Semi-Parametric Batched Global Multi-Armed Bandits with Covariates

Sakshi Arya[*]

*Department of Mathematics, Applied Mathematics, & Statistics,
Case Western Reserve University*

Hyebin Song[†]

Department of Statistics, Pennsylvania State University

March 1, 2025

Abstract

The multi-armed bandits (MAB) framework is a widely used approach for sequential decision-making, where a decision-maker selects an arm in each round with the goal of maximizing long-term rewards. Moreover, in many practical applications, such as personalized medicine and recommendation systems, feedback is provided in batches, contextual information is available at the time of decision-making, and rewards from different arms are related rather than independent. We propose a novel semi-parametric framework for batched bandits with covariates and a shared parameter across arms, leveraging the single-index regression (SIR) model to capture relationships between arm rewards while balancing interpretability and flexibility. Our algorithm, Batched single-Index Dynamic binning and Successive arm elimination (BIDS), employs a batched successive arm elimination strategy with a dynamic binning mechanism guided by the single-index direction. We consider two settings: one where a pilot direction is available and another where the direction is estimated from data, deriving theoretical regret bounds for both cases. When a pilot direction is available with sufficient accuracy, our approach achieves minimax-optimal rates (with $d = 1$) for nonparametric batched bandits, circumventing the curse of dimensionality. Extensive experiments on simulated and real-world datasets demonstrate the effectiveness of our algorithm compared to the nonparametric batched bandit method introduced by Jiang and Ma [2024].

Keywords: single-index regression, contextual bandits, batches, dynamic binning, regret bounds

---

[*]Both authors contributed equally.
[†]Corresponding author: hps5320@psu.edu

1 Introduction

Decision-making is integral to domains such as healthcare, agriculture, industry, and government. In today's data-rich environment, the availability of abundant and continuously growing information empowers us to refine decisions dynamically, even tailoring them to individual contexts. This evolving process, known as sequential decision-making, is the bedrock of data-driven informed choices across diverse sectors.

A foundational framework for sequential decision-making is the multi-armed bandit problem [Lai and Robbins, 1985, Lai, 1987], which seeks to optimize the selection of actions (or arms) to maximize cumulative rewards over time. In this framework, a learner sequentially selects actions and observes their corresponding rewards. In many applications, additional contextual information, or covariates, can significantly enhance decision-making. Incorporating these covariates extends the framework to contextual bandits or multi-armed bandits with covariates (MABC) [Perchet and Rigollet, 2013, Yang and Zhu, 2002].

Standard MABC approaches often assume that arms are independent, which limits their applicability in scenarios where playing one arm reveals insights about others, particularly for similar covariates. This shared informativeness is especially relevant in applications such as clinical trials and personalized recommendations. For example, in clinical trials, treatments with similar chemical compositions are likely to exhibit analogous effects on patients with comparable profiles, such as those in the same age group or with similar disease severity. Similarly, in news recommendation systems, users with similar demographic characteristics, such as age or occupation, often display comparable preferences.

To address this limitation, the Global Multi-Armed Bandits (GMAB) model framework was recently proposed, where arms share a global parameter, making them globally informative [Atan et al., 2015, 2018, Shen et al., 2018]. More specifically, in the GMAB model framework, it is assumed that the expected reward from each arm is a known function of a single global parameter. While this proposal provides an effective framework for leveraging shared information between arms, there are two important limitations: 1. The mean reward functions are assumed to be known, and 2. It cannot incorporate contextual information, as the function is the same for all observations. To address these challenges, we introduce Global Multi-Armed Bandits with Covariates (GMABC), a novel framework

that generalizes the Global Multi-Armed Bandits (GMAB) model by integrating covariates and also allowing the functions to be unknown. In GMABC, we propose that arms are interconnected through a shared global parameter, but the functions linking the global parameter to the rewards can depend on the covariates. Moreover, we do not assume that these functions are known a priori.

In the MABC framework, the relationship between rewards and covariates is typically modeled using regression methods, which can be broadly classified as parametric [Goldenshluger and Zeevi, 2013, Filippi et al., 2010, Chu et al., 2011, Abbasi-Yadkori et al., 2011, Agrawal and Goyal, 2013] or non-parametric [Rigollet and Zeevi, 2010, Wanigasekara and Yu, 2019, Arya and Sriperumbudur, 2023]. Parametric methods assume a predefined relationship (such as linear or generalized linear models), offering interpretability and efficiency when correctly specified, but they can perform poorly under model misspecification. Non-parametric methods, on the other hand, are more flexible and can capture complex relationships but face challenges in high-dimensional contexts, including computational demands and reduced interpretability.

To balance interpretability and flexibility, we adopt a semi-parametric approach using the single-index model [Li and Duan, 1989, Ichimura, 1993, Hardle et al., 1993, Kuchibhotla and Patra, 2020, Dai et al., 2022], where the mean reward depends on a one-dimensional projection of the covariates. This model effectively mitigates the curse of dimensionality while retaining interpretability through the projection coefficients, also known as the *index*. This setting extends generalized linear models (GLMs) by allowing the link function to remain unknown, unlike in standard GLMs where it is predefined. Within the GMABC framework, we assume a shared index parameter across arms, motivated by real-world scenarios such as drug dosage optimization and dynamic pricing, where rewards are correlated for similar covariates.

Another practical consideration is that data in many real-world applications are collected in batches rather than sequentially. For example, clinical trials often proceed in phases, where treatments are allocated for an entire batch and outcomes are analyzed collectively before updating the decision policy. Batched bandits with both fixed and adaptive batch sizes have been studied extensively in the literature [Perchet et al., 2016, Esfandiari et al., 2021, Kalkanli and Ozgur, 2021, Jin et al., 2021]. Theoretical work on

batched bandits has provided regret guarantees for both parametric [Han et al., 2020, Ren et al., 2022] and nonparametric frameworks [Gu et al., 2024, Jiang and Ma, 2024, Feng et al., 2022], highlighting the relevance and challenges in scenarios with a small number of batches ($M \approx 2, 3, 4, 5$), as often seen in clinical trials.

