \le a\}$ for $a > 0$.

Theorem 6.3 (Lower bounds of the probability of error)**.** *We fix $\rho_1, \dots, \rho_d \in [0, 1/2)$. Let $\mathcal{A}$ be an algorithm satisfying Assumption 6.2. Furthermore, we assume that $\mathcal{A}$ satisfies $\sup_{\mathcal{I} \in \Pi_{\le a}} P_{\mathcal{I}}(\widehat{i} \ne i^*_{\mathcal{I}}) \le 1/2$, where $i^*_{\mathcal{I}}$ is the best arm in the problem instance $\mathcal{I}$. Then, there exist universal constants $c_1, \dots, c_5 > 0$ such that for any $a > c_1 K \widetilde{T}^{2\rho_{\min}}$ and T with $T \ge c_2 \sigma_0^2 a^2 \Gamma_{\boldsymbol{\rho}}^2 \alpha^{4\rho_{\max}} \sqrt{\log(KTc_3)}$ we have*

$$\sup_{\mathcal{I} \in \Pi_{\le a}} P_{\mathcal{I}}(\widehat{i} \ne i^*_{\mathcal{I}}) \ge c_4 \exp\left(-c_5 \frac{T}{\sigma_0^2 a \Gamma_{\boldsymbol{\rho}} \alpha^{2\rho_{\max}}}\right). \tag{10}$$

Here, since $\Delta_i(\widetilde{T}) \lesssim \widetilde{T}^{-\rho_{\min}}$, we require the condition that $a \gtrsim K\widetilde{T}^{2\rho_{\min}}$. If the loss model (1) has a constant term, i.e., $\rho_{\min} = 0$, then this is equivalent to $a \gtrsim K$. The assumption that $\sup_{\mathcal{I} \in \Pi_{\le a}} P_{\mathcal{I}}(\widehat{i} \ne i^*_{\mathcal{I}}) \le 1/2$ should be satisfied for any reasonable algorithm and the same condition was assumed in the stationary case (Carpentier and Locatelli, 2016). Our lower bound (10) is of the form $\exp\left(-O\left(\frac{T}{H(\widetilde{T})\Gamma_{\boldsymbol{\rho}} \alpha^{2\rho_{\max}}}\right)\right)$. The factors $\Gamma_{\boldsymbol{\rho}}$ and $\alpha^{2\rho_{\max}}$ are independent of T and $H(\widetilde{T})$, and by the remark in Section 6.2, we have $\alpha^{2\rho_{\max}} = \widetilde{\Omega}(K^{-1})$ for typical classes of algorithms.

Next, we compare Theorem 6.3 with (Mussi et al., 2023, Theorem 6.1, 6.2). For simplicity, we assume that $T = \widetilde{T}$ in this comparison. As explained in the introduction, they proved that in their problem setting, which is more general than ours, unless T satisfies (2), then, no algorithms cannot identify the best arm i^*, where ρ corresponds to $\min\{\rho_i : \rho_i > 0\}$ in our notation. In Theorem 6.3, we provide lower bounds under the assumption that the budget satisfies $T = \Omega(H(\widetilde{T})^2 \Gamma_{\boldsymbol{\rho}}^2)$. Ignoring a constant $\Gamma_{\boldsymbol{\rho}}$ that only depends on $\boldsymbol{\rho}$, we note that $H(\widetilde{T})^2$ can be smaller than the RHS of (2) especially for environments where expected losses are slowly decreasing ($\rho_i < 1/4$ for any i satisfying $\rho_i \ne 0$). Therefore, Theorem 6.3 holds even if the assumption (2) is not satisfied. We provide an example where the slowly decreasing assumption is valid. If we aim to minimize a function $F : \mathcal{X} \to \mathbb{R}$ using learning algorithms corresponding to the set of arms, the loss $y_{i,\tau}$ is related to simple regret of the learning algorithm. The lower bound of the expected simple regret of Bayesian optimization algorithms is given as $\widetilde{\Omega}(T^{-\nu/(2\nu+D)})$ (the case of Matèrn-ν kernels) (Vakili et al., 2021), where $\mathcal{X} \subset \mathbb{R}^D$. If $\nu = 1.5, D > 3$, then we have $\nu/(2\nu + D) < 1/4$.

7 EXPERIMENTS

Environments. We conduct experiments using synthetic environments described as follows. For $d \ge 1$, we select $\rho_1, \dots, \rho_d$ so that $\det \widetilde{\Sigma}$ is large. Specifically, we let $\rho_1 = 0.0$ and maximize $f(\rho_2, \dots, \rho_d) = \det \widetilde{\Sigma}$ on $[0.05, 0.45]^{d-1}$. Here, we note that the matrix $\widetilde{\Sigma}$ is approximately the gram matrix in the optimization problem (4). Let $2 \le m < n \le d$ and $a, b, a' > 0$. We define $y_{i,\tau}$ as $y_{i,\tau} = \begin{cases} a\tau^{-\rho_m} + b\tau^{-\rho_n} + \varepsilon_{i,\tau} & \text{if } i = 1, \\ a'\tau^{-\rho_m} + \varepsilon_{i,\tau} & \text{if } i \ge 2. \end{cases}$ Here, $\varepsilon_{i,\tau} \sim \mathcal{N}(0, \sigma_0^2)$ with $\sigma_0 = 0.01$. We consider two problem instances. One is lower dimensional $d = 4$ and the other one is higher dimensional $d = 50$, which are denoted by $\mathcal{I}_4$ and $\mathcal{I}_{50}$ respectively. In both environments, we assume $K = 5$, and the expected loss $\overline{y}_{i,\tau}$ is the same for $2 \le i \le K$ by definition, and the expected loss of arm 1 is initially larger than the others, but it decreases faster than them. In this experiments, we assume $\widetilde{T} \ge 50$ and $\widetilde{T} = T$. The best arm is the arm 1. We take $a = b = 0.5, a' = 0.8, \rho_m \approx 0.29, \rho_n = 0.45$ in the case of $\mathcal{I}_4$ and in the case of $\mathcal{I}_{50}$, we select a, b, a' so that $H(50)$ is the same as that in $\mathcal{I}_4$. We provide more details in Section G.

Baselines. We compare our algorithms to the state-of-the-art methods for decreasing bandits BAI and a well-known algorithm SH (Karnin et al., 2013) for the stationary BAI. RSH (Rising SH) is a modification of RSR (Rising Successive Rejects) (Mussi et al., 2023), which is a BAI algorithm for decreasing bandits. RSR is a variant of Successive Rejects (Audibert et al., 2010) and it estimates $\overline{y}_{i,\widetilde{T}}$ by the average $\widehat{y}(\tau, \varepsilon)$ of the most recent $\lfloor \epsilon\tau \rfloor$ the observed losses, where $\epsilon \in (0,1)$ is a parameter of the algorithm. In this experiment, we also consider a modification RSH of RSR based on SH. The only difference between our proposed methods and RSH are the definition of the estimators, which is $\hat{y}(\tau, \varepsilon)$ in the case of RSH, and it would not be difficult to see that RSH has a similar theoretical property as that of RSR. RUCBE is a UCB-type algorithm for the same setting as RSR and it is stated that RUCBE can identify the best arm for smaller budgets compared to

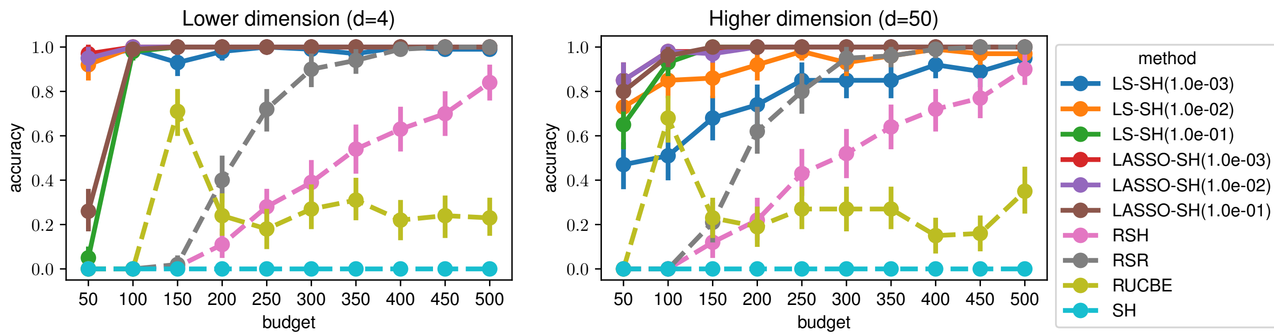

Figure 2: Experiments in synthetic environments

RSR (Mussi et al., 2023).

Results. In Figure 2, we display the accuracies of identifying the best arm in the problem instance $\mathcal{I}_4$ and $\mathcal{I}_{50}$ (the left and right figure, respectively). Here, we conduct the experiments for budgets $T = 50, 100, 150, \ldots, 500$ and repeat each experiment 100 times. The the error bars represent 95%-confidence interval over the repetition. LS-SH and LASSO-SH are our proposed methods (Algorithm 1, 2, respectively). For the parameters λ and l of LS-SH and LASSO-SH, we show results for the values $1\text{e}{-3}, 1\text{e}{-2}, 1\text{e}{-1}$ of a similar order as $\sigma_0 = 0.01$. Regarding the baselines, we select the same best parameter across the budgets, and show results for the best parameter. Figure 2 indicates that the proposed methods achieve higher accuracies with smaller budgets compared to the baselines, which empirically proves our theoretical results. We find that RUCBE performs well at first, but deteriorates gradually. We suspect that this is because RUCBE estimates $\overline{y}_{i,T}$, and if T is larger, then the bias can be larger. Results in the problem instance $\mathcal{I}_{50}$ show that that while LS-SH suffers from high dimensionality, LASSO-SH is less prone to it.

8 CONCLUSION

In this paper, we consider a best arm identification problem in the fixed budget setting, where the expected loss is monotone decreasing function as the number of arm pulls. We proposed two algorithms under sparse and general loss models, and provided upper bounds of the probability of the error. We also provided algorithm-independent lower bounds. Moreover, we conducted experiments in synthetic environments, and empirically verify our theoretical results. The extension to the case of of ϵ-optimal arm (Zhao et al., 2023) or that to the infinite-armed case (Li et al., 2017) would be an interesting possible future work.

Acknowledgments. Aditya Gopalan is grateful to Fujitsu Limited for sponsorship of this research. The first named author appreciates fruitful discussion with N. Namura and Y. Hikima.

References

Abbasi-Yadkori, Y., Bartlett, P., Gabillon, V., Malek, A., and Valko, M. (2018). Best of both worlds: Stochastic & adversarial best-arm identification. In *Conference on Learning Theory*, pages 918–949. PMLR.

Agarwal, A., Luo, H., Neyshabur, B., and Schapire, R. E. (2017). Corralling a band of bandit algorithms. In *Conference on Learning Theory*, pages 12–38. PMLR.

Apostol, T. M. (1999). An elementary view of Euler's summation formula. *The American Mathematical Monthly*, 106(5):409–418.

Audibert, J.-Y., Bubeck, S., and Munos, R. (2010). Best arm identification in multi-armed bandits. In *COLT*, pages 41–53.

Balandat, M., Karrer, B., Jiang, D., Daulton, S., Letham, B., Wilson, A. G., and Bakshy, E. (2020). BoTorch: A framework for efficient Monte-Carlo Bayesian optimization. *Advances in neural information processing systems*, 33:21524–21538.

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Carpentier, A. and Locatelli, A. (2016). Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Conference on Learning Theory*, pages 590–604. PMLR.

Cella, L., Pontil, M., and Gentile, C. (2021). Best model identification: A rested bandit formulation.

