Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

BEST-ARM IDENTIFICATION WITH HIGH-DIMENSIONAL FEATURES

Dohyun Ahn[1], Dongwook Shin[2], and Lewen Zheng[1]

[1]Department of SEEM, The Chinese University of Hong Kong, Shatin, NT, HONG KONG
[2]Department of ISOM, HKUST Business School, Clear Water Bay, Kowloon, HONG KONG

ABSTRACT

Given a collection of stochastic systems (or arms), we focus on the problem of identifying the best system with the highest expected payoff by learning the unknown statistical characteristics of system payoffs via sequential sampling. The distributions of the system payoffs are governed by a linear model consisting of high-dimensional system features that are fixed and known, as well as unknown parameters that are common across all systems. However, due to the high dimensionality, the ordinary least-squares estimator for the unknown parameters exhibits a large variance, leading to significant errors in identifying the best system. Based on the theory of large deviations, we show that this performance degradation can be effectively addressed by using the LASSO estimator with a judiciously chosen regularization parameter. Furthermore, we provide a practical guideline for selecting the regularization parameter and design a dynamic sampling policy that improves the performance in identifying the best system.

1 INTRODUCTION

We explore a best-arm identification problem, where a decision maker aims to identify the best among several competing systems (or arms), where the "best" is defined as the one with the highest expected payoff. It is assumed that the statistical characteristics of random payoffs generated from the systems are a priori unknown but can be learned through sample observations. This is an example of an ordinal optimization problem, where it is common to sequentially allocate samples across the systems in a way that maximizes the likelihood of correctly identifying the best system (Ho et al. 2008).

A conventional approach to address the above problem is to estimate the sample mean (or its variants) of each system's payoff and allocate samples based on these estimates. This nonparametric approach requires allocating sufficient samples to *all* systems, and thus, can become inefficient when the sample size is limited relative to the number of systems (Branke et al. 2007). In this paper, we alternatively focus on a parametric setting where the underlying payoff distributions are characterized by a linear model composed of (1) known system features, which provide detailed information about the systems, (2) unknown parameters that are common across all systems, and (3) random noise associated with each sample. In particular, each system's expected payoff is represented as the product of its system features and the unknown parameters. Accordingly, it is possible to learn the unknown parameters and, consequently, the expected payoffs even when the sample allocation is not well-balanced across the systems.

However, a significant challenge arises in this parametric setting when the features are *high-dimensional*, which is frequently observed in recent practical applications. For example, in online marketplaces, the number of products (e.g., fashion items) has grown significantly, leading to the use of principal features for item descriptions, such as product category, brand, material, design, and price. Furthermore, sociodemographic attributes of customers (e.g., income, gender, and education) are often incorporated as additional features. Learning a parametric model with high-dimensional features can be prohibitively costly because samples are often obtained through controlled experiments on customers.

A possible remedy is to perform variable selection and regularization, such as the LASSO (Tibshirani 1996), to identify a sparse subset of features, which enhances the efficiency of estimation and the

interpretability of the resulting model. Such methods penalize the non-zero coefficients of the unknown parameters, thereby enforcing some coefficients to be zero. This reduces variance but introduces bias in estimating the unknown parameters. While the reduction in variance is beneficial in decreasing the sample size required to estimate the expected payoffs, the associated bias may increase the likelihood of selecting a non-best system. Therefore, it is crucial to strike a balance between bias and variance in these regularized regression methods for effective best-arm identification.

The main contribution of our work is three-fold. First, we characterize the large deviations rate function associated with the LASSO estimator for a static sampling policy. (See Dembo and Zeitouni (2009) for an introduction to large deviations theory.) When the sampling budget is sufficiently large, the problem of identifying the optimal sample allocation, which aims to minimize the probability of selecting a non-best system, reduces to the problem of finding the allocation that maximizes the rate function associated with that probability. Second, we demonstrate that the performance can be significantly improved by using the LASSO estimator instead of the ordinary least-squares estimator, provided that a suitable regularization parameter is chosen. Furthermore, we provide a guideline for how to identify an efficient range of the regularization parameter that leads to superior performance of the LASSO-based sampling policy. Third, we propose a dynamic sampling policy by leveraging the structural properties of the LASSO-based rate function as well as the aforementioned guideline.

In the best-arm identification literature, there has been considerable work in the situation where the mean system performance is described by a linear function of features. For instance, Soare et al. (2014) propose a static sampling algorithm and analyze the associated sample complexity, and Xu et al. (2018) develop a dynamic sampling algorithm that can make a substantial improvement over static ones. In cases with the generalized linear model, Ahn and Shin (2020) and Kazerouni and Wein (2021) propose dynamic sampling policies in fixed-budget and fixed-confidence settings, respectively. More recently, Ahn et al. (2024) extend the results of Ahn and Shin (2020) to the case when the system features are misspecified. Despite the ever-growing importance of high-dimensional data in a wide range of applications, the literature is scant on the implications of high-dimensional features in best-arm identification problems. Only a few studies have investigated the issue of dimensionality in fixed-confidence settings (Camilleri et al. 2021; Zhu et al. 2021). To the best of our knowledge, this work serves as the first attempt to tackle this issue in a fixed-budget setting.

In the field of classical multi-armed bandit problems that aim to minimize cumulative regret, there exists a vast literature on the linear bandit problem, where the mean of the underlying distribution is characterized by a linear function of features; see, for example, Auer et al. (2002), Dani et al. (2008), Rusmevichientong and Tsitsiklis (2010), Chu et al. (2011), and Abbasi-Yadkori et al. (2011). Within this category, a strand of research has emerged that focuses on tackling the issue of dimensionality using LASSO. Several studies by Abbasi-Yadkori et al. (2012), Carpentier and Munos (2012), Kim and Paik (2019), Bastani and Bayati (2020), Oh et al. (2021), and Li et al. (2022) propose different algorithms based on different regret analyses and assumptions. While our work shares a common theme with these studies, it is widely known that cumulative-regret-minimizing algorithms are not well-suited for best-arm identification problems.

Furthermore, it is worth noting that our model is both related to and distinct from standard contextual bandit models (Auer 2002; Abe et al. 2003). In contextual bandit models, context, or side information, is given in each time step and the payoffs of the systems depend on this information. Our model can be extended to include this contextual information (see Remark 1). However, our paper distinguishes itself from this stream of literature by focusing on selecting the system that is universally the best across various contexts, rather than identifying the best system specific to a particular context.

The remainder of the paper is organized as follows. In Section 2, we formulate the best-arm identification problem based on linear models and LASSO regression. Section 3 discusses the impact of the regularization parameter of the LASSO regression on the performance of identifying the best system. In Section 4, we characterize the decay rate of the probability of false selection and gain geometric insights into the selection of the regularization parameter. Building upon these findings, we introduce a heuristic approach for

sequentially adjusting the regularization parameter and propose a dynamic sampling policy that improves the performance of best-arm identification. Section 5 concludes the paper.