In this paper, we explore the batched GMABC problem within a semi-parametric framework, aiming to achieve a "best-of-both-worlds" approach to sequential decision-making. Our methodological contributions are twofold: (1) leveraging arm dependence through a global parameter: we propose a novel global MABC framework tailored for the batched bandit setting, and (2) semi-parametric regression framework: we develop and analyze a single-index regression framework that enhances the interpretability of parametric models while preserving the flexibility of non-parametric approaches.

We propose an algorithm, Batched Index based Dynamic binning and Successive arm elimination (BIDS), based on dynamic binning and successive arm elimination in the batched GMABC framework. We provide regret guarantees for the proposed algorithm in two settings: 1) pilot estimate available: when a reliable pilot estimate of the single-index direction is available from prior studies. 2) no prior information on the single-index direction is available: this requires the estimation of the index vector. We explore the trade-offs in jointly estimating the projection vector and the link function in the absence of prior information. Our contributions include regret bounds for both scenarios and insights into the practical implications of incorporating covariates and batch constraints in MABC under a single-index framework with a global index parameter.

Related literature A related line of work is the *semi-parametric bandits* framework [Greenewald et al., 2017, Krishnamurthy et al., 2018, Kim and Paik, 2019], which differs from our approach in its underlying model. Semi-parametric bandits represent the mean reward function as the sum of a linear function of the arm with a shared parameter and an additive non-linear perturbation that is independent of the action/arm. In contrast, our model follows a single-index approach, making it "semi-parametric" from a regression perspective within the MABC framework. While this line of research assumes a linear treatment effect, our model allows for non-linear treatment effects through unknown link functions specific to each arm. Despite both frameworks aiming to balance the simplicity of

parametric models with the flexibility of nonparametric models, the fundamental modeling assumptions differ. In the semi-parametric bandits literature, the non-linear term is treated as a confounder, necessitating new linear model estimators to handle confounding effects. In contrast, our approach leverages methodology from single-index regression to estimate the shared global parameter. Another relevant theme is dimension reduction in the MABC framework. For instance, Qian and Yang [2016] propose a sufficient dimension reduction framework in MABC, though without theoretical guarantees. Later, Bastani and Bayati [2020] introduce a LASSO bandit for high-dimensional covariates. Other works exploring dimension reduction in contextual bandits include Li et al. [2021, 2022], Qian et al. [2024] and Li et al. [2023], where the first three concentrate on high-dimensional linear bandits, while the last one adopts a localized LASSO approach for nonparametric bandits.

This paper is organized as follows: Section 2 introduces the problem setup for GMABC with a single-index regression framework and outlines the associated model assumptions. In Section 3, we present the Batched Single Index Dynamic Binning and Successive Arm Elimination (BIDS) algorithm, which integrates single-index guided dynamic binning and successive arm elimination as its core components. Sections 3.1 and 3.2 detail the methodology for two distinct scenarios: when a pilot estimate is available and when it is not. Section 4 provides the regret analysis for these two scenarios, with Sections 4.1 and 4.2 addressing the regret rate for both the cases of known and unknown pilot estimates, respectively. The performance of the proposed methodology is demonstrated through both simulated and real data in Sections 5 and 6. Finally, we conclude with a discussion in Section 7.

2   Problem Setup

We begin by presenting the problem setup for the *batched global multi-armed bandit with covariates (GMABC)* problem that we will be working with hereafter. We assume that we have d-dimensional covariates $X_1, X_2, \ldots$ such that $X_t \sim \mathbb{P}_X$ i.i.d. for $t = 1, \ldots, T$. For simplicity of exposition, we focus on the two-arm setting where we select an arm $k \in \{1, \ldots, K\}$ with $K = 2$; however, the generalization to a $K > 2$ arm setting is straightforward.

The model for rewards for each arm $k \in \{1, 2\}$ is given by:

$$Y_t^{(k)} = g^{(k)}(X_t) + \epsilon_t \tag{1}$$

for $t = 1, \ldots, T$, where $g^{(k)} : \mathbb{R}^d \to \mathbb{R}$ are the mean reward functions, and $\{\epsilon_t\}_{t \geq 0}$ is a sequence of independent mean zero random variables. Furthermore, we assume the following single index model structure for $g^{(k)}$:

$$g^{(k)}(x) = f^{(k)}(x^\top \beta_0) \tag{2}$$

for $k = 1, 2$, where $f^{(k)} : \mathbb{R} \to \mathbb{R}$ are 1-dimensional *link functions* and $\beta_0 \in \mathbb{R}^d$ is the unknown *index parameter or direction* shared by both arms. Throughout the paper, we assume $\|\beta_0\|_2 = 1$ for the identifiability of the parameter. Model (1) with (2) is what we refer to as the GMABC regression framework for the underlying sequential decision-making problem.

A *policy* $\pi_t : \mathcal{X} \to \{1, 2\}$ for $t = 1, \ldots, T$ determines an action $A_t \in \{1, 2\}$ at t. Based on the chosen action A_t, a reward $Y_t^{(A_t)}$ is obtained. In the sequential setting without batch constraints, the policy π_t can depend on all the observations $(X_s, Y_s^{(A_s)})$ for $s < t$. In contrast, in a batched setting with M batches, where $0 = t_0 < t_1 < \cdots < t_{M-1} < t_M = T$, for $t \in [t_i, t_{i+1})$, the policy π_t can depend on observations from the previous batches, but not on any observations within the same batch. In other words, policy updates can occur only at the predetermined batch boundaries $t_1, \ldots, t_M$.