In *International Conference on Machine Learning*, pages 1362–1372. PMLR.

Eriksson, D., Pearce, M., Gardner, J., Turner, R. D., and Poloczek, M. (2019). Scalable global optimization via local Bayesian optimization. *Advances in neural information processing systems*, 32.

Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027. PMLR.

Heidari, H., Kearns, M. J., and Roth, A. (2016). Tight Policy Regret Bounds for Improving and Decaying Bandits. In *IJCAI*, pages 1562–1570.

Karnin, Z., Koren, T., and Somekh, O. (2013). Almost optimal exploration in multi-armed bandits. In *International conference on machine learning*, pages 1238–1246. PMLR.

Laguna, M. and Marti, R. (2005). Experimental testing of advanced scatter search designs for global optimization of multimodal functions. *Journal of Global Optimization*, 33:235–255.

Lattimore, T. and Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., and Talwalkar, A. (2017). Hyperband: A novel bandit-based approach to hyperparameter optimization. *The journal of machine learning research*, 18(1):6765–6816.

Li, Y., Jiang, J., Gao, J., Shao, Y., Zhang, C., and Cui, B. (2020). Efficient automatic CASH via rising bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 4763–4771.

Liu, D. C. and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 45(1-3):503–528.

Mannor, S. and Tsitsiklis, J. N. (2004). The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5(Jun):623–648.

Metelli, A. M., Trovo, F., Pirola, M., and Restelli, M. (2022). Stochastic rising bandits. In *International Conference on Machine Learning*, pages 15421–15457. PMLR.

Mussi, M., Montenegro, A., Trovo, F., Restelli, M., and Metelli, A. M. (2023). Best Arm Identification for Stochastic Rising Bandits. *arXiv preprint arXiv:2302.07510*.

Pacchiano, A., Phan, M., Abbasi Yadkori, Y., Rao, A., Zimmert, J., Lattimore, T., and Szepesvari, C. (2020). Model selection in contextual stochastic bandit problems. *Advances in Neural Information Processing Systems*, 33:10328–10337.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Tekin, C. and Liu, M. (2012). Online learning of rested and restless bandits. *IEEE Transactions on Information Theory*, 58(8):5588–5611.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294.

Vakili, S., Khezeli, K., and Picheny, V. (2021). On information gain and regret bounds in gaussian process bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 82–90. PMLR.

Valko, M., Korda, N., Munos, R., Flaounas, I., and Cristianini, N. (2013). Finite-Time Analysis of Kernelised Contextual Bandits. In *Uncertainty in Artificial Intelligence*, page 654.

van de Geer, S. A. and Bühlmann, P. (2009). On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, 3:1360–1392.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Yang, J. and Tan, V. (2022). Minimax Optimal Fixed-Budget Best Arm Identification in Linear Bandits. *Advances in Neural Information Processing Systems*, 35:12253–12266.

Yavas, R. C. and Tan, V. Y. (2023). Fixed-Budget Best-Arm Identification in Sparse Linear Bandits. *arXiv preprint arXiv:2311.00481*.

Zhao, Y. and Huo, X. (2023). A survey of numerical algorithms that can solve the Lasso problems. *Wiley Interdisciplinary Reviews: Computational Statistics*, 15(4):e1602.

Zhao, Y., Stephens, C., Szepesvari, C., and Jun, K.-S. (2023). Revisiting simple regret: Fast rates for returning a good arm. *International Conference on Machine Learning*.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes] See Section 3.2, and assumptions in each Lemma, Proposition, and Theorem.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes] See Sections 4, 5, D.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes] We will submit code as a supplementary material.
2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes] See Section 3.2, and assumptions in each Lemma, Proposition, and Theorem.
   (b) Complete proofs of all theoretical results. [Yes] See Section E.
   (c) Clear explanations of any assumptions. [Yes] See Section 3.2, and remarks after each Proposition and Theorem.
3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes] Code is available at `https://github.com/takemori135/model-based-BAI-for-decreasing-bandits`.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes] See Section G.1.
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes] See Section 7.
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes] See Section G.1.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Yes] See Section G.2.
   (b) The license information of the assets, if applicable. [Yes] See Section G.2.
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]
5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Appendix

A Notation

Table 1: List of Notations

Symbol	Description
T	budget
K	the number of arms
$\widetilde{T}$	time step at which the best arm is defined using the expected loss
$\tau = \tau(i, t)$	the number of times the arm i has been pulled up to time step t
$\theta(i) \in \mathbb{R}^d$	a coefficient vector of the loss model (1) for the i-th arm
$y_{i,\tau}$	loss random variable of the i-th arm observed by the learner at time step t
$\overline{y}_{i,\tau}$	the expected loss $\mathbb{E}\left[y_{i,\tau}\right]$
i^*	the best arm $i^* = \operatorname{argmin}_i \overline{y}_{i,\widetilde{T}}$
$\varepsilon_{i,\tau}$	noise random variable that is σ_0-subgaussian
$\boldsymbol{\rho} = (\rho_1, \dots, \rho_d)$	exponents of $\tau(i, t)$ in the loss model
$\rho_{\min}$	$\min_{1 \le i \le d} \rho_i$
$\rho_{\max}$	$\max_{1 \le i \le d} \rho_i$
$x(\tau)$	$(\tau^{-\rho_1}, \dots, \tau^{-\rho_d})^\top \in \mathbb{R}^d$
$\widetilde{\Sigma} = \widetilde{\Sigma}(\boldsymbol{\rho})$	a symmetric matrix of size d defined as $(\widetilde{\Sigma})_{ij} = 1/(1 - \rho_i - \rho_j)$
$H(\widetilde{T})$	a problem complexity of an instance (see Section 3)
$H_2(\widetilde{T})$	a problem complexity of an instance (see Section 3)
$H^{(\rho)}(T)$	a variant of the problem complexity defined as $\sum_{i \ne i^*} \frac{1}{\Delta_i(T)^{1/\rho}}$
$\Delta_i(\widetilde{T})$	the optimality gap for arm $i \in [K]$
$\widehat{y}^{(\tau)}_{i,\widetilde{T}}$	the LS estimator of $\overline{y}_{i,\widetilde{T}}$ using τ samples
$\gamma_{\boldsymbol{\rho},\lambda}$	the model complexity for the general loss model (Section 4)
B	upper bound of the norms $\\|\theta(i)\\|$
$\widetilde{y}^{(\tau)}_{i,\widetilde{T}}$	the LASSO estimator of $\overline{y}_{i,\widetilde{T}}$ using τ samples
$S \subset [d]$	the sparsity of the loss model
$L > 1$	a constant used in the definition of the compatibility constant
$\phi^2(S, A, L)$	the compatibility constant ($A \in \mathbb{R}^{d \times d}$: semi positive-definite)
$\gamma_{\mathrm{s}}(\boldsymbol{\rho}, S, L)$	the model complexity for the sparse loss model (Section 5)
$\Gamma_{\boldsymbol{\rho}}$	the model complexity for the lower bounds
$\\|A\\|_{\mathrm{F}}$	the Frobenius norm of a matrix A
$\\|A\\|_2$	the spectral norm of a matrix A
$\\|A\\|_\infty$	the sup norm of a matrix A

B APPLICATION to PSEUDO REGRET MINIMIZATION

In this section, we apply our methods to the pseudo regret minimization problem (Cella et al., 2021). In this problem setting, a learner outputs an arm $\widehat{i}$ after T rounds and pseudo regret $R(T)$ is defined as follows:

$$R(T) = \mathbb{E}\left[y_{\widehat{i},\tau}\right] - \mathbb{E}\left[y_{i^*,T}\right],$$

where $i^* = \operatorname{argmin}_i \mathbb{E}\left[y_{i,T}\right]$ and $\tau = \tau(\widehat{i}, T)$. Since the expected loss is decreasing, if the learner selects the right arm (i.e., $\widehat{i} = i^*$), then to minimize $R(T)$ the learner has to maximize $\tau(i^*, T)$, i.e., the leaner has to minimize the total number $\sum_{i \ne i^*} \tau(i, T)$ of pull counts of suboptimal arms.

We consider a class of algorithms satisfying the properties stated in Theorem 4.4 and Theorem 5.6.

Assumption B.1 (Assumption for a BAI algorithm)**.** We assume that an algorithm $\mathcal{A}(T_{\mathrm{e}}, \widetilde{T})$ for the decreasing BAI problem satisfies the following property, where T_{e} is the budget for the BAI problem. There exists constants

$a, b > 0, T_0(\boldsymbol{\rho}) > 0$ and $\gamma_{\boldsymbol{\rho}} > 0$ such that for any $T_\mathrm{e} \geq T_0(\boldsymbol{\rho})$ and $\widetilde{T} \geq T_\mathrm{e}$, the probability of the error of identifying the best arm at $\widetilde{T}$ is bounded as follows:

$$P\left(\widehat{i} \neq \mathrm{argmin}_i\, \overline{y}_{i,\widetilde{T}}\right) \leq b \exp\left(-\frac{T_\mathrm{e}}{a\gamma_{\boldsymbol{\rho}} H_2(\widetilde{T})}\right).$$

Then, we consider a method that follows the explore-then-commit framework as displayed in Algorithm 3.

Algorithm 3 ETC for Decreasing Bandits

1: **Input**: time interval T, period for exploration T_e, number of arms K, $\mathcal{A}$: algorithm for the decreasing bandits BAI problem
2: Run $\mathcal{A}(T_\mathrm{e}, \widetilde{T})$ with $\widetilde{T} = T$ using the budget T_e and let $\widehat{i}$ be the returned arm.
3: For the rest rounds ($t = T_\mathrm{e} + 1, \dots, T$), select the same arm $\widehat{i}$.

Then, we can easily prove the following result.

Proposition B.2. *We assume that there exist constants $\rho \in (0, 1/2)$ and $\alpha > 0$ such that $H(T) \leq \alpha T^{2\rho}$. Moreover, we assume that ρ and α are known to the learner. We assume $\theta(i^*) \in \mathbb{R}^d_{\geq 0}$ and define $\tilde{\rho}$ so that*

$$\mathbb{E}\left[y_{i^*,\tau}\right] - \lim_{\tau\to\infty} \mathbb{E}\left[y_{i^*,\tau}\right] = \Theta\left(\tau^{-\widetilde{\rho}}\right), \tag{11}$$

that is $\widetilde{\rho} = \min_{m \in M} \rho_m$, where $M = \{m \in [d] : \theta_m(i^*) \neq 0, \rho_m > 0\}$. Let $\mathcal{A}$ be an algorithm for the decreasing BAI satisfying Assumption 6.2 with constants $a, b, T_0(\boldsymbol{\rho}), \gamma_{\boldsymbol{\rho}} > 0$. Let $\nu > 0$ be a parameter satisfying $\nu > 1 + \widetilde{\rho}$. If we run Algorithm 3 with $T_\mathrm{e} = a\gamma_{\boldsymbol{\rho}}\alpha T^{2\rho}(\nu \log T + \log b)$ and T is sufficiently large, then the expected pseudo regret is upper bounded as follows:

$$\mathbb{E}\left[R(T)\right] \leq T^{-\nu} + \overline{y}_{i^*,T-T_\mathrm{e}} - \overline{y}_{i^*,T} = \widetilde{O}(T^{-\widetilde{\rho}-1+2\rho}).$$

Remark B.3. In the proposition, we assume that $H(T) \lesssim T^{2\rho}$ with $0 < \rho < 1/2$, which is a natural assumption by our model assumption (Section 3). Since $-1 + 2\rho < 0$, the proposition states that the expected pseudo regret $\mathbb{E}\left[R(T)\right]$ decreases faster than $\mathbb{E}\left[y_{i^*,\tau}\right] - \lim_{\tau\to\infty} \mathbb{E}\left[y_{i^*,\tau}\right]$. Compared to (Cella et al., 2021), this proposition provides a more interpretable result in a more generalized setting.