2 PROBLEM FORMULATION

This paper proposes a novel method for addressing the best-arm identification problem with high-dimensional features. To provide clear and transparent analysis, we specifically focus on the case where the payoff of each system can be represented as a linear combination of these features. The objective is to minimize the probability of false selection by leveraging the theory of large deviations coupled with the LASSO estimator. To that end, we first introduce model preliminaries and discuss parametric inference via the LASSO estimator in Sections 2.1 and 2.2, respectively, and then formulate our main problem in Section 2.3.

Before delving into the details, let us briefly introduce the basic notations used in this paper. Let $\mathbb{R}^n$ be the n-dimensional Euclidean space. All vectors are column vectors. For any vector $\mathbf{u} \in \mathbb{R}^n$, u_i and $\mathbf{u}^\top$ denote its i-th component and its transpose, respectively, and we use $\|\mathbf{u}\|_1$ and $\|\mathbf{u}\|_2$ to represent its ℓ_1 and ℓ_2 norms, respectively, i.e., $\|\mathbf{u}\|_1 := \sum_{i=1}^n |u_i|$ and $\|\mathbf{u}\|_2 := \sqrt{\sum_{i=1}^n |u_i|^2}$. The 'hat' notation will be reserved for sample-based estimates. For any condition A, $\mathbb{I}\{A\}$ is an indicator function that equals one if A is true and zero otherwise. For any set Γ, its interior is denoted by Γ°. For any square matrix $\mathbf{M}$, its trace is denoted by $\mathrm{tr}(\mathbf{M})$.

2.1 Model Preliminaries

A decision maker faces K stochastic systems (or arms), indexed by $k = 1, \ldots, K$, where system k is characterized by a deterministic feature vector $\mathbf{x}_k \in \mathbb{R}^d$ (with $K \ll d$) that is known to the decision maker. In stage $t = 1, 2, \ldots$, the decision maker may take a payoff sample from system k, denoted by $y_{k,t}$. The probability distribution of a sample $y_{k,t}$ is determined by the linear model

$$y_{k,t} = \mathbf{x}_k^\top \theta_0 + \varepsilon_{k,t}, \tag{1}$$

where $\varepsilon_{k,t}$ is a random noise that is independent and identically distributed across all t, with $\mathsf{E}[\varepsilon_t] = 0$. It is assumed that $\varepsilon_{k,t}$ and $\varepsilon_{k',t'}$ are independent for any $t \neq t'$ whether or not $k = k'$. The parameter $\theta_0 \in \Theta$ is assumed to be common across the systems and unknown to the decision maker, where Θ is a compact set in $\mathbb{R}^d$. Without loss of generality, we assume that system 1 has the highest mean, i.e.,

$$\mathbf{x}_1^\top \theta_0 > \mathbf{x}_k^\top \theta_0$$

for all $k = 2, \ldots, K$, and is referred to as the best system (or equivalently, the best arm).

Suppose that a fixed sampling budget T is given, indicating that a total of T independent samples can be drawn from the K systems. A sampling policy π is defined as a sequence of random variables, $\pi := \{\pi_1, \pi_2, \ldots, \pi_T\}$, taking values in the index set $\{1, 2, \ldots, K\}$, where $\{\pi_t = k\}$ means a sample from system k is taken in stage t. Let $\mathscr{F}_t$ denote the σ-field generated by the sampling decisions and samples taken up to stage t, i.e., by $\{(\pi_s, y_{\pi_s,s}) : s = 1, \ldots, t\}$, where $\mathscr{F}_0$ is the nominal σ-field associated with the underlying probability space. The set of non-anticipating policies is denoted by Π, where the sampling decision in stage t is determined based on all the sampling decisions and samples observed in previous stages, i.e., $\{\pi_t = k\} \in \mathscr{F}_{t-1}$ for all $k = 1, \ldots, K$ and $t = 1, \ldots, T$. The number of samples drawn from system k up to stage t is defined as $N_{k,t}^\pi := \sum_{s=1}^t \mathbb{I}\{\pi_s = k\}$, and the associated sampling ratio is given by $\alpha_{k,t}^\pi := N_{k,t}^\pi / t$.

2.2 LASSO Estimation

Consider a linear model described in (1). Given a sampling policy π, a standard approach to estimating the mean values of $y_{1,t}, \ldots, y_{K,t}$ based on the sample observations up to stage t is to estimate θ_0 via the

method of least squares, i.e., by solving the following optimization problem

$$\min_{\theta\in\Theta}\left\{\frac{1}{2t}\sum_{s=1}^{t}\left(y_{\pi_s,s}-\mathbf{x}_{\pi_s}^\top\theta\right)^2\right\}.$$

Let $\mathscr{X}_t=(\mathbf{x}_{\pi_1},\mathbf{x}_{\pi_2},\ldots,\mathbf{x}_{\pi_t})^\top$ and $\mathscr{E}_t=(\varepsilon_{\pi_1,1},\varepsilon_{\pi_2,2},\ldots,\varepsilon_{\pi_t,t})^\top$. If the feature dimension is sufficiently small such that $\mathrm{rank}(\mathscr{X}_t)=d\le K$, this problem yields the unique solution $\hat{\theta}_t^\pi=\theta_0+(\mathscr{X}_t^\top\mathscr{X}_t)^{-1}\mathscr{X}_t^\top\mathscr{E}_t$, and the estimates for the mean values $\mathbf{x}_1^\top\theta_0,\ldots,\mathbf{x}_K^\top\theta_0$ of $y_{1,t},\ldots,y_{K,t}$ can be written as $\mathbf{x}_1^\top\hat{\theta}_t^\pi,\ldots,\mathbf{x}_K^\top\hat{\theta}_t^\pi$. It is important to note that in this case, the mean-squared prediction error satisfies the following relationship:

$$\begin{aligned}\mathsf{E}\left[\left\|\mathscr{X}_t\hat{\theta}_t^\pi-\mathscr{X}_t\theta_0\right\|^2\right]&=\mathsf{E}\left[\left\|\mathscr{X}_t(\mathscr{X}_t^\top\mathscr{X}_t)^{-1}\mathscr{X}_t^\top\mathscr{E}_t\right\|^2\right]\\&=\mathsf{E}\left[\mathscr{E}_t^\top\mathscr{X}_t(\mathscr{X}_t^\top\mathscr{X}_t)^{-1}\mathscr{X}_t^\top\mathscr{E}_t\right]\\&=\mathsf{E}\left[\mathrm{tr}\left(\left(\mathscr{X}_t^\top\mathscr{X}_t\right)^{-1}\mathscr{X}_t^\top\mathsf{E}\left[\mathscr{E}_t\mathscr{E}_t^\top\middle|\mathscr{X}_t\right]\mathscr{X}_t\right)\right]\\&\ge d\min_{k=1,\ldots,K}\mathsf{E}\left[\varepsilon_{k,t}^2\right],\end{aligned}$$

where the inequality holds because $\mathsf{E}\left[\mathscr{E}_t\mathscr{E}_t^\top\middle|\mathscr{X}_t\right]$ is a diagonal matrix whose s-th diagonal element is $\mathsf{E}\left[\varepsilon_{\pi_s,s}^2\right]\ge\min_{k=1,\ldots,K}\mathsf{E}\left[\varepsilon_{k,t}^2\right]$. The above relationship implies that the quality of the least-squares estimation degrades as the feature dimension increases. Furthermore, as alluded to earlier, our focus is on practical settings where the features are high-dimensional, i.e., $K\ll d$. In such settings, the least-squares estimator is not even uniquely determined. To address these issues in high-dimensional situations, researchers have explored various approaches to *regularize* the least-squares estimator. As alluded to earlier, we focus on the LASSO estimator for θ_0. This estimator incorporates the ℓ_1-penalty into the least-squares estimator and is defined below:

Definition 1 (LASSO) Given a sampling policy π and a regularization parameter $\lambda\ge 0$, the LASSO estimator based on the sample observations up to stage t is

$$\hat{\theta}_t^\pi(\lambda)\in\operatorname*{argmin}_{\theta\in\Theta}\left\{\frac{1}{2t}\sum_{s=1}^{t}\left(y_{\pi_s,s}-\mathbf{x}_{\pi_s}^\top\theta\right)^2+\lambda\|\theta\|_1\right\}.\tag{2}$$

In the special case of $\lambda=0$, the LASSO estimator $\hat{\theta}_t^\pi(\lambda)$ reduces to the least-squares estimator. As the value of the regularization parameter λ increases, the LASSO regression imposes a stronger penalty on the non-zero components of the parameter θ. This penalty has the effect of enforcing some components in θ to zero, effectively excluding the associated features from the model. Therefore, λ plays a role in controlling the bias-variance trade-off. A small value of λ can result in a high-variance, low-bias model (potentially overfitting), while a large value of λ discourages the learning of a more complex or flexible model, and thus, can lead to a low-variance, high-bias model (potentially underfitting). We further discuss the impact of the regularization parameter on the performance in the best-arm identification problem in Section 3. Given the sample observations up to stage t based on a policy π, the predicted mean values based on the LASSO estimator are denoted by

$$\hat{\mathbf{y}}_t^\pi(\lambda):=\mathbf{X}^\top\hat{\theta}_t^\pi(\lambda)=\left(\mathbf{x}_1^\top\hat{\theta}_t^\pi(\lambda),\ldots,\mathbf{x}_K^\top\hat{\theta}_t^\pi(\lambda)\right)\in\mathbb{R}^K,\tag{3}$$

where $\mathbf{X}:=(\mathbf{x}_1,\ldots,\mathbf{x}_K)\in\mathbb{R}^{d\times K}$ is called the design matrix.

2.3 Main Problem

In this subsection, we formalize the best-arm identification problem that will be investigated throughout the paper. Given a sampling policy π, the decision maker aims to identify the best system with high probability based on the LASSO estimator with a regularization parameter λ. Formally, we define an event $\mathrm{FS}_T^\pi(\lambda) = \{\hat{\mathbf{y}}_T^\pi(\lambda) \in Y_{\mathrm{FS}}\}$, where $\hat{\mathbf{y}}_T^\pi(\lambda)$ is the vector of the predicted means defined in (3) and

$$Y_{\mathrm{FS}} \coloneqq \left\{\mathbf{y} \in \mathbb{R}^K : y_1 \leq y_k \text{ for some } k = 2, \ldots, K\right\}. \tag{4}$$

The probability of falsely selecting a non-best system, denoted by $\mathsf{P}(\mathrm{FS}_T^\pi(\lambda))$, is a widely used criterion for the efficiency of a sampling policy, but its precise evaluation is not analytically tractable as well documented in the literature (Kim and Nelson 2006). Alternatively, our objective is to design a sampling policy π that "asymptotically" minimizes $\mathsf{P}(\mathrm{FS}_T^\pi(\lambda))$ in the large-T regime by leveraging the theory of large deviations. The details of the associated asymptotic analysis will be discussed in Section 4.

Remark 1 In our model in (1), the feature vector $\mathbf{x}_k$ represents system-specific factors (e.g., brand, material, color, size, texture, design, and price for fashion items). In real-world situations, samples may also depend on the sociodemographic features of customers (e.g., income, gender, and education). In this case, one may consider a linear function

$$y_{k,t} = \theta_0^\top \mathbf{x}_k + \beta^\top \mathbf{w}_t + \varepsilon_{k,t}, \tag{5}$$

where $\mathbf{w}_t$ is an i.i.d. random covariate vector representing the features of a customer in stage t. If the mean of $\mathbf{w}_t$ is known, then the best-arm identification problem based on (5) can be equivalently transformed into the problem based on (1) by replacing $\varepsilon_{k,t}$ in (1) with $\tilde{\varepsilon}_{k,t} \coloneqq \beta^\top(\mathbf{w}_t - \mathsf{E}[\mathbf{w}_t]) + \varepsilon_{k,t}$. Hence, in this paper, we restrict our focus to the base setting in Section 2.1.

3 IMPACT OF THE REGULARIZATION PARAMETER

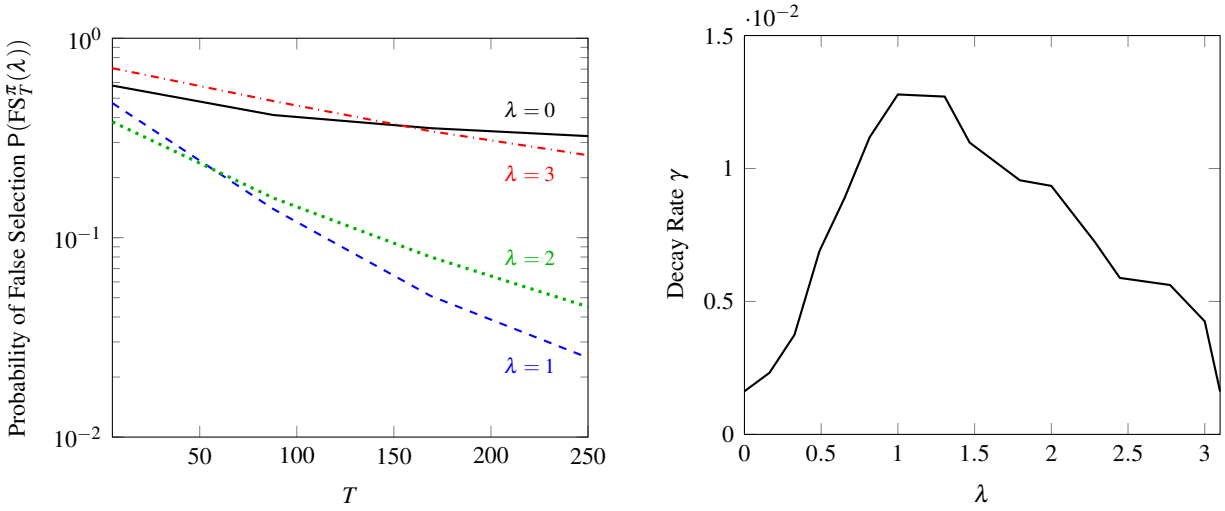
Recall that the penalty term in the LASSO regression decreases the variance because it reduces the essential dimension of the parameter (by forcing some parameters to zero) while introducing a bias compared to the least-squares estimator. In the context of best-arm identification, a lower-variance estimator allows a more precise comparison of the mean values between arms, whereas if the bias is significant, the probability of false selection may not converge to zero even when the sampling budget grows to infinity.