Let $\mathcal{G} = \{t_0, t_1, \ldots, t_M\}$ represent a partition of time $\{0, 1, \ldots, T\}$ into M intervals, and $\pi = (\pi_t)_{t=1}^T$ be the sequence of policies applied at each time step. The overarching objective of the decision-maker is to devise an M-batch policy $(\mathcal{G}, \pi)$ that minimizes the expected *cumulative regret*, defined as $\mathcal{R}_T(\pi) = E[R_T(\pi)]$, where

$$R_T(\pi) = \sum_{t=1}^T g^{(*)}(X_t) - g^{(\pi_t(X_t))}(X_t) = \sum_{t=1}^T f^{(*)}(X_t^\top \beta_0) - f^{(\pi_t(X_t))}(X_t^\top \beta_0) \tag{3}$$

where $g^{(*)}(x) = \max_{k \in \{1, 2\}} g^{(k)}(x)$ is the expected reward from the optimal choice of arms given a context x. The cumulative regret serves as a pivotal metric, quantifying the differ-

ence between the cumulative reward attained by π and that achieved by an optimal policy, assuming perfect foreknowledge of the optimal action at each time step.

We make the following assumptions on the reward functions.

Assumption 1 (Smoothness)**.** We assume that the link function $f^{(k)} : \mathbb{R} \to \mathbb{R}$ for each arm is (η, L)-smooth, that is, there exists $\eta \in (0, 1]$ and $L > 0$ such that for $k \in \{1, 2\}$,

$$|f^{(k)}(u) - f^{(k)}(u')| \leq L|u - u'|^{\eta},$$

holds for $u, u' \in \mathbb{R}$.

Assumption 2 (Margin)**.** Reward functions satisfy the margin condition with parameter $\alpha > 0$, that is, there exists $\delta_0 \in (0, 1)$ and $D_0 > 0$ such that

$$\mathbb{P}_X(0 < |f^{(1)}(X^\top \beta_0) - f^{(2)}(X^\top \beta_0)| \leq \delta) \leq D_0 \delta^{\alpha},$$

holds for all $\delta \in [0, \delta_0]$.

Remark 1. The margin parameter measures the complexity of the problem. A small α means that the two functions are quite close to each other in many regions. Throughout this paper, we assume that $\alpha\eta \leq 1$, because in the $\alpha\eta > 1$ regime, the context information becomes irrelevant as one arm dominates the other (e.g., see Perchet and Rigollet [2013]).

Define $\mathbb{B}_2(r; c) = \{v \in \mathbb{R}^d; \|v - c\|_2 \leq r\}$ denote the ℓ_2 ball of radius r centered at c. The next Assumption 3 concerns the required conditions for the distribution of the reward $Y^{(k)}$ and covariate X:

Assumption 3. The reward $Y_t^{(k)}$ satisfies $|Y_t^{(k)}| \leq 0.5$ for all $t = 1, \ldots, T$, $k \in \{1, 2\}$. The probability measure $\mathbb{P}_X$ is absolutely continuous with respect to the Lebesgue measure, and its support set $\text{Supp}(\mathbb{P}_X)$ is bounded, i.e., there exists $R_X < \infty$ such that $\text{Supp}(\mathbb{P}_X) \subseteq \mathbb{B}_2(R_X; 0)$. Moreover, there exists $R_0 > 0$ such that for any $v \in \mathbb{B}_2(R_0; \beta_0)$ and $\|v\|_2 = 1$, $\mathbb{P}_{X^\top v}$ is supported on an interval $\mathcal{I}_v \subseteq \mathbb{R}$, and the density function $f_{X^\top v}$ on $\mathcal{I}_v$ is bounded above and below by some constants $\bar{c}_X > 0$ and $\underline{c}_X > 0$ independent of v.

The boundedness assumption for rewards is made for technical reasons to apply concentration bounds. The constant 0.5 is chosen for simplicity of exposition, but can easily

be replaced with other (large) constants. For the distribution $\mathbb{P}_X$ of X, we assume that $\mathbb{P}_X$ has a density, its support is bounded in $\mathbb{R}^d$, and the density of the projection of X onto a direction near β_0 is non-vanishing and supported on an interval in $\mathbb{R}$. Essentially, the last condition allows us to obtain information on $f^{(k)}$ from all regions given a sufficiently accurate working direction. Similar assumptions have been made in other non-parametric bandit settings for $\mathbb{P}_X$ [Perchet and Rigollet, 2013, Jiang, 2019], where the covariate distribution $\mathbb{P}_X$ is supported on a hypercube and its density does not vanish within that hypercube.

To provide a concrete illustration of $\mathbb{P}_X$ satisfying Assumption 3, consider X following a truncated multivariate normal distribution $N(\mathbf{0}, \Sigma)$ constrained within a unit hypercube $\mathcal{H} = \prod_{j=1}^{d} 1\{|x_j| \leq 0.5\}$, i.e., whose density is proportional to $\exp(-\frac{1}{2}x^\top \Sigma^{-1}x)1\{x \in \mathcal{H}\}$. We can find $R_0, \bar{c}_X$, and $\underline{c}_X$ that satisfy Assumption 3. See Lemma 1 for details. The proof for the Lemma is provided in Appendix S1.