Proof of Proposition B.2. We assume that T satisfies $T \geq T_0(\boldsymbol{\rho})$ and we let $i^* = \mathrm{argmin}_i\, \overline{y}_{i,T}$. Let $\mathcal{E}$ be the event defined as $\{\omega : \widehat{i} \neq i^*\}$. By Assumption 6.2 and assumptions of the proposition, we have $P(\mathcal{E}) \leq T^{-\nu}$. On $\mathcal{E}^c$, the pseudo regret is given as $\overline{y}_{i^*,T-T_\mathrm{e}} - \overline{y}_{i^*,T}$. Thus, the expected pseudo regret is given as

$$\begin{aligned}\mathbb{E}\left[R(T)\right] &\leq T^{-\nu} + \overline{y}_{i^*,T-T_\mathrm{e}} - \overline{y}_{i^*} \\ &= T^{-\nu} + O((T - T_\mathrm{e})^{-\widetilde{\rho}} - T^{-\widetilde{\rho}}) = \widetilde{O}(T^{-\widetilde{\rho}-1+2\rho}).\end{aligned}$$

Here, the first equality follows from (11) and the second equality follows from $T_\mathrm{e} = \widetilde{O}(T^{2\rho})$ and the assumption on ν. □

C Upper and Lower Bounds

In this section, we discuss the gap between the upper bound provided in Section 4 and the lower bound provided in Section 6. Since $\widetilde{\Sigma}_{\tau,\lambda}$ and $\widetilde{\Sigma}_\tau$ can be approximated by $\widetilde{\Sigma}$, the gap of model complexities defined in Section 4 (general loss model) and Section 6 (lower bounds) is approximately given as $d\,\kappa(\widetilde{\Sigma})$, where $\kappa(\widetilde{\Sigma}) := \lambda_{\max}(\widetilde{\Sigma})/\lambda_{\min}(\widetilde{\Sigma})$ is the condition number of the matrix. Therefore, the question "how large is the gap between lower and upper bounds?" is approximately equivalent to "how ill-conditioned is the matrix $\widetilde{\Sigma}$?". The matrix $\widetilde{\Sigma}$ can be ill-conditioned, and it is related to the numerical instability of the estimators. This was the motivation of introducing the sparse loss model. We provide numerical examples below. As the table shows, the matrix $\widetilde{\Sigma}$ becomes ill-conditioned as d increases. We also provide numerical examples of (upper bounds of) the compatibility constant $\phi^2(S, \widetilde{\Sigma}, L)$ in the case of $|S| = 2$ (due to the non-convexity of the problem, we can only compute upper bounds of the compatibility constants). In Section 5, we use the compatibility constants instead of the minimum eigenvalues. The table indicates that the compatibility constants can be much larger than minimum eigenvalues (numerical examples of Proposition 5.2).

d	3	4	5	6	7
$\lambda_{\min}(\widetilde{\Sigma})$	4e-1	4e-1	2e-1	4e-2	9e-3
$d\,\kappa(\widetilde{\Sigma})$	2e+2	1e+3	5e+3	3e+4	2e+5
$\phi^2(S, \widetilde{\Sigma}, L)$	7e-1	9e-1	5e-1	3e-1	1e-1

D COMPUTATIONAL COMPLEXITY

We briefly discuss computational complexities of the proposed methods (Algorithm 1, 2). First, we consider the computational complexity of Algorithm 1. For each arm i and phase $r = 0, \dots, \lceil \log_2 K \rceil - 1$, the computational complexity of $\widehat{y}_{i,\widetilde{T}}^{(\tau_r)}$ is given as $O(d^2\tau_r + d^3)$, where $\tau_r = \sum_{s=0}^{r} t_s$. Thus, in total, the computational complexity of Algorithm 1 is given as $O(d^2 KT/\log_2 K + d^3 K \log_2 K)$.

Next, let us consider the computational complexity of Algorithm 2. We assume an oracle for solving the convex optimization problem (the LASSO optimization problem) (4) and denote by $C(\tau, d)$ its computational complexity. For the computational complexities of algorithms for LASSO, we refer to (Beck and Teboulle, 2009; Zhao and Huo, 2023). For each arm i and phase r, the computational complexity for $\widetilde{y}_{i,\widetilde{T}}^{(\tau)}$ is given as $O(d^2\tau_r + C(\tau_r, d))$, where $\tau_r \approx \frac{2^{r+1}T}{K\log_2 K}$. Thus, in total, the computational complexity is given as $O\left(d^2 KT/\log_2 K + K \sum_{r=0}^{\lceil \log_2 K \rceil - 1} C(\tau_r, d)\right)$.

E PROOFS

In this section, we provide proofs omitted in the main paper. In Section E.1, we prove Lemma 4.1. In Section E.2, we provide proofs of the results in Section 4 other than Lemma 4.1. In Section E.3, we provide proofs of the results in Section 5. In Section E.4, we provide proofs of the results in Section 6.

Additional Notations. We introduce additional notation used in the proofs. For a matrix $A = (a_{ij})$, we denote by $\|A\|_{\mathrm{F}}$ the Frobenius norm of a matrix A and define $\|A\|_\infty$ as $\max_{i,j} |a_{ij}|$. For a matrix $A \in \mathbb{R}^{d\times d}$, we denote by $\|A\|_2 \geq 0$ the spectral norm of A, i.e., $\|A\|_2^2 = \lambda_{\max}(A^\top A)$. If A is symmetric, and positive semi-definite, then $\|A\|_2 = \lambda_{\max}(A)$. For any A, we have $\|A\|_2 \leq \|A\|_{\mathrm{F}}$.

E.1 LEMMA FOR GENERALIZED HARMONIC NUMBERS

For $\tau \in \mathbb{Z}_{\geq 1}$ and $\rho > 0$, we define

$$H(\tau, \rho) = \sum_{s=1}^{\tau} s^{-\rho}.$$

Proof of Lemma 4.1. The statement of the lemma follows from the following inequality and we provide a proof below.

$$\left| H(\tau, \rho) - \left(\frac{1}{2} + \frac{\tau^{-\rho}}{2} + \frac{\tau^{1-\rho} - 1}{1-\rho} \right) \right| \leq \frac{1 - \tau^{-\rho}}{2}. \tag{12}$$

From the Euler-Maclaurin formula (see e.g. Apostol (1999)), we have the following (one can find the following formula in a standard text book on the Riemann zeta function).

$$\sum_{s=1}^{\tau} s^{-\rho} = \frac{1}{2} + \frac{\tau^{-\rho}}{2} + \frac{\tau^{1-\rho} - 1}{1-\rho} - \rho \int_1^{\tau} \frac{x - \lfloor x \rfloor - 1/2}{x^{\rho+1}} dx.$$

Then, the inequality (12) follows from $\rho \left| \int_1^{\tau} \frac{x - \lfloor x \rfloor - 1/2}{x^{\rho+1}} dx \right| \leq (1 - \tau^{-\rho})/2$. ◻

E.2 GENERAL LOSS MODEL

In this section, we provide proofs omitted in Section 4. First, we prove Proposition 4.2.

Proof of Proposition 4.2. By definition, γ depends only on λ and $\boldsymbol{\rho}$. We note that $\lambda_{\min}(\widetilde{\Sigma}_{\tau,\lambda}) > 0$ for any τ by the linear independence. To prove that $\gamma < \infty$, it is sufficient to prove $\lambda_{\min}(\widetilde{\Sigma}_{\tau,\lambda})$ is lower bounded by a constant independent of $\tau \in \mathbb{Z}_{\geq 1}$ for sufficiently large τ. We have the following:

$$\begin{aligned}\sqrt{d}\left|\lambda_{\min}(\widetilde{\Sigma}_{\tau,\lambda}) - \lambda_{\min}(\widetilde{\Sigma})\right| &\leq \|\widetilde{\Sigma}_{\tau,\lambda} - \widetilde{\Sigma}\|_{\mathrm{F}} \\ &\leq d\left(\lambda + \frac{2-2\rho}{1-2\rho}\right)\tau^{2\rho-1}.\end{aligned}$$

Here, $\rho = \max \rho_i$, the first inequality follows from the Wielandt-Hoffman Theorem, and the second inequality follows from Lemma 4.1. Since $\rho < 1/2$, if t_0 is sufficiently large, then we have

$$\frac{1}{2}\sqrt{d}\left(\lambda + \frac{2-2\rho}{1-2\rho}\right)t_0^{2\rho-1} < \lambda_{\min}(\widetilde{\Sigma}).$$

Thus, for any $\tau \geq t_0$, we have $\lambda_{\min}(\widetilde{\Sigma}_{t,\lambda}) \geq \lambda_{\min}(\widetilde{\Sigma})/2$. Since $\lambda_{\min}(\widetilde{\Sigma}) > 0$ by the assumption, we have our assertion. □

Next, we prove Proposition 4.3. In the stochastic bandit literature, the following well-known result is often used to construct an estimator. We refer to e.g. (Valko et al., 2013, Lemma 1).

Proposition E.1. *Let $x_1, \cdots, x_t \in \mathbb{R}^d$ be random vectors, and $\varepsilon_1, \ldots, \varepsilon_t$ be σ_0-subgaussian random variables. Let $\theta \in \mathbb{R}^d$ and define $y_s = \langle \theta, x_s \rangle + \varepsilon_s$ for $s = 1, \ldots, t$. For $\lambda \geq 0$, we put $V_t = \lambda 1_d + \sum_{s=1}^t x_s x_s^\top$. For $x \in \mathbb{R}^d$, we define $\mu_t(x) = xV_t^{-1}\sum_{s=1}^t y_s x_s$ and $\sigma_t(x) = \sqrt{x^\top V_t^{-1} x}$. We assume that $x_1, \ldots, x_t$ are independent of $\varepsilon_1, \ldots, \varepsilon_t$, random variables $\{\varepsilon_s\}_{s=1}^t$ are independent.*

1. *Suppose $\lambda = 0$ and $x_1, \ldots, x_t$ span R^d. Then, $\mu_t(x) - \langle \theta, x \rangle$ is $\sigma_0\sigma_t(x)$-subgaussian.*
2. *Suppose $\lambda > 0$. Then, for each $x \in \mathbb{R}^d$, we have the following inequality with probability at least $1-\delta$:*

$$|\mu_t(x) - \langle \theta, x \rangle| \leq \left(\sigma_0\sqrt{2\log 2/\delta} + \sqrt{\lambda}\|\theta\|\right)\sigma_t(x).$$

Proof. This follows from the proof of (Valko et al., 2013, Lemma 1). □

By Proposition 4.2 and Proposition E.1, we can prove Proposition 4.3 as follows.