To illustrate the above-mentioned phenomenon, we conduct a numerical experiment with $K = 5$ systems, where each system is characterized by a feature vector of $d = 200$ dimensions. The true parameter vector $\theta_0 \in \mathbb{R}^{200}$ is sparse, with only four non-zero entries, all taking the value of 1. The feature vectors are generated randomly from the standard normal distribution. The expected payoffs of the five systems are given by

$$\mathbf{X}^\top \theta_0 = (2.72, 2.63, -3.30, 1.07, 1.40).$$

The noise terms $\varepsilon_{1,t}, \ldots, \varepsilon_{K,t}$ are assumed to follow the standard normal distribution. The sampling policy π is designated as an equal allocation rule, where each system is sampled with equal probability in every stage, i.e., $\mathsf{P}(\pi_t = 1) = \cdots = \mathsf{P}(\pi_t = K) = 1/K$ for all t.

We estimate the probability of false selection $\mathsf{P}(\mathrm{FS}_T^\pi(\lambda))$ with different values of the sampling budget $T \in [5, 250]$ and the regularization parameter $\lambda \in \{0, 1, 2, 3\}$, using the Monte Carlo simulation with 1,000 trials. The results are illustrated in Figure 1a. When using the least-squares regression without regularization (i.e., $\lambda = 0$), the estimated parameter lies in the vicinity of the true parameter θ_0. However, due to the high variance of the estimator, the probability of false selection converges slowly to zero as the sampling budget grows. On the other hand, for a large regularization parameter ($\lambda = 3$), the LASSO estimator is significantly biased, resulting in a marginal improvement in the performance of identifying the best system, compared to the case of the least-squares regression. The best performance is achieved with a moderate regularization parameter ($\lambda = 1$), which strikes a balance between bias and variance.



(a) The probability of false selection $\mathsf{P}(\mathrm{FS}_T^\pi(\lambda))$ for different values of T and λ.

(b) The estimated decay rate γ of the probability of false selection, where $\gamma = -\lim_{T\to\infty} \log \mathsf{P}(\mathrm{FS}_T^\pi(\lambda))/T$.

Figure 1: Performance of the LASSO estimator and the impact of regularization.

It can be seen from Figure 1a that the probability of false selection converges at an exponential rate, that is, $\mathsf{P}(\mathrm{FS}_T^\pi(\lambda)) \approx C\exp(-\gamma T)$, which implies that $\log \mathsf{P}(\mathrm{FS}_T^\pi(\lambda))$ would be approximately linear in T with slope $-\gamma$. We refer to γ as the decay rate of the probability of false selection, and we plot its estimates under different values of λ in Figure 1b. Specifically, for each value of λ, we first estimate $\log \mathsf{P}(\mathrm{FS}_T^\pi(\lambda))$ via the Monte Carlo simulation with 1,000 trials, where T ranges from 5 to 250. We then use linear regression to estimate the corresponding value of γ. Several observations can be made from Figure 1b. Firstly, there exists a strictly positive regularization parameter λ^* that maximizes the decay rate, which holds true even for a large sample size. This is in stark contrast to the standard practice, where the regularization parameter is typically selected to decrease to zero at the rate of $1/\sqrt{t}$ as the sample size t increases. Secondly, when a regularization parameter is chosen around λ^*, the performance based on the LASSO estimator is significantly superior to that based on the least-squares estimator. Thirdly, for a substantially high value of the regularization parameter ($\lambda > 3$), the performance based on the LASSO estimator may become worse than that based on the least-squares estimator. In the remainder of the paper, we provide a rigorous mathematical analysis that demonstrates these observations.

4 MAIN RESULTS

In this section, we begin by analyzing the asymptotic behavior of the LASSO-based predicted mean values $\hat{\mathbf{y}}_t^\pi(\lambda)$ in (3). This analysis allows us to characterize the decay rate of the probability of false selection. We then explore how the regularization parameter affects the decay rate, providing a guideline for choosing this parameter. Finally, utilizing the insights gained from the convergence rate analysis and the guideline, we design a dynamic sampling policy to achieve optimal performance.

4.1 Large Deviations Principle

Fix a vector α in the interior of the $(K-1)$-simplex $\Delta := \{\alpha \in \mathbb{R}^K : \sum_{k=1}^K \alpha_k = 1, \alpha_k \geq 0\}$ and consider a static sampling policy π^α such that $N_{k,T}^{\pi^\alpha} = \alpha_k T$ for each k, ignoring integrality constraints. We call α the allocation vector as it determines how the sampling budget T is allocated to each system. In this section, we fix $\pi = \pi^\alpha$ and omit α in the superscript for ease of exposition. Denote by $\varphi_k(u) = \log \mathsf{E}[e^{u\varepsilon_{k,t}}]$ the

cumulant generating function of the random variable $\varepsilon_{k,t}$. We define a function $\Lambda : \mathbb{R}^d \to \mathbb{R}$ given by

$$\Lambda(\xi;\alpha) \coloneqq \sum_{k=1}^{K} \alpha_k \varphi_k(\xi^\top \mathbf{x}_k),$$

and we denote its convex conjugate by

$$\Lambda^*(\mathbf{z};\alpha) \coloneqq \sup_{\xi\in\mathbb{R}^d} \left\{ \mathbf{z}^\top \xi - \Lambda(\xi;\alpha) \right\}.$$

Furthermore, for any $\lambda \in \mathbb{R}$ and $\mathbf{z} \in \mathbb{R}^d$, it will be convenient to define the following set:

$$\Theta(\alpha,\lambda,\mathbf{z};\theta_0) \coloneqq \operatorname*{argmin}_{\theta\in\Theta} \left\{ \frac{1}{2} \sum_{i=1}^{K} \alpha_i \|(\theta_0-\theta)^\top \mathbf{x}_i\|_2^2 + \mathbf{z}^\top(\theta_0-\theta) + \lambda\|\theta\|_1 \right\}. \tag{6}$$

The first term in the objective function represents the Kullback-Leibler divergence between the true model (parameterized by θ_0) and an arbitrary model (parameterized by θ). Recalling that minimizing the sum of squared errors is equivalent to minimizing the Kullback-Leibler divergence (Murphy 2022), it can be easily checked that if we let $\mathbf{z} = \hat{\mathbf{z}}_T$ be the aggregated mean-zero noises, i.e.,

$$\hat{\mathbf{z}}_T \coloneqq \frac{1}{T} \sum_{t=1}^{T} \sum_{k=1}^{K} \varepsilon_{k,t} \mathbf{x}_k \mathbb{I}\{\pi_t = k\}, \tag{7}$$

then for fixed $\lambda \geq 0$ and under the said static sampling policy, $\Theta(\alpha,\lambda,\hat{\mathbf{z}}_T;\theta_0)$ is the set of the LASSO estimators $\hat{\theta}_T^\pi(\lambda)$ at the end of the sampling horizon. The following lemma characterizes the large deviations principle for the LASSO estimator $\hat{\theta}_T^\pi(\lambda)$.