Lemma 1. *Suppose $X \sim N_T(0, \Sigma)$ whose density is given by*

$$
f_X(x) = \begin{cases} \frac{1}{Z(\Sigma)} \exp\{-\frac{1}{2}x^\top \Sigma^{-1}x\} & x \in \mathcal{H} \\ 0 & otherwise \end{cases}
$$

with $Z(\Sigma) = \int_{x \in \mathbb{R}^d} e^{-\frac{1}{2}x^\top \Sigma^{-1}x}1\{x \in \mathcal{H}\}dx$ where $\mathcal{H} = \prod_{j=1}^{d} 1\{|x_j| \leq 0.5\}$. Then we can find $R_0 > 0$ such that for any $v \in \mathbb{B}_2(R_0; \beta_0)$ and $\|v\|_2 = 1$, the density of $\mathbb{P}_{X^\top v}$ is bounded above and below by some constants $\bar{c}_X > 0$ and $\underline{c}_X > 0$. independent of v, on its support $\mathcal{I}_v$, which is an interval in $\mathbb{R}$.

3   BIDS Algorithm for Batched GMABC

In this section, we propose an algorithm, which we call as Batched single Index Dynamic Binning and Successive arm elimination (BIDS), for the batched GMABC problem that leverages the shared parameter structure across arms and the underlying single-index models. Our approach adopts a successive elimination strategy combined with adaptive binning [Perchet et al., 2016, Jiang and Ma, 2024], but is specifically tailored to leverage the underlying single-index model structure. The main idea of our approach is to partition the

covariate space $\mathcal{X}$ based on its one-dimensional projection along the specified index esti-mate. Once the partition is formed, decisions within each bin of the covariate space can be made by treating the problem as a standard stochastic bandit problem without covariates, with the average regret within each bin estimated as a constant.

To form a partition, an index vector β is required to determine the direction along which $x \in \mathbb{R}^d$ is projected. We consider two settings: one where a pilot estimate $\beta \in \mathbb{R}^d$ is provided with reasonable accuracy, and another where no pilot estimate is available. When a pilot estimate β is available, for instance from previous studies or other preliminary analyses, we propose the BIDS algorithm based on partitioning of the covariate space guided by the direction of β (Algorithm 1). In the absence of a pilot estimate, we begin with an initial phase where we first collect i.i.d. observations from each arm in a cyclic manner. These observations are then used to estimate the index vector. Once the direction is estimated, the BIDS algorithm applied in the first setting can be utilized. In Section 3.1, we discuss the BIDS algorithm with a given direction β. In Section 3.2, we present an algorithm to estimate the index vector during the initial phase when β is not available.

To enhance readability, we summarize some relevant notations in Table 1. A more extensive list is provided in Table S1 in the Supplementary Material.

Category	Notation	Description
Problem setup	T	Total time horizon
	K	Number of arms
	M	Number of batches
	$\{t_0, t_1, \ldots, t_M\}$	Batch end points
	$R_T(\pi)$ and $\mathcal{R}_T(\pi)$	Cumulative regret and expected cumulative regret of policy π
Parameters	β_0	True index parameter in the single index model
	η	Smoothness parameter in $(0, 1]$
	α	Margin parameter
Algorithmic Variables	β	Working direction
	$\mathcal{I}_\beta := [L_\beta, U_\beta]$	Interval of projected covariates along β
	t_{init}	Initial batch size used when pilot unknown
	$\mathcal{A}_i$	Partition of $\mathcal{I}_\beta = [L_\beta, U_\beta]$ at layer i
	$\mathcal{T}_\mathcal{A}$	$\cup_{i=1}^M \mathcal{A}_i$
	$\mathcal{B}_i$	Partition of $\mathcal{X}$ induced by $\mathcal{A}_i$
	$\mathcal{T}$	$\cup_{i=1}^M \mathcal{B}_i$
	b_i	Number of splits at layer i
	n_i	Number of equal width intervals at layer i
	$w_i = |\mathcal{I}_\beta|/n_i$	Bin width for the bin $A_i \in \mathcal{A}_i$
	$C = C_A(\beta)$	Bin in $\mathcal{X}$ corresponding to $A \in \mathcal{T}_\mathcal{A}$
	$|C|_\mathcal{T}$	the width of A for $C = C_A(\beta)$
	$p(C) = p(C_A(\beta))$	Parent bin of C defined by A
	$\text{child}(C)$	Child bin of C defined by A
	$\mathcal{L}_t, \mathcal{L}^{(i)}$	Set of active bins at time t / at batch i
	$\mathcal{I}_C$	Set of active arms in bin C

Table 1: Summary of some relevant notations used in the paper

3.1 Index based dynamic binning and arm elimination

Given a pilot direction $\beta \in \mathbb{R}^d$ such that $\|\beta\|_2 = 1$, the dynamic binning strategy employed in our algorithm can be explained through a tree-based interpretation as follows.

Hierarchical partitioning and tree structure   We build a tree $\mathcal{T}$ of depth M (recall, M is the number of batches). Each layer of the tree $\mathcal{T}$ consists of a partition of $\mathcal{X}$, the support of $\mathbb{P}_X$, defined by the direction β and the number of splits at each layer $\{b_l\}_{l=0}^{M-1}$. Let $\mathcal{I}_\beta = \{x^\top \beta ; x \in \mathcal{X}\}$, which is an interval by Assumption 3, i.e., let $\mathcal{I}_\beta = [L_\beta, U_\beta] \subseteq \mathbb{R}$. For the ith layer of the tree $\mathcal{T}$, $i \in \{1, \ldots, M\}$, we first introduce a partition $\mathcal{A}_i$ of $[L_\beta, U_\beta]$, which consists of bins obtained by splitting $[L_\beta, U_\beta]$ into $n_i := \prod_{l=0}^{i-1} b_l$ equal width intervals:

$$A_i := \begin{cases} [L_\beta + (v-1)w_i, L_\beta + vw_i) & v = 1, 2, \ldots, n_i - 1 \\ [L_\beta + (n_i - 1)w_i, U_\beta] & v = n_i \end{cases}$$