Proof of Proposition 4.3. We define $\sigma_\tau(\widetilde{T})$ as

$$\sigma_\tau(\widetilde{T}) = \sqrt{x^\top(\widetilde{T})V_\tau^{-1}x(\widetilde{T})}.$$

Let $\widetilde{\Sigma}_{\tau,\lambda}$ be the symmetric matrix defined as $(\widetilde{\Sigma}_\tau)_{ij} = \tau^{\rho_i+\rho_j-1}\left(\lambda\delta_{ij} + H(\tau, \rho_i + \rho_j)\right)$. Then by definition, we have $\widetilde{\Sigma}_{\tau,\lambda}$ is given as

$$\frac{1}{\tau}\mathrm{diag}(\tau^{\rho_1}, \ldots, \tau^{\rho_d})V_\tau\mathrm{diag}(\tau^{\rho_1}, \ldots, \tau^{\rho_d}).$$

Therefore, we can rewrite $\sigma_\tau(\widetilde{T})$ as follows:

$$\begin{aligned}\sigma_\tau^2(\widetilde{T}) &= x^\top(\widetilde{T})V_\tau^{-1}x(\widetilde{T}) \\ &= x^\top(\widetilde{T}/\tau)\mathrm{diag}(\tau^{-\rho_1}, \ldots, \tau^{-\rho_d})V_\tau^{-1}\mathrm{diag}(\tau^{-\rho_1}, \ldots, \tau^{-\rho_d})x(\widetilde{T}/\tau) \\ &= \frac{1}{\tau}x^\top(\widetilde{T}/\tau)\widetilde{\Sigma}_{\tau,\lambda}^{-1}x(\widetilde{T}/\tau).\end{aligned}$$

By definition of $\gamma_{\boldsymbol{\rho}}$ (Proposition 4.2) and the fact that $\widetilde{\Sigma}_{\tau,\lambda} - \widetilde{\Sigma}_\tau$ is positive semi-definite, we have

$$\sigma_\tau^2(\widetilde{T}) \leq \frac{\gamma_{\boldsymbol{\rho}}\|x(\widetilde{T}/\tau)\|^2}{d\tau} \leq \frac{\gamma_{\boldsymbol{\rho}}}{\tau}. \tag{13}$$

Here the second inequality follows from $\tau \leq \widetilde{T}$. Thus, we have our assertion by Proposition E.1. □

The first statement of Theorem 4.4 can be proved by Proposition 4.3 and the proof of (Karnin et al., 2013, Theorem 4.1) as follows.

Proof of Theorem 4.4 1. In the proof, we put $n = \lceil \log_2 K \rceil$ and $m = \lfloor \log_2 K \rfloor$. We also let

$$R = \sigma_0 \sqrt{\gamma_{\boldsymbol{\rho}}}.$$

and we simply denote $\Delta_i(\widetilde{T})$ by Δ_i in this proof. We assume that the arms are sorted so that $\overline{y}_{1,\widetilde{T}} < \overline{y}_{2,\widetilde{T}} \leq \cdots \leq \overline{y}_{K,\widetilde{T}}$. If $\widetilde{T} \geq \tau$, then, by Proposition 4.3 and the assumption that noise random variables are independent, $\widehat{y}^{(\tau)}_{1,\widetilde{T}} + \widehat{y}^{(\tau)}_{i,\widetilde{T}} - \overline{y}_{1,\widetilde{T}} - \overline{y}_{i,\widetilde{T}}$ is $\sqrt{2}R/\sqrt{\tau}$-subgaussian (Lattimore and Szepesvári, 2020, Lemma 5.4). Thus, by (Lattimore and Szepesvári, 2020, Theorem 5.3), for any $2 \leq i \leq K$, we have

$$P\left(\widehat{y}^{(\tau)}_{1,\widetilde{T}} > \widehat{y}^{(\tau)}_{i,\widetilde{T}}\right) \leq \exp\left(-\frac{\tau \Delta_i^2}{R^2}\right). \tag{14}$$

For each phase, $r = 0, \ldots, n-1$, using τ_r samples, we compute $\widehat{y}^{(\tau_r)}_{i,\widetilde{T}}$, where $\tau_r = \sum_{s=0}^{r} t_s$, and $t_r = \left\lfloor \frac{T}{|A_r| n} \right\rfloor$. We note that by the assumption $\widetilde{T} \geq \frac{T}{\lceil \log_2 K \rceil}$, we have $\widetilde{T} \geq \tau_r$ for any r. Since $t_r \geq 1$, we note that $t_r \geq \frac{T}{2|A_r| n}$ and $|A_r| \leq 2^{n-r}$, and we have the following:

$$\tau_r \geq \frac{2^{-n+r-1} T}{n}. \tag{15}$$

For $r = 0, \ldots, n-1$, we denote by $\mathcal{E}_r$ the event on which the best arm i^* is eliminated in the phase r. We shall prove that

$$P(\mathcal{E}_r) \leq 4 \exp\left(-\frac{T}{2^4 n R^2 H_2(\widetilde{T})}\right). \tag{16}$$

If we can prove inequality (16), then we have the statement of the theorem by taking a union bound.

First, we assume $|A_r| = 2$ (the case of the last phase $r = n-1$). Then, by the inequality (14), for any $i \in A_r$ with $i \neq 1$, we have

$$P(\mathcal{E}_r) \leq P\left(\widehat{y}^{(\tau_r)}_{1,\widetilde{T}} > \widehat{y}^{(\tau_r)}_{i,\widetilde{T}}\right) \leq \exp\left(-\frac{2^{-3}T}{nR^2}\frac{\Delta_i^2}{2}\right) \leq \exp\left(-\frac{2^{-3}T}{nR^2}\frac{\Delta_2^2}{2}\right) \leq \exp\left(-\frac{2^{-3}T}{nR^2 H_2}\right),$$

where the second inequality follows from (15) with $r = n-1$. In particular, the inequality (16) holds.

Next, let us assume $|A_r| > 2$. Let A'_r be a subset of A_r such that $A_r \setminus A'_r$ is the set of the top $\lceil |A_r|/4 \rceil$ arms in A_r in terms of loss $\overline{y}_{i,\widetilde{T}}$. We also define a random set A''_r as

$$A''_r = \left\{ i \in A'_r : \widehat{y}^{(\tau_r)}_{i,\widetilde{T}} < \widehat{y}^{(\tau_r)}_{i^*,\widetilde{T}} \right\}.$$

Then, noting that for any $i \in A'_r$, we have $\Delta_i \geq \Delta_{i_r}$, where $i_r = \lceil |A_r|/4 \rceil$, and that the inequality (15) holds, we have

$$\begin{aligned}
\mathbb{E}\left[|A''_r|\right] &= \sum_{i \in A'_r} P\left(\widehat{y}^{(\tau_r)}_{i,\widetilde{T}} < \widehat{y}^{(\tau_r)}_{i^*,\widetilde{T}}\right) \\
&\leq |A'_r| \max_{i \in A'_r} \exp\left(-\frac{2^{-n+r-1} T \Delta_i^2}{nR^2}\right) \\
&\leq |A'_r| \exp\left(-\frac{2^{-n+r-1} T \Delta_{i_r}^2}{nR^2}\right) \\
&\leq |A'_r| \exp\left(-\frac{2^{-4}T}{nR^2}\frac{\Delta_{i_r}^2}{i_r}\right) \leq |A'_r| \exp\left(-\frac{2^{-4}T}{nR^2 H_2}\right),
\end{aligned}$$

where $i_r = \lceil |A_r|/4 \rceil$. Here, the first inequality follows from (14), the second inequality follows from the fact that $\Delta_i \le \Delta_{i_r}$ for any $i \in A'_r$, which holds by definition, and the third inequality follows from $i_r \ge 2^{m-2-r}$. If the event $\mathcal{E}_r$ holds, then we have $A''_r \supseteq A'_r \cap A_{r+1}$. Thus, we have

$$P(\mathcal{E}_r) \le P(A'_r \cap A_{r+1} \subseteq A''_r) \le P(|A'_r \cap A_{r+1}| \le |A''_r|) \le P(|A'_r|/4 \le |A''_r|).$$

Here, the third inequality holds since $|A'_r \cap A_{r+1}| \ge |A_{r+1}| - \lceil |A_r|/4 \rceil \ge |A'_r|/4$. By this inequality and the Markov's inequality, we have (16). This completes the proof. □

To prove the second statement of Theorem 4.4, we introduce the following lemma. In this lemma, similarly to Algorithm 2, we consider an arbitrary estimator of $\overline{y}_{i,\widetilde{T}}$ at line 5 in Algorithm 1.

Lemma E.2. *For each phase r in Algorithm 1, We assume that there is an estimator of $\breve{y}^{(\tau_r)}_{i,\widetilde{T}}$ of $\overline{y}_{i,\widetilde{T}}$ such that the following inequality holds for any $i \in [K]$ and any phase $r = 0, \dots, \lceil \log_2 K \rceil - 1$ with probability at least $1-\delta$:*

$$\left| \breve{y}^{(\tau_r)}_{i,\widetilde{T}} - \overline{y}_{i,\widetilde{T}} \right| \le C \sqrt{\frac{\log(\alpha/\delta)}{\tau_r}}. \tag{17}$$

Here $\alpha, C > 0$ We let

$$p = \alpha \exp\left(-\frac{T}{2^6 C^2 n H_2} \right),$$

and assume that $p < 1$. For any $\delta > p$, with probability at least $1-\delta$, a modification Algorithm 1 that uses the estimator $\breve{y}^{(\tau_r)}_{i,\widetilde{T}}$ at line 5 in Algorithm 1 returns the best arm i^*.

Proof. In the proof, we put $n = \lceil \log_2 K \rceil$ and $m = \lfloor \log_2 K \rfloor$. In the proof, we denote $\Delta_i(\widetilde{T})$ by Δ_i and we sort the arms so that $\overline{y}_{1,\widetilde{T}} < \overline{y}_{2,\widetilde{T}} \le \overline{y}_{3,\widetilde{T}} \le \dots \le \overline{y}_{K,\widetilde{T}}$. Therefore, $i^* = 1$. We note that $\tau_r = \sum_{s=0}^{r} t_s$, and $t_r = \left\lfloor \frac{T}{|A_r| n} \right\rfloor$. Since $t_r \ge 1$, $t_r \ge \frac{T}{2|A_r|n}$ and $|A_r| \le 2^{n-r}$, we have

$$\tau_r \ge \frac{2^{-n+r-1} T}{n}. \tag{18}$$

Let $\mathcal{E}'$ be an event on which (17) holds.

$$\mathcal{E}' = \left\{ \omega \in \Omega : \left| \breve{y}^{(\tau_r)}_{i,\widetilde{T}} - \overline{y}_{i,\widetilde{T}} \right| \le C \sqrt{\frac{\log(\alpha/\delta)}{\tau_r}}, \text{ for } i \in [K], r = 0, \dots, n-1 \right\},$$

where Ω is the sample space of the probability space on which we are working. Then, we have $P(\mathcal{E}') \ge 1 - \delta$.

We assume that (17) holds for any i and r. We also assume that i^* is eliminated in the r-th phase. We prove the following:

$$\log \frac{\alpha}{\delta} \ge \frac{T}{2^6 C^2 n H_2}. \tag{19}$$

First, we assume that $|A_r| > 2$. Let A'_r be a subset of A_r such that $A_r \setminus A'_r$ is the set of the top $\lceil |A_r|/4 \rceil$ arms in A_r in terms of loss $\overline{y}_{i,\widetilde{T}}$. We define A''_r by $A''_r = \left\{ i \in A'_r : \breve{y}^{(\tau_r)}_{i,\widetilde{T}} < \breve{y}^{(\tau_r)}_{i^*,\widetilde{T}} \right\}$. Since i^* is eliminated at the r-th phase, the half of arms in A_r is better than i^* in terms of $\breve{y}^{(\tau_r)}_{i,\widetilde{T}}$ by the argument in the proof of Theorem 4.4, we see that $A''_r \ne \emptyset$. By the definition of A''_r and (17), we have for any $i \in A''_r$

$$2C \sqrt{\frac{\log(\alpha/\delta)}{\tau_r}} \ge \Delta_i.$$

By the definition of A'_r, we have $\Delta_i \ge \Delta_{i_r}$ for any $i \in A'_r$, where $i_r = \lceil |A_r|/4 \rceil$. Thus, noting that (18) holds and $i_r \ge 2^{m-r-2}$, for $i \in A''_r$, we have

$$\log \frac{\alpha}{\delta} \ge \frac{\tau_r \Delta_i^2}{4C^2} = \frac{\tau_r i_r}{4C^2} \frac{\Delta_i^2}{i_r} \ge \frac{T}{2^6 C^2 n} \frac{\Delta_i^2}{i_r} \ge \frac{T}{2^6 C^2 n} \frac{\Delta_{i_r}^2}{i_r} \ge \frac{T}{2^6 C^2 n H_2}.$$

Therefore, the inequality (19) holds. We can easily prove (19) in the case of $|A_r| = 2$. Thus, we have $\delta \le 2dKn \exp\left(-\frac{T}{2^6 C^2 n H_2}\right)$.