Lemma 1 Suppose that $\mathbf{X} = (\mathbf{x}_1,\ldots,\mathbf{x}_K)$ has full column rank. Fix a static sampling policy $\pi = \pi^\alpha$ for some $\alpha \in \Delta^\circ$. Assume that the function $\Lambda(\cdot)$ is finite in an open ball containing the origin. Then, for almost all $\lambda \geq 0$, the predicted mean values $\hat{\mathbf{y}}_T^\pi(\lambda)$ in (3) satisfy a large deviations principle with rate $I(\cdot;\alpha,\lambda)$ given by

$$I(\mathbf{y};\alpha,\lambda) \coloneqq \inf_{\mathbf{z}\in\mathbb{R}^d} \left\{ \Lambda^*(\mathbf{z};\alpha) : \mathbf{X}^\top \hat{\theta} = \mathbf{y},\, \hat{\theta} \in \Theta(\alpha,\lambda,\mathbf{z};\theta_0) \right\}. \tag{8}$$

Recall that $\mathrm{FS}_T^\pi(\lambda) = \{\hat{\mathbf{y}}_T^\pi(\lambda) \in Y_{\mathrm{FS}}\} \subset \mathbb{R}^K$, where Y_{FS} is defined in (4). The large deviations principle for the LASSO-based predicted mean values in Lemma 1 enables us to characterize the asymptotic behavior of the probability of false selection $\mathsf{P}(\mathrm{FS}_T^\pi(\lambda))$:

Theorem 1 Suppose that the assumptions in Lemma 1 hold, and fix a static sampling policy $\pi = \pi^\alpha$ with some $\alpha \in \Delta^\circ$. Let $\gamma(\alpha,\lambda) \coloneqq \inf_{\mathbf{y}\in Y_{\mathrm{FS}}}\{I(\mathbf{y};\alpha,\lambda)\}$ for any $\lambda \geq 0$. Then, the probability of false selection based on the LASSO estimator with a regularization parameter $\lambda \geq 0$ satisfies

$$\lim_{T\to\infty} \frac{1}{T} \log \mathsf{P}\big(\mathrm{FS}_T^\pi(\lambda)\big) = -\gamma(\alpha,\lambda).$$

The preceding theorem confirms our numerical observation in the previous section that $\log \mathsf{P}(\mathrm{FS}_T^\pi(\lambda))$ is asymptotically linear as the sampling budget T grows large. Specifically, the function $\gamma(\alpha,\lambda)$ in Theorem 1 corresponds to the decay rate γ of the probability of false selection discussed in Section 3. Accordingly, we refer to $\gamma(\alpha,\lambda)$ as the rate function for the probability of false selection. For a fixed regularization parameter, the standard approach can be applied to maximize the (large deviations) rate function asymptotically; see, e.g., Chen and Ryzhov (2023) and Ahn et al. (2024). However, as demonstrated in Section 3, the rate function $\gamma(\alpha,\lambda)$ depends crucially on the regularization parameter λ. In the following subsection, we investigate the relationship between the rate function and the regularization parameter for a fixed allocation vector, which leads to a suitable choice of the parameter. This procedure is then leveraged to construct a dynamic sampling policy that is designed to maximize the decay rate $\gamma(\alpha,\lambda)$.

4.2 Choice of Regularization Parameter and the Proposed Sampling Policy

Geometric Insights. To demonstrate the "right" choice of the regularization parameter λ, we consider a simple setting with $K = 2$ systems, where their feature vectors are of dimension $d = 2$ and the true parameter θ_0 are given by:

$$\mathbf{x}_1 = (\cos(\pi/6), \cos(-\pi/3))^\top,\ \mathbf{x}_2 = (\sin(\pi/6), \sin(-\pi/3))^\top,\ \text{and}\ \theta_0 = (1,0)^\top.$$

Thus, the corresponding expected payoffs are $\mathbf{X}^\top\theta_0 = (\cos(\pi/6), \cos(-\pi/3)) \approx (0.866, 0.5)$.

Recall that for fixed $\alpha \in \Delta^\circ$ and $\lambda \geq 0$, the set of LASSO estimators is given by $\Theta(\alpha, \lambda, \hat{\mathbf{z}}_T; \theta_0)$ defined as in (6). Note that the LASSO estimators depend on the random samples only through the *aggregate noise*, $\hat{\mathbf{z}}_T$, defined in (7). In light of this, we transform the problem $\inf_{\mathbf{y}\in Y_{\text{FS}}}\{I(\mathbf{y}; \alpha, \lambda)\}$ in Theorem 1 into an optimization problem with respect to $\mathbf{z}$. Given that $I(\mathbf{y}; \alpha, \lambda)$ is represented by (8), the entire feasible set of the said problem can be expressed equivalently as the *false-selection zone*, denoted by $Z_{\text{FS}}(\alpha, \lambda)$, which is defined as follows:

$$Z_{\text{FS}}(\alpha, \lambda) := \left\{\mathbf{z} \in \mathbb{R}^d : \mathbf{x}_1^\top\hat{\theta} \leq \mathbf{x}_2^\top\hat{\theta} \text{ for some } \hat{\theta} \in \Theta(\alpha, \lambda, \mathbf{z}; \theta_0)\right\}. \tag{9}$$

Observe that $\hat{\mathbf{z}}_T \in Z_{\text{FS}}(\alpha, \lambda)$ is equivalent to the existence of a LASSO estimator $\hat{\theta}_t^\pi(\lambda)$ that leads to a false selection of system 2 as the best system. Accordingly, the rate function $\gamma(\alpha, \lambda)$ for the probability of false selection, characterized in Theorem 1, can be rewritten as follows:

$$\gamma(\alpha, \lambda) = \inf_{\mathbf{z}\in Z_{\text{FS}}(\alpha,\lambda)}\{\Lambda^*(\mathbf{z}; \alpha)\}, \tag{10}$$

and we denote an optimal solution to the right-hand side, if it exists, by $\mathbf{z}^*(\alpha, \lambda)$, i.e.,

$$\mathbf{z}^*(\alpha, \lambda) \in \operatorname*{argmin}_{\mathbf{z}\in Z_{\text{FS}}(\alpha,\lambda)}\{\Lambda^*(\mathbf{z}; \alpha)\}. \tag{11}$$