where

$$w_i = \frac{U_\beta - L_\beta}{n_i} = (U_\beta - L_\beta)(\prod_{\ell=0}^{i-1} b_\ell)^{-1} \tag{4}$$

for each layer $i = 1, 2, \ldots, M$. Then we define $\mathcal{B}_i$ for $i = 1, 2, \ldots, M$, which consists of bins $C_{A_i}(\beta)$ defined as:

$$C_{A_i}(\beta) = \{x \in \mathcal{X} : x^\top \beta \in A_i\}.$$

We note that $\mathcal{B}_i$ is a partition of $\mathcal{X}$. First, observe that for $x \in \mathcal{X}$, $x^\top \beta \in [L_\beta, U_\beta]$ with probability 1 by the definition of $\mathcal{I}_\beta$. Then there exists an $j \in \{1, \ldots, n_i\}$ such that $x^\top \beta \in A_j$. Thus, $\mathcal{X} \subseteq \cup_{j=1}^{n_i} C_{A_j}(\beta)$. On the other hand, each $C_{A_j}(\beta) \subseteq \mathcal{X}$, and therefore $\cup_{j=1}^{n_i} C_{A_j}(\beta) \subseteq \mathcal{X}$. Moreover, for $j \neq k$, $A_j \cap A_k = \emptyset$ and $C_{A_j}(\beta) \cap C_{A_k}(\beta) = \{x \in \mathcal{X} \mid x^\top \beta \in A_j \cap A_k\} = \emptyset$. Then we let the tree $\mathcal{T}$ be the collection of $\mathcal{B}_i$'s, i.e., $\mathcal{T} = \cup_{i=1}^M \mathcal{B}_i$. For future reference, we define $\mathcal{T}_{\mathcal{A}} = \cup_{i=1}^M \mathcal{A}_i$ as well. Note that by the setup, for $C \in \mathcal{T}$, we have $C = C_A(\beta)$ for some set $A \in \mathcal{T}_{\mathcal{A}}$. We will sometimes need to refer to the width of A that defines C. For $C \in \mathcal{T}$, define $|C|_{\mathcal{T}}$ as $|C|_{\mathcal{T}} = |A|$ where $C = C_A(\beta)$.

For $A \in \mathcal{T}_{\mathcal{A}}$, we define its child and parent sets as follows. Since $A \in \mathcal{T}_{\mathcal{A}}$, we

have $A \in \mathcal{A}_i$ for some $i \in \{1, \ldots, M - 1\}$. Define its *child* set as $\text{child}(A) := \{A' \in \mathcal{A}_{i+1}; A' \subseteq A\}$. The *parent* of A is then defined as $p(A) = \{A' \in \mathcal{A}_{i-1}; A \in \text{child}(A')\}$ for $i \in \{1, \ldots, M\}$. Similarly, for a bin $C_A(\beta) \in \mathcal{B}_i$, we define its child and parent as $\text{child}(C_A(\beta)) = \{C_{A'}(\beta); A' \in \text{child}(A)\}$ and $p(C_A(\beta)) = \{C_{A'}(\beta); A \in \text{child}(A')\}$. For $C \in \mathcal{T}$ (or $\mathcal{T}_\mathcal{A}$), we define $p^k(C) = p(p^{k-1}(C))$ to be the kth ancestor of C for $k \geq 2$. Then we let $\mathcal{P}(C) = \{C' \in \mathcal{T} \text{ (or } \mathcal{T}_\mathcal{A}) : C' = p^k(C) \text{ for some } k \geq 1\}$ be the set of all ancestors of C. By definition, if $A' = p(A)$ then $C_{A'}(\beta) = p(C_A(\beta))$.

Our proposed algorithm, Algorithm 1 (BIDS), proceeds in batches and each batch has two key terms, a list of *active bins* $\mathcal{L}_t$ at time t and the corresponding *active arms* $\mathcal{I}_C$ for each $C \in \mathcal{L}_t$. Before the first batch, $\mathcal{L}_1 = \mathcal{B}_1$, i.e., the list of active bins $\mathcal{L}_1$ contains all bins in layer 1, and $\mathcal{I}_C = \{1, 2\}$ for all $C \in \mathcal{L}_1$, i.e., each bin contains both active arms. In each batch, observations are drawn cyclically from each of the active arms. At the end of the batch, all the rewards in the batch are revealed. Using this information, we perform an arm elimination procedure to update the active arms set $\mathcal{I}_C$. Specifically, for each active arm set with multiple active arms, we eliminate arms that are "statistically worse than the best arm". Then, if any active bin still has more than one active arm, this suggests that the current bin is not fine enough for the decision-maker to tell the difference between the two arms. As a result, we split any active bin that still has more than one active arm into its children sets $\text{child}(C)$ in $\mathcal{T}$. Finally, we update the set of active bins and repeat this process at the end of each batch.

Since the set of active bins is only updated at the end of each batch, $\mathcal{L}_t$ only changes in the beginning of a new batch. That is, $\mathcal{L}_t$ is different from $\mathcal{L}_{t-1}$ only when $t = t_1 + 1, \ldots, t_{M-1} + 1$. We let $\mathcal{L}^{(i)} = \mathcal{L}_{t_{i-1}+1}$ to denote the list of active sets during the ith batch for $i = 1, \ldots, M$, and $\mathcal{L}^{(0)} = \emptyset$. We will say that a set $C \in \mathcal{T}$ is *born* at batch i if $C \notin \mathcal{L}^{(i-1)}$ and $C \in \mathcal{L}^{(i)}$. This happens if $p(C)$ was split at the end of batch $i - 1$. We note that by the set-up of algorithm, the sets that are born at the beginning of batch i always belong to $\mathcal{B}_i$. This is because when $i = 1$, $\mathcal{L}^{(1)} = \mathcal{B}_1$ by the set-up of the algorithm and therefore all sets that are born at batch 1 belong to $\mathcal{B}_1$. Then the sets that are born at batch i are always children of the sets that were born at $i - 1$.