Let us assume $\delta > \alpha \exp\left(-\frac{T}{2^6 C^2 n H_2}\right)$. If we assume that the event $\mathcal{E}'$ holds and i^* is eliminated by the algorithm, then we have deduced a contradiction. This completes the proof. □

The second statement of Theorem 4.4 can be proved by Lemma E.2.

Proof of Theorem 4.4 2. In the proof, we put $n = \lceil \log_2 K \rceil$. By Proposition 4.3 (ii), and taking a union bound for $i \in [K]$ and $r = 0, \dots, n-1$, we have the following inequality with probability at least $1 - \delta$:

$$|\widehat{y}^{(\tau)}_{i,\widetilde{T}} - \overline{y}_{i,\widetilde{T}}| \le \left(\sigma_0 \sqrt{2 \log(2nK/\delta)} + \sqrt{\lambda} B\right) \frac{\sqrt{\gamma_{\boldsymbol{\rho}}}}{\sqrt{\tau}}.$$

If we take $\lambda = \lambda(\delta)$ as in Theorem 4.4 (ii), then we have

$$|\widehat{y}^{(\tau)}_{i,\widetilde{T}} - \overline{y}_{i,\widetilde{T}}| \le 2\sigma_0 \sqrt{2\gamma_{\boldsymbol{\rho}}} \frac{\sqrt{\log(2nK/\delta)}}{\sqrt{\tau}}.$$

Then, the statement (ii) of Theorem 4.4 follows from this inequality and Lemma E.2. □

E.3 SPARSE MODEL

This section provides proofs omitted in Section 5.

First, we prove Proposition 5.5. We introduce a lemma below for it.

Lemma E.3 (van de Geer and Bühlmann (2009) Corollary 10.1)**.** *For two symmetric matrices $\Sigma_1, \Sigma_2 \in \mathbb{R}^{d \times d}$, we define $\|\Sigma_1 - \Sigma_2\|_\infty = \sup_{1 \le i,j \le d} |(\Sigma_1)_{ij} - (\Sigma_2)_{ij}|$. Then, we have*

$$\phi(S, \Sigma_1, L) \ge \phi(S, \Sigma_2, L) - (L+1)\sqrt{\|\Sigma_1 - \Sigma_2\|_\infty |S|}.$$

This lemma states that a function $\Sigma \mapsto \phi(S, \Sigma, L)$ is Lipschitz continuous. We prove Proposition 5.5 as follows.

Proof of Proposition 5.5. (i) By Lemma 4.1, we have

$$\|\widetilde{\Sigma}_\tau - \widetilde{\Sigma}\|_\infty < \frac{2 - 2\rho_{\max}}{1 - 2\rho_{\max}} \tau^{2\rho_{\max} - 1}.$$

By Proposition 5.2 and the assumption $\det \widetilde{\Sigma} > 0$, we have $\phi^2(S, \widetilde{\Sigma}, L) > 0$. We take τ_1 so that for any τ with $\tau \ge \tau_1$ satisfies the following:

$$(L+1)\sqrt{|S|} \sqrt{\frac{2 - 2\rho_{\max}}{1 - 2\rho_{\max}}} \tau^{\rho_{\max} - 1} < \phi(S, \widetilde{\Sigma}, L)/2.$$

Then, by Lemma E.3, we have

$$\phi^2(S, \widetilde{\Sigma}_\tau, L) \ge \phi^2(S, \widetilde{\Sigma}, L)/4, \tag{20}$$

if $\tau \ge \tau_1$. Since $\phi^2(S, \widetilde{\Sigma}, L) > 0$, we see that an integer τ_0 satisfying the condition in Definition 5.4 exists and $\tau_0 \le \tau_1$.

If we assume $x(1), \dots, x(d)$ are linearly independent and $\tau \ge d$, then $\phi^2(S, \widetilde{\Sigma}_\tau, L) > 0$ by Proposition 5.2. Therefore, we see that $\tau_0 \le d$.

(ii) By definition $\gamma_{\mathrm{s}}(\boldsymbol{\rho}, L, S)$ depends only on $\boldsymbol{\rho}, L, S$. In (1), we have proved that we can take $\tau_1 \ge 1$ so that if $\tau \ge \tau_1$, then (20) holds. Therefore, we see that

$$\sqrt{\gamma_{\mathrm{s}}} \le \max\left\{ \sup_{\tau_0 \le \tau \le \tau_1} \frac{|S|}{\phi^2(S, \widetilde{\Sigma}_\tau, L)}, \frac{4|S|}{\phi^2(S, \widetilde{\Sigma}, L)} \right\} < \infty.$$

□

To prove Proposition 5.3, we introduce a result on the confidence set of the estimator $\widehat{\theta}_{\lambda,\tau}(i)$. This can be proved in the same manner as (Bühlmann and Van De Geer, 2011, Theorem 6.1), but, we provide a proof for the sake of completeness.

Proposition E.4. *Let $1 \le i \le K$ be an arm. For any $\delta \in (0,1)$, we have the following inequality with probability at least $1-\delta$:*

$$\|\widehat{\theta}_{\lambda,\tau}(i) - \theta'(i)\|_1 \le C\sqrt{\frac{\log(2d/\delta)}{\tau}},$$

where $\lambda = \frac{(L+1)\beta}{L-1}\sqrt{\frac{2\log(2d/\delta)}{\tau}}$, $C = \frac{2\sqrt{2}\beta L|S|}{(L-1)\phi^2(S,\widetilde{\Sigma}_\tau,L)}$, and $\beta = 2\sqrt{(3-2\rho_{\max})/(1-2\rho_{\max})}$.

If we assume Proposition E.4 holds, then we can prove Proposition 5.3 as follows.

Proof of Proposition5.3. We assume that Proposition E.4 holds. Since $\widetilde{T} \ge \tau$ and $\rho_i \ge 0$ for $1 \le i \le d$, we have $\|x'(\widetilde{T})\|_\infty \le 1$. Then, noting that $\left|\widetilde{y}^{(\tau)}_{i,\widetilde{T}} - \overline{y}_{i,\widetilde{T}}\right| \le \|x'(\widetilde{T})\|_\infty \|\widehat{\theta}_{\lambda,\tau} - \theta'\|_1$, our assertion immediately follows from Proposition E.4. □

To prove Proposition E.4, we introduce the following lemma.

Lemma E.5. *For $1 \le m \le d$, let $(X'_\tau)^{(m)}$ be the m-th row of the matrix $X'_\tau = (x'(1),\ldots,x'(\tau))^\top \in \mathbb{R}^{\tau\times d}$. We define $\boldsymbol{\varepsilon}_\tau$ by $\boldsymbol{\varepsilon}_\tau = (\varepsilon_s)_{1\le s\le d} \in \mathbb{R}^d$. For a constant $\lambda_0 > 0$, we define an event $\mathcal{G}$ as follows:*

$$\mathcal{G} = \left\{\max_{1\le m\le d} 2\left|\boldsymbol{\varepsilon}_\tau \cdot (X'_\tau)^{(m)}\right|/\tau \le \lambda_0\right\}. \tag{21}$$

Then, we have

$$P(\mathcal{G}) \ge 1 - 2d\exp\left(-\frac{\lambda_0^2\tau}{2\beta^2}\right),$$

where $\beta = 2\sqrt{(3-2\rho_{\max})/(1-2\rho_{\max})}$ as before.

Proof. Let $\epsilon_m = 2\boldsymbol{\varepsilon}_\tau \cdot (X'_\tau)^{(m)}$. Since $\varepsilon_1,\ldots,\varepsilon_\tau$ are independent, ϵ_m is R-subgaussian (c.f. (Lattimore and Szepesvári, 2020, Lemma 5.4)), where R is given as

$$R = \frac{2}{\tau}\tau^{\rho_m}\sqrt{\sum_{s=1}^{\tau} s^{-2\rho_m}} = \frac{2}{\sqrt{\tau}}\sqrt{\tau^{2\rho_m-1}H(\tau,2\rho_m)} \le \frac{2}{\sqrt{\tau}}\sqrt{\frac{3-2\rho_m}{1-2\rho_m}}.$$

Here the last inequality follows from Lemma 4.1. Therefore, ϵ_m is $\beta/\sqrt{\tau}$-subgaussian. Then, the assertion of the lemma follows by the concentration inequality for subgaussian random variables (Lattimore and Szepesvári, 2020, Theorem 5.3) and taking a union bound for $m = 1,\ldots,d$. □

Proof of Proposition E.4. First, we note that the statement of the proposition can be rephrased as follows: Let $\lambda > 0$ and $\widehat{\theta}_{\lambda,\tau}$ be a solution of Eq. (4) with a regularizer $\lambda > 0$. Let $\mathcal{E}$ be an event defined as follows:

$$\mathcal{E} = \left\{\|\widehat{\theta}_{\lambda,\tau} - \theta'\|_1 \le \frac{2L|S|\lambda}{(L+1)\phi^2(S,\widetilde{\Sigma}_\tau,L)}\right\}.$$

Assume that $0 \le \rho_i < 1/2$ for all $1 \le i \le d$. Then, we have

$$P(\mathcal{E}) \ge 1 - 2d\exp\left(-\frac{(L-1)^2\lambda^2\tau}{2\beta^2(L+1)^2}\right).$$

In this proof, we assume that the event $\mathcal{G}$ defined by (21) holds and we simply denote $\widehat{\theta}_{\lambda,\tau}$ by $\widehat{\theta}$. Let $\boldsymbol{\varepsilon}_\tau$ by $\boldsymbol{\varepsilon}_\tau = (\varepsilon_s)_{1\le s\le d} \in \mathbb{R}^d$. By the basic inequality (Bühlmann and Van De Geer, 2011, Lemma 6.1), which can be directly derived from the definition of $\widehat{\theta}$, we have

$$\left\|X'_\tau(\widehat{\theta} - \theta')\right\|_2^2/\tau + \lambda\|\widehat{\theta}\|_1 \le 2\boldsymbol{\varepsilon}_\tau^\top X'_\tau(\widehat{\theta} - \theta')/\tau + \lambda\|\theta'\|_1.$$

By

$$2\boldsymbol{\varepsilon}_\tau^\top X'_\tau(\widehat{\theta} - \theta')/\tau \le \|2\boldsymbol{\varepsilon}_\tau^\top X'_\tau/\tau\|_\infty \|\widehat{\theta} - \theta'\|_1,$$

and the definition of $\mathcal{G}$, we have

$$\left\| X'_\tau(\widehat{\theta} - \theta') \right\|_2^2 /\tau + \lambda\|\widehat{\theta}\|_1 \le \lambda_0\|\widehat{\theta} - \theta'\|_1 + \lambda\|\theta'\|_1. \tag{22}$$