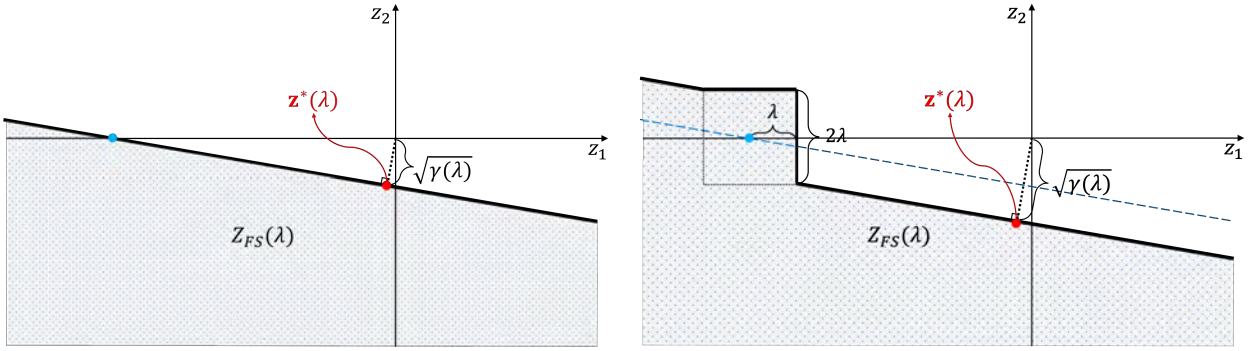
To provide a clearer description of the function $\Lambda^*(\mathbf{z}; \alpha)$ in (11), we set $\alpha_1 = \alpha_2 = 1/2$ and assume that the noise terms follow the standard normal distribution, i.e., $\varepsilon_{1,t}, \varepsilon_{2,t} \sim N(0,1)$. Then, $\Lambda^*(\mathbf{z}; \alpha)$ is given by

$$\Lambda^*(\mathbf{z}; \alpha) = \frac{1}{2}\mathbf{z}^\top\Sigma\left(\alpha; \sigma^2\right)^{-1}\mathbf{z} = z_1^2 + z_2^2, \tag{12}$$

where $\Sigma(\alpha; \sigma^2) := \sum_{k=1}^K \alpha_k\sigma_k^2\mathbf{x}_k\mathbf{x}_k^\top$ and σ_k^2 is the variance of $y_{k,t}$ for $k = 1, \ldots, K$.

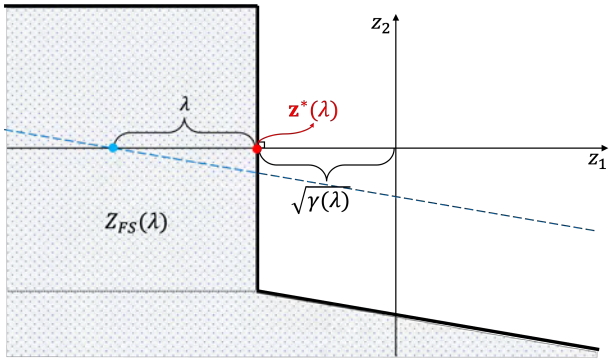
In the following discussion, we fix the allocation vector α and often omit the argument α in $\gamma(\alpha, \lambda)$, $Z_{\text{FS}}(\alpha, \lambda)$, and $\mathbf{z}^*(\alpha, \lambda)$ for simple exposition. Note that the rate function for the probability of false selection is $\gamma(\lambda) = \Lambda^*(\mathbf{z}^*(\lambda))$, and thus, the distance between the origin and the false-selection zone $Z_{\text{FS}}(\lambda)$ is equal to $\sqrt{\gamma(\lambda)}$. This implies that the farther the false-selection zone $Z_{\text{FS}}(\lambda)$ is away from the origin, the faster the probability of false selection decays to zero as the sampling budget increases to infinity.

Figure 2 illustrates the false-selection zone $Z_{\text{FS}}(\lambda)$ and the optimal solution $\mathbf{z}^*(\lambda)$ for different values of the regularization parameter λ. Figure 2a corresponds to the benchmark case of the least-squares estimator, where $\lambda = 0$. For small $\lambda > 0$, the false-selection zone is characterized as in Figure 2b, where the square region with width λ pertains to the regularization term $\lambda\|\theta\|_1$ in (2). In this case, we obtain a higher decay rate due to the increased distance compared to the benchmark case. If we further increase λ, however, Figure 2c shows that the location of $\mathbf{z}^*(\lambda)$ shifts from the lower sloping boundary of the zone to the right side of the square. In this regime, a larger regularization parameter λ brings the false-selection zone closer to the origin, resulting in a slower decay rate of the probability of false selection. If the regularization parameter λ is too high, the LASSO estimator may perform worse than the least-squares estimator, as described in Figure 2d.

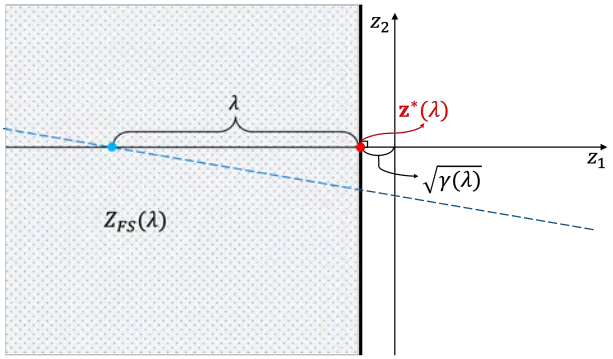


(a) Least-squares regression ($\lambda = 0$)

(b) LASSO regression with small λ



(c) LASSO regression with medium λ



(d) LASSO regression with large λ

Figure 2: A schematic illustration of the false-selection zone $Z_{\text{FS}}(\lambda)$ and the decay rate $\gamma(\lambda)$ of the probability of false selection with different values of the regularization parameter $\lambda \geq 0$.

It is worth emphasizing that the blue dashed lines in Figures 2b to 2d correspond to the hyperplane that characterizes the false-selection zone when using the least-squares estimator, which is illustrated in Figure 2a. This hyperplane has a normal vector of $\mathbf{z}^*(0)$ and is parallel to the lower sloping boundaries in Figures 2b and 2c. From this observation, it becomes evident that for any given $\lambda \geq 0$, if the two vectors $\mathbf{z}^*(\lambda)$ and $\mathbf{z}^*(0)$ are parallel (as shown in Figures 2a and 2b), then increasing λ would lead to a higher decay rate. Conversely, when they are not parallel (as depicted in Figures 2c and 2d), decreasing λ would result in improved performance. Therefore, the optimal choice of the regularization parameter can be determined by comparing $\mathbf{z}^*(\lambda)$ and $\mathbf{z}^*(0)$. We highlight that this observation holds for any normally distributed noise terms with high-dimensional features, not limited to standard normality with two-dimensional features, because the objective function $\Lambda^*(\cdot;\alpha)$ in (11) remains a quadratic function centered at the origin and the associated feasible region $Z_{\text{FS}}(\alpha,\lambda)$ does not depend on the distribution of the noise terms. We defer a rigorous discussion on this generalization for further research.