Remark 2 (Unique batch elimination event for each set)**.** For a set C which was born at batch i, by the construction of the algorithm, the batch elimination procedure will be

performed for C at the end of batch i, and $C \in \mathcal{L}^{(j)}$ for all $j > i$ if and only if C has exactly one active arm after the batch elimination procedure at the end of batch i. In particular, at the end of batch i, the batch elimination procedure is performed only for those bins that are born at the beginning of batch i. As a consequence, each bin undergoes at most one batch elimination event.

Batch elimination procedure For each "newly" born $C \in \mathcal{B}_i$, for $i = 1, \ldots, M$, we obtain reward information from each active arm during batch i and perform a batch elimination event at the end of batch i. Specifically, during batch i, we obtain average rewards on C from active arms by pulling each arm in a fixed, cyclic order whenever $X_t \in C$. At the end of batch i, we perform a batch elimination procedure.

More precisely, for $i = 1, \ldots, M$, let $\tau_{C,i}(s) = \inf\{n \geq \tau_{C,i}(s-1) + 1; X_n \in C\}$ for $s = 1, 2, \ldots$ be the sth time at which the sequence X_t is in C during the batch i where we let $\tau_{C,i}(0) = t_{i-1}$. Let $m_{C,i} = \sum_{t=t_{i-1}+1}^{t_i} 1\{X_t \in C\}$ be the total number of visits of X_t to C during batch i. For the sth visit to C, we pull an observation from the arm k such that $k \equiv s \pmod{K}$. In particular, when $K = 2$, we pull $k = 1$ arm on odd-numbered visits, and pull $k = 2$ arm on even-numbered visits. Let $\tau_{C,i}^{(k)} = \{\tau_{C,i}(s); 1 \leq s \leq m_{C,i}, s \bmod K \equiv k\}$ be the set of time points t during batch i where X_t visits C, and the arm k is pulled for $k = 1, \ldots, K$. Define the average rewards for C from arm $k \in \{1, 2\}$ during batch $i \in \{1, \ldots, M\}$ as:

$$\bar{Y}_{C,i}^{(k)} = \frac{1}{|\tau_{C,i}^{(k)}|} \sum_{t \in \tau_{C,i}^{(k)}} Y_t^{(k)}. \tag{5}$$

Once $\bar{Y}_{C,i}^{(k)}$ for $k \in \{1, 2\}$ are obtained, we check whether,

$$\max_{l \in \{1,2\}} \bar{Y}_{C,i}^{(l)} - \bar{Y}_{C,i}^{(k)} > U(m_{C,i}, T, C), \tag{6}$$

where we define,

$$U(m, T, C) := 4\sqrt{\frac{2\log(2T|C|_{\mathcal{T}})}{m}}, \tag{7}$$

where we recall the definition $|C|_{\mathcal{T}} = |A|$ for a set A such that $C = C_A(\beta)$. In particular, for $C \in \mathcal{B}_i$, $|C|_{\mathcal{T}} = |A_i|$ for $A_i \in \mathcal{A}_i$. We eliminate k from the set of active arms for C if k

satisfies (6). Also, note that $m_{C,i}$ has expectation given by,

$$m_{C,i}^* = \mathbb{E}[m_{C,i}] = (t_i - t_{i-1})\mathbb{P}_X(X \in C).$$

In Lemma S-3, we show that $m_{C,i}$ concentrates around its expectation $m_{C,i}^*$ uniformly over all $C \in \mathcal{T}$.

Toy example of hierarchical partitioning and active set updates We use a simple illustrative example to demonstrate hierarchical partitioning and active bin updates. To visually demonstrate the process, we simulated a simple linear model example $Y_t = \beta_1 X_{t,1} + \beta_2 X_{t,2} + \epsilon_t$ with 2-dimensional covariates $X_t = (X_{t,1}, X_{t,2}) \in \mathbb{R}^2$ for $t = 1, \ldots, 25$ and $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2 = 1)$.

Plots (a)–(d) in Figure 1 illustrate the partitioning of $\mathcal{X}$ at a specific layer using the projection along the given (estimated) direction. Plot (a) shows a simulated dataset $(X_{t,1}, X_{t,2}, Y_t)_{t=1}^{25}$. Plot (b)–(c) show the projection of X_t's along the given direction (red dotted line). The solid black circles represent the projections of the data points onto the estimated direction, with the corresponding covariates linked by grey dotted lines. The partitioning is based on partitioning the projections, specifically in this example splitting the interval $[-5, 7]$ into four sub-intervals (which constitute $\mathcal{A}_i$), which are visualized using different colors in Figure 1(d). The resulting partitioning of the covariates $(\mathcal{B}_i)$ in $\mathbb{R}^2$ is visualized by mapping the projections back to hollow circles filled with the corresponding colors, connected by dotted lines. The same process holds for all layers, $i = 1, \ldots, M$.