By the triangle inequality, we have the following:

$$\begin{aligned}\|\widehat{\theta}\|_1 &= \|\widehat{\theta}_S\|_1 + \|\widehat{\theta}_{S^c}\|_1 \\ &\ge \|\theta'_S\|_1 + \|\widehat{\theta}_{S^c}\|_1 - \|\widehat{\theta}_S - \theta'_S\|_1.\end{aligned}$$

Thus, by (22) and noting that $\|\theta'\|_1 = \|\theta'_S\|_1$, it follows that

$$\left\| X'_\tau(\widehat{\theta} - \theta') \right\|_2^2 /\tau + \lambda\left(\|\widehat{\theta}_{S^c}\|_1 - \|\widehat{\theta}_S - \theta'_S\|_1\right) \le \lambda_0\|\widehat{\theta} - \theta'\|_1.$$

Since $\|\widehat{\theta} - \theta'\|_1 = \|\widehat{\theta}_{S^c}\|_1 + \|\widehat{\theta}_S - \theta'_S\|_1$, we have

$$\left\| X'_\tau(\widehat{\theta} - \theta') \right\|_2^2 /\tau + (\lambda - \lambda_0)\|\widehat{\theta}_{S^c}\|_1 \le (\lambda + \lambda_0)\|\widehat{\theta}_S - \theta'_S\|_1. \tag{23}$$

Assume $\lambda > \lambda_0$ and let $L = \frac{\lambda+\lambda_0}{\lambda-\lambda_0}$. By (23), we see that $\widehat{\theta} - \theta' \in \mathcal{R}(S, L)$. Since $\widetilde{\Sigma}_\tau = (X'_\tau)^\top X'_\tau/\tau$ and by the definition of the compatibility constant $\phi^2(S, \widetilde{\Sigma}_\tau, L)$ and (23), we have

$$\|\widehat{\theta}_S - \theta'\|_1 \le \frac{2|S|L\lambda}{(L+1)\phi^2(S, \widetilde{\Sigma}_\tau, L)}.$$

The statement of the proposition follows from this inequality and Lemma E.5. □

Finally, we prove the main result of Section 5.

Proof of Theorem 5.6. In the proof, we put $n = \lceil \log_2 K \rceil$. In Algorithm 2, we use the estimator (6) with the regularizer $\lambda_r = \frac{(L+1)\beta}{L-1}\sqrt{\frac{2\log(2dKn/\delta)}{\tau_r}}$ in each phase r, then with probability at least $1-\delta$, we have the following inequality for $i \in [K]$ and $r = 0, \dots, n-1$:

$$|\widetilde{y}_{i,t_r}(T) - \mathbb{E}\left[y_{i,T}\right]| \le C\sqrt{\frac{\log(2dKn/\delta)}{\tau_r}}. \tag{24}$$

Noting that the assumptions $t_0 \ge \tau_0$ and $\widetilde{T} \ge \left\lfloor \frac{T}{n} \right\rfloor$ hold, this follows from by taking a union bound in Proposition 5.3 and the definition of γ_{s}. Here $C = 2\sqrt{2}\beta(1 - L^{-1})\sqrt{\gamma_s}$. The statement of the theorem follows from this inequality and Lemma E.2. □

E.4 LOWER BOUNDS

In this section, we provide proofs omitted in Section 6. First, we prove Proposition 6.1.

Proof of Proposition 6.1. Let $\widetilde{\lambda} = \sup_{\tau \in \mathbb{Z}_{\ge 1}} \lambda_{\max}(\widetilde{\Sigma}_\tau)$. Then, by definition, $\widetilde{\lambda}$ depends only on $\rho_1, \dots, \rho_d$, so does Γ. It is enough to prove $0 < \widetilde{\lambda} < \infty$. By the assumption that $\det \widetilde{\Sigma} \ne 0$ and the definition of $\widetilde{\lambda}$, we have $\widetilde{\lambda} > 0$. For a matrix $A \in \mathbb{R}^{d\times d}$, we denote by $\|A\|_2 \ge 0$ the spectral norm of A, i.e., $\|A\|_2^2 = \lambda_{\max}(A^\top A)$. Since for any A, we have $\|A\|_2 \le \|A\|_{\mathrm{F}}$, $\widetilde{\Sigma}$ and $\widetilde{\Sigma}_\tau$ are positive-semi definite, and by the triangle inequality of the spectral norm, we have the following:

$$\begin{aligned}\lambda_{\max}(\widetilde{\Sigma}_\tau) &\le \lambda_{\max}(\widetilde{\Sigma}) + \|\widetilde{\Sigma} - \widetilde{\Sigma}_\tau\|_2 \\ &\le \lambda_{\max}(\widetilde{\Sigma}) + \|\widetilde{\Sigma} - \widetilde{\Sigma}_\tau\|_{\mathrm{F}} \\ &\le \lambda_{\max}(\widetilde{\Sigma}) + \frac{2d(1-\rho_{\max})}{1-2\rho_{\max}}\tau^{-1/2+\rho_{\max}}.\end{aligned}$$

Here the last inequality follows from Lemma 4.1. Since $\rho_{\max} \in [0, 1/2)$ by assumption, we have our assertion. □

Since we omitted details on construction of environments in Section 6, we will provide them as follows. Given an environment, for $1 \le i \le K, 1 \le \tau \le T$, we denote by $y_{i,\tau}$ the observed loss (random variable) when arm i has been pulled τ-times. For a finite set $\Theta = \{\theta(i)\}_{1 \le i \le K} \subset \mathbb{R}^d$ of vectors, we define a probability model of $\{y_{i,\tau}\}_{1 \le i \le K, 1 \le \tau \le T}$ by

$$y_{i,\tau} = \sum_{m=1}^{d} \theta_m(i)\tau^{-\rho_m} + \varepsilon_{i,\tau} = \theta(i) \cdot x(\tau) + \varepsilon_{i,\tau}.$$

Here $\{\varepsilon_{i,\tau}\}_{1 \le i \le K, 1 \le \tau \le T}$ are independent random variable and each $\varepsilon_{i,\tau}$ follows the normal distribution $\mathcal{N}(0, \sigma_0^2)$, and $x(\tau)$ is defined as

$$x(\tau) = (\tau^{-\rho_1}, \ldots, \tau^{-\rho_d})^\top \in \mathbb{R}^d.$$

We denote by $(\Omega, \mathcal{F}, \mathcal{P}_\Theta)$ the corresponding probability space. The probability space $(\Omega, \mathcal{F}, P_\Theta)$ defines a problem instance for the best arm identification. By abuse of terminology, we often identify the problem instance with the probability measure P_Θ.

We let $\theta(1) \in \mathbb{R}^d_{\ge 0}$. For $2 \le i \le K$, we let $\Delta(i) \in \mathbb{R}^d_{\ge 0}$ so that $\theta(1) - \Delta(i) \in \mathbb{R}^d_{\ge 0}$. For example, by assuming $\rho_1 \le \rho_2 \le \cdots \le \rho_d$, we take $\theta(1) \in \mathbb{R}^d_{\ge 0}$ as

$$(\theta(1))_m = \begin{cases} \frac{1}{2} \text{ if } m = 1, \\ \frac{1}{2(d-1)} \text{ if } m \ge 2 \end{cases} \tag{25}$$

and take $\Delta(i) \in \mathbb{R}^d_{\ge 0}$ so that $(\Delta(i))_m \in [0, (\theta(1))_m)$ for $m = 1, \ldots, d$. Then, $(\theta(1) \pm \Delta(i)) \cdot x(\tau) \in [0, 1]$ is satisfied for any i, τ. We define $\theta(i) = \theta(1) + \Delta(i)$ and $\theta'(i) = \theta(1) - \Delta(i)$ for $2 \le i \le K$. We also define

$$d_{i,\widetilde{T}} = \Delta(i) \cdot x(\widetilde{T}),$$

for $2 \le i \le K$. Then, for $1 \le n \le K$, we define a finite sequence $\Theta(n)$ of vectors as $\Theta(i) = (\theta(1; n), \ldots, \theta(K; n))$, where

$$\theta(i; n) = \begin{cases} \theta(i) & \text{if } n = 1, \\ \theta(i) & \text{if } n \ge 2 \text{ and } i \ne n, \\ \theta'(i) & \text{if } n \ge 2 \text{ and } i = n. \end{cases}$$

For $1 \le n \le K$, we simply denote the probability measure (problem instance) $P_{\Theta(n)}$ by P_n. We note that in the problem instance P_n, arm n is the best arm by construction.

For a fixed problem instance P, we define a complexity H_P of the problem as

$$H_P = \sum_{1 \le i \le K, k \ne i^*} \frac{1}{(\overline{y}_{i,\widetilde{T}} - \overline{y}_{i^*,\widetilde{T}})^2},$$

where $i^* = \operatorname{argmin}_{1 \le i \le K} \overline{y}_{i,\widetilde{T}}$. For $1 \le n \le K$, we define $H(n)$ as H_{P_n}. Then, by construction, we see that $\max_{1 \le n \le K} H(n) = H(1)$. We note that $H(1)$ is given as $\sum_{i=2}^K d_{i,\widetilde{T}}^{-2}$ by construction. If we define $\theta(1)$ by (25), and define $\Delta(i)$ as

$$(\Delta(i))_m = \begin{cases} 1/2 - \eta^{-1/2} & \text{if } m = 1, \\ (\theta(1))_m & \text{if } 2 \le m \le d, \end{cases} \tag{26}$$

then, we have $H(1) = K\eta\widetilde{T}^{2\rho_1}$, where η is an arbitrary real number satisfying $\eta > 4$.

We fix an algorithm $\mathcal{A}$ and denote by T_i the number of times the algorithm selects arm i up to time step T. Then, each T_i is a random variable and satisfies $\sum_{i=1}^K T_i = T$. For $1 \le t \le T$, we denote by $\mathcal{F}_t$ the σ-algebra generated by observed losses by the algorithm up to time step t.

We denote by $f_{i,t}^n$ the distribution function on $\mathbb{R}$ associated to the random variable $y_{i,t}$ in the problem instance P_n. By definition, $f_{i,t}^n$ is the distribution function of $\mathcal{N}(\theta(i; n) \cdot x(t), \sigma_0^2)$, i.e., explicitly, $f_{i,t}^n(\xi) = \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left(-\frac{(\xi - \theta(i;n)\cdot x(t))^2}{2\sigma_0^2}\right)$. Following (Audibert et al., 2010; Carpentier and Locatelli, 2016), for $1 \le t \le T$, $2 \le i \le K$, we define

$$\widehat{\mathrm{KL}}_{i,t} = \frac{1}{t} \sum_{s=1}^{t} \log\left(\frac{f_{i,t}^1(y_{i,t})}{f_{i,t}^i(y_{i,t})}\right).$$

By the change of measure argument (Audibert et al., 2010), we have the following (we refer to (Audibert et al., 2010) for the proof).

Lemma E.6. *For any event $\mathcal{E} \in \mathcal{F}_T$, we have*

$$P_i(\mathcal{E}) = \mathbb{E}_1 \left[\mathbb{1}_{\mathcal{E}} \exp\left(-T_i \widehat{\mathrm{KL}}_{i,T_i} \right) \right].$$

The empirical KL-divergence $\widehat{\mathrm{KL}}_{i,t}$ is an estimator of the average KL-divergence $\overline{\mathrm{KL}}_{i,t}$ defined below.

Lemma E.7. *For $1 \le i \le K$ and $1 \le t \le T$, we define $\overline{\mathrm{KL}}_{i,t}$ as*

$$\overline{\mathrm{KL}}_{i,t} = \frac{1}{t} \sum_{s=1}^{t} \mathrm{KL}(f^1_{i,s}, f^i_{i,s}).$$

For $\delta > 0$, we define an event $\mathcal{G} = \mathcal{G}(\delta) \in \mathcal{F}_T$ as

$$\mathcal{G} = \left\{ \omega \in \Omega : |\widehat{\mathrm{KL}}_{i,T_i} - \overline{\mathrm{KL}}_{i,T_i}| \le \sqrt{\frac{8 \log(KT/\delta)}{\sigma_0^2 T_i}}, \quad 1 \le \forall i \le K \right\}.$$

Then, we have $P_1(\mathcal{G}) \ge 1 - \delta$.