Choosing the Regularization Parameter. Based on the aforementioned observation, we provide a guideline for choosing the regularization parameter λ that can lead to (near-)optimal performance under the normality assumption. In particular, for a fixed step size $\delta > 0$ small enough and for any given $\lambda \geq 0$ and $\alpha \in \Delta^\circ$, we consider a simple updating rule, denoted by $\mathscr{G}(\lambda, \delta, Z_{\text{FS}}(\alpha,\cdot), \Lambda^*(\cdot;\alpha))$:

1. Solve (11) to obtain $\mathbf{z}^*(\alpha,\lambda)$ and $\mathbf{z}^*(\alpha,0)$.
2. If $\mathbf{z}^*(\alpha,\lambda)$ and $\mathbf{z}^*(\alpha,0)$ are parallel, we set $\lambda \leftarrow \lambda + \delta$.
3. Otherwise, we set $\lambda \leftarrow \lambda - \delta$.

Algorithm 1 Dynamic sampling policy $\pi(n_0, m, \delta)$

Take n_0 payoff samples from each system
Set $t = Kn_0$, $\alpha_t^\pi = (1/K, \ldots, 1/K)^\top$, and $\lambda = 0$
while $t < T$ **do**
  Find the LASSO estimator $\hat{\theta}_t^\pi$ and the sample variance vector $\hat{\sigma}_t^2$
  Construct $\hat{Z}_{\mathrm{FS},t}(\alpha_t^\pi, \cdot)$ and $\hat{\Lambda}_t^*(\cdot; \alpha_t^\pi)$ by setting $\theta_0 = \hat{\theta}_t^\pi$ in (9) and $\sigma^2 = \hat{\sigma}_t^2$ in (12), respectively
  Update λ via the updating rule $\mathscr{G}(\lambda, \delta, \hat{Z}_{\mathrm{FS},t}(\alpha_t^\pi, \cdot), \hat{\Lambda}_t^*(\cdot; \alpha_t^\pi))$
  Compute $\hat{\alpha}_t = \operatorname{argmax}_{\alpha \in \Delta} \hat{\gamma}_t(\alpha, \lambda)$, where $\hat{\gamma}_t(\alpha, \lambda)$ is defined in (10) with $Z_{\mathrm{FS}}(\alpha, \lambda) = \hat{Z}_{\mathrm{FS},t}(\alpha, \lambda)$ and $\Lambda^*(\cdot; \alpha) = \hat{\Lambda}_t^*(\cdot; \alpha)$
  Generate a sample vector $(n_1, \ldots, n_K)$ from the multinomial distribution with parameters $(m, \hat{\alpha}_t)$
  Take n_k samples from each system k, and set $t = t + m$
end while

Proposed Sampling Policy. The preceding updating rule prescribes a suitable choice of the regularization parameter λ for a given allocation vector α. However, to maximize the decay rate $\gamma(\alpha, \lambda)$, it is essential to optimize the allocation vector α as well as the associated regularization parameter λ. The optimal solution is not known a priori because the true parameter θ_0 is unknown. However, building on our findings from the static policy, we propose a dynamic sampling policy that adjusts the sampling ratio α_t^π and the associated regularization parameter in a way that approaches the optimal solution as t increases. In each stage, this policy sequentially

1. estimates sample counterparts of the false-selection zone $Z_{\mathrm{FS}}(\alpha, \lambda)$ and the convex conjugate function $\Lambda^*(\mathbf{z}; \alpha)$ using all sample observations up to that stage;
2. updates the regularization parameter via the above-mentioned guideline using these estimates;
3. constructs a "plug-in" estimator of the decay rate $\gamma(\alpha, \lambda)$;
4. estimates the optimal allocation by maximizing the plug-in estimator with respect to α; and
5. determines the sample allocation in the next stage using the estimated optimal allocation.

The detailed procedure is described in Algorithm 1. Note that solving a bilevel optimization problem in (10) could be computationally costly, particularly when both K and d are large. To address this issue, this algorithm utilizes batch-based sampling, with two parameters n_0 and m representing the initial sample size for each system and the batch size, respectively. This approach helps reduce the number of times the bilevel optimization problem needs to be solved. For the sake of clarity, we assume that m is a multiple of $T - Kn_0$. The parameter δ in the algorithm corresponds to the step size specified in the guideline for updating the regularization parameter.

We note that the sampling policy in Algorithm 1 aims to present our fundamental ideas in a constructive manner, rather than providing an algorithm with optimal performance. A conceptually simple extension involves making the step size δ a decreasing (possibly randomized) function δ_t based on the number of samples, t. The development of a more refined algorithm with theoretical performance guarantees and the associated numerical experiments are left for future research.

5 CONCLUDING REMARKS

This paper investigates the problem of best-arm identification when the underlying random variables are characterized by a linear model with high-dimensional features. When the best system is selected based on LASSO, we analyze the decay rate of the probability of false selection as the sampling budget grows. Our analysis reveals that utilizing the LASSO estimator, with an appropriate choice of the regularization parameter, can significantly improve the performance of identifying the best system compared to the case of using the ordinary least-squares estimator. Furthermore, building on geometric insights into the impact of the regularization parameter on overall performance, we develop a tractable guideline for selecting this

parameter. Finally, we propose a dynamic sampling policy that integrates this guideline and allows for an effective allocation of the sampling budget to maximize the aforementioned decay rate.

This work opens up several potential directions for future research. Firstly, our analysis in Section 4.2 is conducted with the aim of gaining insights, and for this purpose, we assume that noise terms follow a normal distribution. While this assumption is often made in both practice and theoretical contexts, it would be valuable to extend the analysis to encompass cases with non-normal noise terms. In such cases, the closed form of the function Λ^* is not available and must be estimated via simulation. Thus, it is essential to explore the impact of the corresponding estimation error on the overall performance of the proposed algorithm. This extension would establish a more general guideline for utilizing LASSO in best-arm identification problems. Secondly, one might incorporate other regularized linear regression models, such as ridge regression, into the problem of best-arm identification and conduct a comparative analysis of the performance across different regression models.

ACKNOWLEDGMENTS

D. Ahn acknowledges the support of the Early Career Scheme from the Research Grants Council of Hong Kong (Project No. 24210420) and CUHK Direct Grant for Research (Project No. 4055206). D. Shin acknowledges the support of the General Research Fund from the Research Grants Council of Hong Kong (Project No. 16501821).

REFERENCES

Abbasi-Yadkori, Y., D. Pál, and C. Szepesvári. 2011. "Improved Algorithms for Linear Stochastic Bandits". In *Advances in Neural Information Processing Systems*, edited by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, 2312–2320. New York: Curran Associates, Inc.