Figure 2 demonstrates the hierarchical refinement process of bins and active set updates. Consider the tree structure shown in Figure 2, which represents a tree with depth $M = 3$. At level 1, there are $b_1 = 4$ splits, at level 2, there are $b_2 = 2$ splits, and at level 3, there are $b_3 = 2$ splits. At each level, green bins indicate that arm elimination happened at the end of that batch within those bins, and they remain active in all subsequent layers without further splitting. On the other hand, red bins are further divided into child bins that subsequently become active. For instance, in the first layer, within the interval $[-2, 1]$, an arm elimination occurred, leaving a single active arm. Since this bin is not further split, it remains active throughout the decision-making process. In contrast, an arm elimination
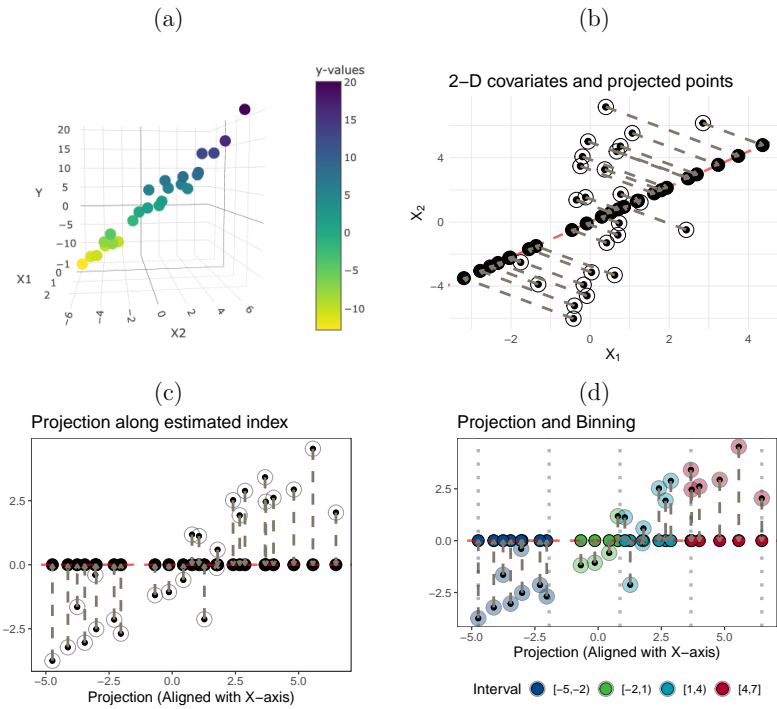
Figure 1: (a) 3-D representation of the data (toy example) such that y depends on a linear combination of the covariates. (b) Projecting $X \in \mathbb{R}^2$ (circles with holes) in the single-index direction (red dotted line with black filled circles as projected points) and grey lines connecting the original points to the projected points. (c) Rotating (b) to align the SIR direction with the x-axis. (d) Binning projected SIR direction into bins, different colors represent covariates that fall in a particular bin.

did not occur for $[4, 7]$, so it was further split into $[4, 5.5)$ and $[5.5, 7]$ in level 2, with both bins initially active. Note in layer 2, batch elimination procedure was only performed for $[-5, -3.5), [-3.5, -2), [4, 5.5)$, and $[5.5, 7]$, while $[-2, 1)$ and $[1, 4)$ were not considered for arm elimination as they already only have a single active arm. Once a bin turns green (i.e., it contains exactly one active arm) at any layer, it remains green and unchanged in the subsequent layers.
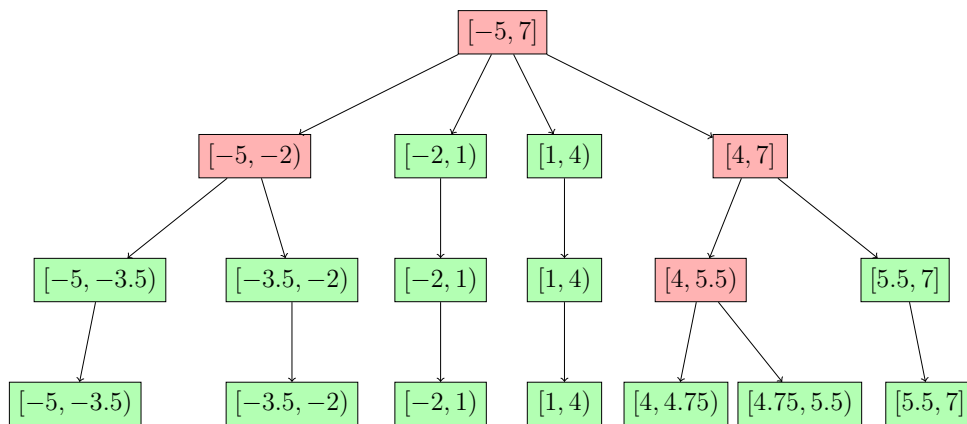

Figure 2: Binning with successive elimination with $M = 3$: the green intervals correspond to the active bins with exactly one arm at each level.

In Algorithm 1, we summarize the BIDS algorithm, which performs hierarchical partitioning based on projection along a given index vector and dynamic binning through successive arm elimination and active set updates.

3.2 Estimation of single-index vector without a pilot estimate

In this subsection, we discuss the process of estimating the single-index vector using a separate initial phase when no pilot estimate is available. We divide the time horizon $1, \ldots, T$ into two phases: an initial phase (first batch), during which we draw i.i.d. samples from each arm $k \in \{1, 2, \ldots, K\}$, and a second phase where we run the BIDS algorithm (Algorithm 1) using the estimated direction.