Proof. We define an event $\mathcal{G}'$ as

$$\mathcal{G}' = \left\{ \omega \in \Omega : |\widehat{\mathrm{KL}}_{i,t} - \overline{\mathrm{KL}}_{i,T_i}| \le \sqrt{\frac{8 \log(KT/\delta)}{\sigma_0^2 t}}, \quad 1 \le \forall i \le K, 1 \le \forall t \le T \right\}.$$

Since $\mathcal{G}' \subseteq \mathcal{G}$, it is sufficient to prove that $P_1(\mathcal{G}') \ge 1 - \delta$. We let $\overline{y}_{i,t}(n) = \theta(i;n) \cdot x(t) \in [0,1]$. Then by definition, we have

$$\widehat{\mathrm{KL}}_{i,t} - \overline{\mathrm{KL}}_{i,t} = \frac{1}{\sigma_0^2 t} \sum_{s=1}^{t} (\overline{y}_{i,t}(1) - \overline{y}_{i,t}(i))(y_{i,t} - \overline{y}_{i,t}(1)).$$

Since we consider $P_1(\mathcal{G}')$, we assume $y_{i,t}$ follows distribution $f^1_{i,t}$, i.e., $y_{i,t} = \overline{y}_{i,t}(1) + \varepsilon_{i,t}$. Then, the inequality $P_1(\mathcal{G}') \ge 1 - \delta$ can be proved by (Lattimore and Szepesvári, 2020, Theorem 5.3) and taking a union bound. This completes the proof. □

Next, we provide an upper bound of the average KL-divergence $\overline{\mathrm{KL}}_{i,t}$ as below.

Lemma E.8. *For $2 \le i \le K$ and $\alpha \widetilde{T} \le t \le \widetilde{T}$, we have*

$$\overline{\mathrm{KL}}_{i,t} \le \frac{\alpha^{-2\rho_{\max}} d^2_{i,\widetilde{T}}}{2\Gamma \sigma_0^2}.$$

Proof. We have the following:

$$\begin{aligned}
\overline{\mathrm{KL}}_{i,t} &= \frac{1}{2\sigma_0^2 t} \sum_{s=1}^{t} \left(\theta(i;1) \cdot x(s) - \theta(i;i) \cdot x(s) \right)^2 \\
&= \frac{1}{2\sigma_0^2 t} \sum_{s=1}^{t} (\Delta(i)^\top x(s))^2 \\
&= \frac{1}{2\sigma_0^2 t} \operatorname{Tr} \sum_{s=1}^{t} x(s) x^\top(s) \Delta(i) \Delta(i)^\top \\
&= \frac{1}{2\sigma_0^2 t} \operatorname{Tr} D \sum_{s=1}^{t} x(s) x^\top(s) D D^{-1} \Delta(i) \Delta(i)^\top D^{-1},
\end{aligned}$$

where $D = D(\rho_1, \dots, \rho_d; t) := \mathrm{diag}(t^{\rho_1}, \dots, t^{\rho_d})$. Here the first equality follows from the definition and the fact that $\mathrm{KL}(\mathcal{N}(\mu, \sigma_0^2), \mathcal{N}(\mu', \sigma_0^2)) = (\mu - \mu')^2/(2\sigma_0^2)$, and the second equality follows from the construction of the problem instances. Thus, we have

$$\overline{\mathrm{KL}}_{i,t} = \frac{1}{2\sigma_0^2} \operatorname{Tr} \Sigma_t' D^{-1} \Delta(i) \Delta(i)^\top D^{-1},$$

where $\Sigma_t' \in \mathbb{R}^{d \times d}$ is defined by Eq. (5). Since the spectral norm $\| \cdot \|_2$ is dual to the trace norm, we have

$$\overline{\mathrm{KL}}_{i,t} \leq \frac{1}{2\sigma_0^2} \lambda_{\max}(\Sigma_t') \operatorname{Tr} D^{-1} \Delta(i) \Delta(i)^\top D^{-1}.$$

Since $\Delta(i) \in \mathbb{R}^d_{\geq 0}$, we have $\operatorname{Tr} D^{-1} \Delta(i) \Delta(i)^\top D^{-1} \leq (\Delta(i) \cdot x(t))^2$. By the assumption $t \geq \alpha \widetilde{T}$, we have

$$(\Delta(i) \cdot x(t))^2 \leq \alpha^{-2\rho_{\max}} (\Delta(i) \cdot x(\widetilde{T}))^2 = \alpha^{-2\rho_{\max}} d_{i,\widetilde{T}}^2.$$

Therefore, we have assertion of the lemma by Eq. (9). □

In the following lemma, we define an event $\mathcal{E}_i$ $(2 \leq i \leq K)$ on which the algorithm $\mathcal{A}$ returns $\widehat{i} = 1$. In the problem instance P_i, the arm 1 is not the best arm, and we provide a lower bound of $P_i(\mathcal{E}_i)$ as follows.

Lemma E.9. *For $1 \leq k \leq K$, we define*

$$t_k = \mathbb{E}_1 \left[T_k \right].$$

For $\delta, \delta' \in (0, 1)$ and $2 \leq i \leq K$, we define an event $\mathcal{E}_i \in \mathcal{F}_T$ as

$$\mathcal{E}_i = \left\{ \omega \in \mathcal{G}(\delta) : \widehat{i} = 1,\ T_i \leq t_i/\delta' \right\}.$$

Then, we have

$$P_i(\mathcal{E}_i) \geq \exp\left(-\frac{t_i d_{i,\widetilde{T}}^2 \alpha^{-2\rho_{\max}}}{2\Gamma \sigma_0^2 \delta'} - \sqrt{\frac{8 \log(KT/\delta)}{\sigma_0^2 T}} \right) P_1(\mathcal{E}_i).$$

Proof. Since $\mathcal{E}_i \in cF_t$,

$$\begin{aligned}
P_i(\mathcal{E}_i) &= \mathbb{E}_1 \left[\mathbb{1}_{\mathcal{E}_i} \exp\left(-T_i \widehat{\mathrm{KL}}_{i,T_i} \right) \right] \\
&\geq \mathbb{E}_1 \left[\mathbb{1}_{\mathcal{E}_i} \exp\left(-T_i \overline{\mathrm{KL}}_{i,T_i} - \sqrt{\frac{8 \log(KT/\delta)}{\sigma_0^2 T_i}} \right) \right] \\
&\geq \mathbb{E}_1 \left[\mathbb{1}_{\mathcal{E}_i} \exp\left(-T_i \overline{\mathrm{KL}}_{i,T_i} - \sqrt{\frac{8 \log(KT/\delta)}{\sigma_0^2 T}} \right) \right] \\
&\geq \mathbb{E}_1 \left[\mathbb{1}_{\mathcal{E}_i} \exp\left(-t_i \overline{\mathrm{KL}}_{i,T_i}/\delta' - \sqrt{\frac{8 \log(KT/\delta)}{\sigma_0^2 T}} \right) \right] \\
&\geq \mathbb{E}_1 \left[\mathbb{1}_{\mathcal{E}_i} \exp\left(-\frac{t_i d_{i,\widetilde{T}}^2 \alpha^{-2\rho_{\max}}}{2\Gamma \sigma_0^2 \delta'} - \sqrt{\frac{8 \log(KT/\delta)}{\sigma_0^2 T}} \right) \right] \\
&= \exp\left(-\frac{t_i d_{i,\widetilde{T}}^2 \alpha^{-2\rho_{\max}}}{2\Gamma \sigma_0^2 \delta'} - \sqrt{\frac{8 \log(KT/\delta)}{\sigma_0^2 T}} \right) P_1(\mathcal{E}_i).
\end{aligned}$$

Here the first equality follows from Lemma E.6, the fist inequality follows from Lemma E.7, the second inequality follows from $T_i \leq T$, the third inequality holds since we have $T_i \leq t_i/\delta'$ on $\mathcal{E}_i$, the fourth inequality follows from Lemma E.8. This completes the proof. □

Finally, we provide a proof of Theorem 6.3.

Proof of Theorem 6.3. This can be proved in a similar manner to Carpentier and Locatelli (2016). For the sake of completeness, we provide a proof. For given $a > 4K\widetilde{T}^{2\rho_{\min}}$, we take $\Delta(i)$ so that $a = H(1)$. This is possible if we define $\theta(1)$ and $\Delta(i)$ by (25) and (26). Since in the problem instance P_1, the arm 1 is the best arm, by the assumption $\sup_{\mathcal{I}\in\Pi_{\le a}} P_{\mathcal{I}}(\widehat{i} \neq i^*_{\mathcal{I}}) \le 1/2$, we have

$$P_1(\widehat{i} \neq 1) \le 1/2.$$

By Markov's inequality, we have

$$P_1(T_i \ge t_i/\delta') \le \delta' \mathbb{E}_1\left[T_i\right]/t_i = \delta'.$$

By these inequalities and Lemma E.7, we have

$$P_1(\mathcal{E}_i) \ge 1 - \delta - \delta' - 1/2 = 1/2 - \delta - \delta'.$$

Thus, by selecting δ, δ' so that $1/2 - \delta - \delta' > 0$ (e.g., $\delta = \delta' = 1/6$), we have the following inequality for $2 \le i \le K$ by Lemma E.9:

$$P_i(\mathcal{E}_i) \gtrsim \exp\left(-\frac{t_i d_{i,\widetilde{T}}^2 \alpha^{-2\rho_{\max}}}{2\Gamma\sigma_0^2\delta'} - \sqrt{\frac{8\log(KT/\delta)}{\sigma_0^2 T}}\right). \tag{27}$$

Here $\gtrsim$ hides a (universal) constant. Since $\sum_{i=2}^K d_{i,\widetilde{T}}^{-2} = H(1)$, there exists $2 \le i' \le K$ such that

$$t_{i'} \le \frac{T}{H(1) d_{i',\widetilde{T}}^2}.$$

By (27), we have

$$P_{i'}(\mathcal{E}_{i'}) \gtrsim \exp\left(-\frac{\alpha^{-2\rho_{\max}} T}{2\Gamma H(1)\sigma_0^2\delta'} - \sqrt{\frac{8\log(KT/\delta)}{\sigma_0^2 T}}\right).$$

Therefore, we have the following inequality with $a = H(1)$:

$$\sup_{\mathcal{I}\in\Pi_{\le a}} P_{\mathcal{I}}(\widehat{i} \neq i^*_{\mathcal{I}}) \gtrsim \exp\left(-\frac{\alpha^{-2\rho_{\max}} T}{2\Gamma a\sigma_0^2\delta'} - \sqrt{\frac{8\log(KT/\delta)}{\sigma_0^2 T}}\right).$$

Our assertion follows from this inequality. ☐

F GENERALIZATION OF THE UPPER BOUNDS

In Theorems 4.4, 5.6, we assume that $\widetilde{T}$ satisfies $\widetilde{T} \ge \tau_r \approx T/\log_2 K$, where $r = \lceil \log_2 K \rceil - 1$, $\tau_r = \sum_{s=0}^r t_r$, $t_r = \left\lfloor \frac{T}{|A_r|\lceil \log_2 K\rceil} \right\rfloor$ as in Algorithm 1. We provide a generalization of the theorems to the case when $\widetilde{T} < \tau_r$.