Abbasi-Yadkori, Y., D. Pal, and C. Szepesvari. 2012. "Online-to-Confidence-Set Conversions and Application to Sparse Stochastic Bandits". In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, edited by N. D. Lawrence and M. Girolami, 1–9. Cambridge, Massachusetts: Proceedings of Machine Learning Research.

Abe, N., A. W. Biermann, and P. M. Long. 2003. "Reinforcement Learning with Immediate Rewards and Linear Hypotheses". *Algorithmica* 37:263–293.

Ahn, D. and D. Shin. 2020. "Ordinal Optimization with Generalized Linear Model". In *2020 Winter Simulation Conference (WSC)*, 3008–3019 https://doi.org/https://doi.org/10.1109/WSC48552.2020.9384070.

Ahn, D., D. Shin, and A. Zeevi. 2024. "Feature Misspecification in Sequential Learning Problems". *Management Science*. Forthcoming. Available at https://papers.ssrn.com/abstract_id=3860650.

Auer, P. 2002. "Using Confidence Bounds for Exploitation-Exploration Trade-offs". *Journal of Machine Learning Research* 3(Nov):397–422.

Auer, P., N. Cesa-Bianchi, and P. Fischer. 2002. "Finite-Time Analysis of the Multiarmed Bandit Problem". *Machine Learning* 47:235–256.

Bastani, H. and M. Bayati. 2020. "Online Decision Making with High-Dimensional Covariates". *Operations Research* 68(1):276–294 https://doi.org/10.1287/opre.2019.1902.

Branke, J., S. E. Chick, and C. Schmidt. 2007. "Selecting a Selection Procedure". *Management Science* 53(12):1916–1932.

Camilleri, R., K. Jamieson, and J. Katz-Samuels. 2021. "High-dimensional Experimental Design and Kernel Bandits". In *Proceedings of the 38th International Conference on Machine Learning*, edited by M. Meila and T. Zhang, 1227–1237. Cambridge, Massachusetts: Proceedings of Machine Learning Research.

Carpentier, A. and R. Munos. 2012. "Bandit Theory Meets Compressed Sensing for High Dimensional Stochastic Linear Bandit". In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, edited by N. D. Lawrence and M. Girolami, 190–198. Cambridge, Massachusetts: Proceedings of Machine Learning Research.

Chen, Y. and I. O. Ryzhov. 2023. "Balancing Optimal Large Deviations in Sequential Selection". *Management Science* 69(6):3457–3473.

Chu, W., L. Li, L. Reyzin, and R. Schapire. 2011. "Contextual Bandits with Linear Payoff Functions". In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, edited by G. Gordon, D. Dunson, and M. Dudík, 208–214. Cambridge, Massachusetts: Proceedings of Machine Learning Research.

Dani, V., T. P. Hayes, and S. M. Kakade. 2008. "Stochastic Linear Optimization under Bandit Feedback". In *Proceedings of the 21st Annual Conference on Learning Theory*, 355–366. July 9th-12th, Helsinki, Finland.

Dembo, A. and O. Zeitouni. 2009. *Large Deviations Techniques and Applications*. New York: Springer Science & Business Media.

Ho, Y.-C., Q.-C. Zhao, and Q.-S. Jia. 2008. *Ordinal Optimization: Soft Optimization for Hard Problems*. New York: Springer Science & Business Media.

Kazerouni, A. and L. M. Wein. 2021. "Best Arm Identification in Generalized Linear Bandits". *Operations Research Letters* 49(3):365–371.

Kim, G.-S. and M. C. Paik. 2019. "Doubly-Robust Lasso Bandit". In *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett. New York: Curran Associates, Inc.

Kim, S. H. and B. L. Nelson. 2006. "Selecting the Best System". In *Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. L. Nelson, Volume 13, Chapter 17, 501–534. Boston: Elsevier.

Li, W., A. Barik, and J. Honorio. 2022. "A Simple Unified Framework for High Dimensional Bandit Problems". In *Proceedings of the 39th International Conference on Machine Learning*, edited by K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, 12619–12655. Cambridge, Massachusetts: Proceedings of Machine Learning Research.

Murphy, K. P. 2022. *Probabilistic Machine Learning: An Introduction*. MIT press.

Oh, M.-H., G. Iyengar, and A. Zeevi. 2021. "Sparsity-Agnostic Lasso Bandit". In *Proceedings of the 38th International Conference on Machine Learning*, edited by M. Meila and T. Zhang, Volume 139, 8271–8280. Cambridge, Massachusetts: Proceedings of Machine Learning Research.

Rusmevichientong, P. and J. N. Tsitsiklis. 2010. "Linearly Parameterized Bandits". *Mathematics of Operations Research* 35(2):395–411.

Soare, M., A. Lazaric, and R. Munos. 2014. "Best-Arm Identification in Linear Bandits". In *Advances in Neural Information Processing Systems*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Volume 27, 828–836. Cambridge, Massachusetts: MIT Press.

Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso". *Journal of the Royal Statistical Society. Series B (Methodological)* 58(1):267–288.

Xu, L., J. Honda, and M. Sugiyama. 2018. "A Fully Adaptive Algorithm for Pure Exploration in Linear Bandits". In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, edited by A. Storkey and F. Perez-Cruz, 843–851. Cambridge, Massachusetts: Proceedings of Machine Learning Research.

Zhu, Y., D. Zhou, R. Jiang, Q. Gu, R. Willett and R. Nowak. 2021. "Pure Exploration in Kernel and Neural Bandits". In *Advances in Neural Information Processing Systems*, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Volume 34, 11618–11630. New York: Curran Associates, Inc.

AUTHOR BIOGRAPHIES

DOHYUN AHN is an Assistant Professor in the Department of Systems Engineering and Engineering Management at the Chinese University of Hong Kong. He received his B.S., M.S., and Ph.D. degrees in Industrial & Systems Engineering, all from KAIST. His research interests include, but are not limited to, quantitative risk management, Monte Carlo simulation methodologies, decision making under uncertainty, and network analysis in finance and operations. His email address is dohyun.ahn@cuhk.edu.hk and his website is https://sites.google.com/view/dohyun/.

DONGWOOK SHIN is an Assistant Professor in the Department of Information Systems, Business Statistics, and Operations Management at the HKUST Business School. He received his Ph.D. degree from Columbia Business School. His research centers on optimal learning theory at the interface between machine learning and sequential decision making, with applications to problems arising in the context of operations management. His email address is dwshin@ust.hk and his website is http://dwshin.people.ust.hk/.

LEWEN ZHENG is a Ph.D. candidate in the Department of Systems Engineering and Engineering Management at the Chinese University of Hong Kong. He received his B.S. degree from the School of Mathematics at Sun Yat-sen University in Guangzhou, China. His research interests include simulation methods and their applications to finance. His email address is lwzheng@se.cuhk.edu.hk and his website is https://sites.google.com/view/lewenzheng/.