Theorem F.1. *We assume $\|\theta_i\| \le B$ for any $1 \le i \le K$ with $B > 0$. We define $\alpha_{\widetilde{T}} = \max(1, \tau_r/\widetilde{T})$, where $r = \lceil \log_2 K \rceil - 1$. Then, the probability of the error of Algorithm 1 satisfies the following:*

$$P(\widehat{i} \neq i^*) \le 4\lceil \log_2 K\rceil \exp\left(-\frac{T}{2^4 \alpha_{\widetilde{T}}^{2\rho_{\max}} \gamma_{\boldsymbol{\rho},\lambda}(B^2 + \frac{\sigma_0^2}{2\lambda}) H_2 \lceil \log_2 K\rceil}\right),$$

where $H_2 = H_2(\widetilde{T})$.

Next, we assume the same setting as in Section 5. Similarly to Theorem 5.6, we have the following.

Theorem F.2. *We define $\alpha_{\widetilde{T}} = \max(1, \tau_r/\widetilde{T})$, where $r = \lceil \log_2 K \rceil - 1$. Let $\gamma_{\mathrm{s}} = \gamma_{\mathrm{s}}(\boldsymbol{\rho}, S, L)$ be the complexity of the model and $\tau_0(S, L) \ge 1$ be the integer in Definition 5.4. We put*

$$p = 2dK\lceil \log_2 K\rceil \exp\left(-\frac{T}{c\alpha_{\widetilde{T}}^{2\rho_{\max}} \beta^2 \gamma_{\mathrm{s}} H_2 \lceil \log_2 (K)\rceil}\right).$$

where $c = 2^9(1 - L^{-1})^2$, $H_2 = H_2(\widetilde{T})$ and β is a constant depending only on $\rho_{\max}$ given in Proposition E.4. We define $l(\delta)$ as $\frac{(L+1)\beta}{L-1}\sqrt{2\log(2dK\lceil\log_2 K\rceil/\delta)}$. We assume that $p < 1$ and $t_0 \geq \tau_0(S, L)$, where $t_0 = \left\lfloor \frac{T}{K\lceil\log_2 K\rceil} \right\rfloor$. Then, for any $\varepsilon > 0$ satisfying $(1 + \varepsilon)p \in (0, 1)$, the probability of the error of Algorithm 2 with $l = l(\delta)$, $\delta = (1 + \varepsilon)p$ is upper bounded as $P(\widehat{i} \neq i^*) \leq (1 + \varepsilon)p$.

Since the proof is the same, we only provide a proof of Theorem F.1.

Proof of Theorem F.1. In Theorem 5.6, we need the condition that $\widetilde{T} \leq \tau_{\lceil\log_2 K\rceil-1}$ because it is necessary to bound the norm $\|x(\widetilde{T}/\tau)\|$ by $\sqrt{d}$. in (13). In Algorithm 1, for each $s = 0, \ldots, \lceil\log_2 K\rceil - 1$ phase, it constructs an estimator $\widehat{y}^{(\tau_s)}_{i,\widetilde{T}}$ using τ_s samples. We note that $\|x(\widetilde{T}/\tau_s)\| \leq \|x(\widetilde{T}/\tau_r)\| \leq \alpha_{\widetilde{T}}^{\rho_{\max}}\sqrt{d}$, where $r = \lceil\log_2 K\rceil - 1$. Then, instead of Proposition 4.3, we see that $\widehat{y}^{(\tau_s)}_{i,\widetilde{T}} - \overline{y}_{i,\widetilde{T}}$ is $\alpha_{\widetilde{T}}^{\rho_{\max}}\sqrt{\frac{\gamma_{\rho,\lambda}(B^2+\sigma_0^2/(2\lambda))}{\tau}}$-subgaussian for $s = 0, \ldots, \lceil\log_2 K\rceil - 1$. By the same proof of Theorem 5.6, we have the assertion of the theorem. □

G APPENDIX TO EXPERIMENTS

In this section, we provide details omitted in Section 7.

G.1 DETAILS OF THE EXPERIMENTS

Code

We provide code and data for reproducing experiments in a supplementary material.

Selection of $\boldsymbol{\rho}$

As described in Section 7, we let $\rho_1 = 0$ and select $\rho_2, \ldots, \rho_d$ as

$$\mathrm{argmax}_{(\rho_2,\ldots,\rho_d)\in[0.05,0.45]^{d-1}} \det \widetilde{\Sigma}(\boldsymbol{\rho}).$$

Here, we consider the maximization on $[0.05, 0.45]^{d-1}$, since errors in Lemma 4.1 become worse if $2\rho_i$ is close to 1 ($2 \leq i \leq d$). If $d = 4$, then $\boldsymbol{\rho}$ is given as $(\rho_2, \rho_3, \rho_4) = (0.2935618930156164, 0.41565736966993305, 0.45)$. For other dimensions, the values are stored in the file data/rhos.json in the code directory. For optimization, we used the L-BFGS-B method (Liu and Nocedal, 1989), which is available in the SciPy library (Virtanen et al., 2020).

Environments

As mentioned in Section 7, we consider the loss model as follows:

$$y_{i,\tau} = \begin{cases} a\tau^{-\rho_m} + b\tau^{-\rho_n} + \varepsilon_{i,\tau} & \text{if } i = 1, \\ a'\tau^{-\rho_m} + \varepsilon_{i,\tau} & \text{if } i \geq 2. \end{cases}$$

Here, $i \in [K]$ with $K = 5$, $\varepsilon_{i,\tau} \sim \mathcal{N}(0, \sigma_0^2)$ with $\sigma_0 = 0.01$. In the problem instance $\mathcal{I}_4$, we let $a = b = 0.5, a' = 0.8, \rho_m = \rho_2, \rho_n = \rho_4$. In the problem instance $\mathcal{I}_{50}$, we select a, b, a' so that $H(50)$ is the same as that in $\mathcal{I}_4$. Specifically, we let $a =' 0.8, b = 0.2, a' = 0.8954915084635513$ and $\rho_m = 0.14086303741335188, \rho_n = 0.3763414216486129$. In Figures 3, 4, we show expected losses $\mathbb{E}[y_{i,\tau}]$ in the problem instance $\mathcal{I}_4$ and $\mathcal{I}_{50}$. We also show problem complexities $H(\widetilde{T})$ for the problem instance $\mathcal{I}_4$ and $\mathcal{I}_{50}$ in Figure 5 and 6, respectively. As we mentioned in Section 7, the expected loss of arm 1 is initially larger than that of the other arms, but it decreases faster than the other arms. Also, if $\widetilde{T} \geq 50$, the best arm is the arm 1.

Choice of the Parameters

Next, we detail how we selected parameters of the algorithms. For the proposed methods (Algorithm 1 and Algorithm 2), we show experimental results for $\lambda = 1\mathrm{e}{-3}, 1\mathrm{e}{-2}, 1\mathrm{e}{-1}$ and $l = 1\mathrm{e}{-3}, 1\mathrm{e}{-2}, 1\mathrm{e}{-1}$, which is a similar order as $\sigma_0 = 1\mathrm{e}{-2}$. For the baselines, in Section 7, we only show the results for the "best" parameter. Here, the best parameter is defined as follows. For a parameter κ of a baseline algorithm, we repeat the experiments for 100

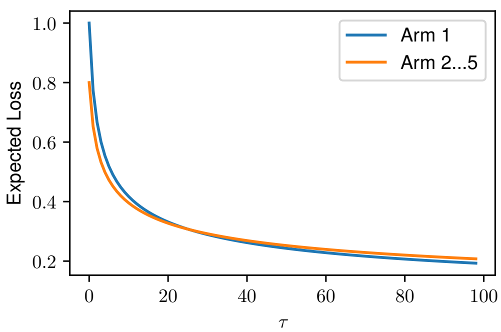

Figure 3: Expected losses in the problem instance $\mathcal{I}_4$ ($K = 5, d = 4$)

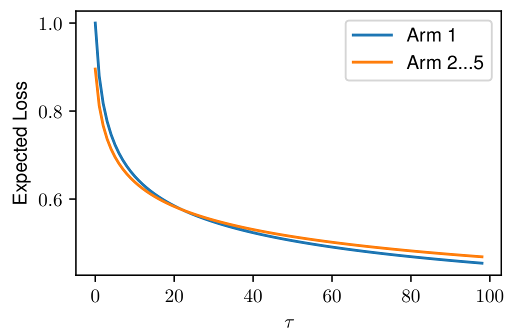

Figure 4: Expected losses in the problem instance $\mathcal{I}_{50}$ ($K = 5, d = 50$)

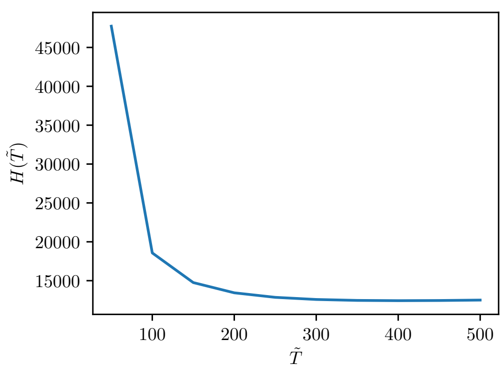

Figure 5: $H(\widetilde{T})$ in the problem instance $\mathcal{I}_4$ ($K = 5, d = 4$)

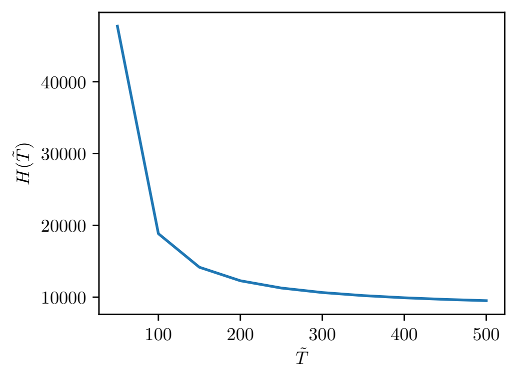

Figure 6: $H(\widetilde{T})$ in the problem instance $\mathcal{I}_{50}$ ($K = 5, d = 50$)

times for each budget $T = 50, 100, \ldots, 500$. For each budget T, we compute accuracy $\mathrm{acc}(\kappa, T) = \sum 1(\widehat{i} = i^*)/100$ over the repetition and define a metric $m(\kappa)$ by the median of $\mathrm{acc}(\kappa, T)$ for $T = 50, 100, \ldots, 500$. For the baselines, we only show the results for a parameter κ with the largest median $m(\kappa)$. A set of parameters of the baselines are selected as follows. For each arm i, RSH and RSR use the estimator $\widehat{y}_i(\tau, \varepsilon)$, which is the average of the most recent $\lfloor \varepsilon\tau \rfloor$ observed losses. For these algorithms, we conducted experiments for $\varepsilon = 0.1, 0.2, 0.3, 0.4, 0.5$. For RUCBE, we take $\varepsilon = 0.25$ as in (Mussi et al., 2023) and conducted experiments for $a = 0.1, 0.2, \ldots, 0.9$, where a is the exploration parameter.

Computational Resources

We conducted the experiments using the Intel Xeon Gold 6148 processor with 30GB RAM.

G.2 EXISTING ASSETS AND LICENSE

We implemented the algorithms using the Scikit-Learn library (Pedregosa et al., 2011), which is licensed under the BSD 3-Clause "New" or "Revised" License. To generate Figure 1, we optimized a test function using a Bayesian optimization algorithm (Eriksson et al., 2019). For the experiment, we use the BoTorch (Balandat et al., 2020) implementation of the Bayesian optimization algorithm, where BoTorch is licensed under the MIT license.