More specifically, in the initial phase, we draw i.i.d. samples cyclically for $k = 1, \ldots, K$, assigning arm k when $t(\bmod K) \equiv k$. For example, when $K = 2$, we draw from arm 1 if t is odd and from arm 2 if t is even. Formally, we construct i.i.d. datasets $\mathcal{D}_{\text{init}}^{(k)} = (X_t, Y_t^{(k)})_{t \in \mathbb{T}_k}$, where $\mathbb{T}_k = \{1 \le t \le t_{\text{init}} \mid t(\bmod K) \equiv k\}$ represents the set of time points at which arm k is selected during the initial phase. Once these i.i.d. datasets are available, any single-index

15

Algorithm 1 BIDS algorithm

1: **Input:** No. of batches M, grid $\{t_i\}_{i=0}^{M}$, split factors $\{b_i\}_{i=0}^{M-1}$, working direction: β
2: Initialize active bins: $\mathcal{L}^{(1)} \leftarrow \mathcal{B}_1$.
3: Initialize active arms: $\mathcal{I}_C \leftarrow \{1, 2, \ldots, K\}$ for all $C \in \mathcal{L}^{(1)}$
4: **for** $i = 1, \ldots, M$ **do**
5:     **for** $t = t_{i-1} + 1, \ldots, t_i$ **do**   ▷ Draw Observations (during batch i)
6:         Find $C \in \mathcal{L}^{(i)}$ such that $X_t \in C$.
7:         Pull an arm from $\mathcal{I}_C$ in a cyclic manner (let s be the number of visits to C up to the current time. set $Y_t = Y_t^{(k)}$, for $k \equiv s\,(\mathrm{mod}\,K)$.)
8:     **end for**
9:     **if** $t = t_i$ and $i < M$ **then**   ▷ Batch Elimination (at the end of batch i)
10:         Rewards during batch i, $Y_{t_{i-1}+1}, \ldots, Y_{t_i}$, are revealed.
11:         Initialize $\mathcal{L}^{(i+1)} = \{\}$.
12:         **for** $C \in \mathcal{L}^{(i)}$ **do**   ▷ Iterate over active bins
13:             **if** $|\mathcal{I}_C| = 1$ **then**   ▷ If only one active arm remains in C
14:                 $\mathcal{L}^{(i+1)} = \mathcal{L}^{(i+1)} \cup \{C\}$
15:                 Break (Proceed to the next bin C)
16:             **else** $|\mathcal{I}_C| > 1$ ▷ If more than one active arm remains
17:                 $\bar{Y}_{C,i}^{\max} = \max_{k \in \mathcal{I}_C} \bar{Y}_{C,i}^{(k)}$
18:                 **for** k in $\mathcal{I}_C$ **do**   ▷ successive arm elimination
19:                     **if** $\bar{Y}_{C,i}^{\max} - \bar{Y}_{C,i}^{(k)} > U(m_{C,i}, T, C)$ **then**
20:                         $\mathcal{I}_C = \mathcal{I}_C \setminus \{k\}$
21:                     **end if**
22:                 **end for**
23:                 **if** $|\mathcal{I}_C| > 1$ **then**   ▷ If arm elimination did not occur,
24:                     $\mathcal{I}_{C'} = \mathcal{I}_C$, for $C' \in \mathrm{child}(C)$ ▷ split the bin into children bins
25:                     $\mathcal{L}^{(i+1)} = \mathcal{L}^{(i+1)} \cup \{C'; C' \in \mathrm{child}(C)\}$ ▷ update the active bins
26:                 **end if**
27:             **end if**
28:         **end for**
29:     **end if**
30: **end for**

---

regression (SIR) algorithm can be employed to estimate the direction β_0. For example, in Section S4 of the Supplementary Material, we demonstrate this process using the Sliced Average Derivative Estimation (SADE) method from Babichev and Bach [2018].

Let $\hat{\beta}^{(k)}$ denote the estimate of β_0 obtained using $\mathcal{D}_{\mathrm{init}}^{(k)}$ for $k = 1, \ldots, K$. Since single-index models estimate the direction up to a rotation, we cannot simply combine these vectors by taking their (weighted) average. We propose to first estimate the projection matrix $\mathcal{P}_0 = \beta_0 \beta_0^\top$ of β_0 by computing a (weighted) average of the projection matrices from each arm with weights ω_k, i.e., $\hat{\mathcal{P}} = \sum_{k=1}^{K} \omega_k \hat{\beta}^{(k)}(\hat{\beta}^{(k)})^\top$, then we obtain the final vector

$\hat{\beta}$ by computing the first eigenvector of the estimated matrix $\hat{\mathcal{P}}$. In our simulations and real-data illustrations in Sections 5 and 6, we use the average with equal weights $\omega_k = 1/K$ for datasets corresponding to each of the K arms.

We summarize the procedure for estimating the single index vector during the initial phase in Algorithm 2 as follows:

Algorithm 2 Initial Direction Estimation

1: **Input:** Number of samples in the initial phase t_{init}, weights for each arm $(\omega_k)_{k=1}^K$, an SIR algorithm $\text{SIR}(\cdot)$
2: **for** $t = 1, \dots, t_{\text{init}}$ **do**
3:     Pull arm k if $k \equiv t(\text{mod} K)$.
4: **end for**
5: **for** $k = 1, \dots, K$ **do**
6:     Define the indices assigned to arm k: $\mathbb{T}_k = \{t; 1 \le t \le t_{\text{init}}, t(\text{mod} K) = k\}$
7:     Compute $\hat{\beta}^{(k)} \leftarrow \text{SIR}((X_t, Y_t^{(k)})_{t \in \mathbb{T}_k})$
8: **end for**
9: Compute the estimated projection matrix $\hat{\mathcal{P}} = \sum_{k=1}^K \omega_k \hat{\beta}^{(k)} (\hat{\beta}^{(k)})^\top$ of $\mathcal{P}_0$.
10: Return $\hat{\beta}$, the eigenvector corresponding to the largest eigenvalue of $\hat{\mathcal{P}}$.

---

4   Regret upper bounds

First, recall that our adaptive binning is performed by partitioning the projected space, where the projection is based on the pilot index vector. As a result, the regret depends on how accurate the initial index vector is. To quantify this accuracy, we make the following assumption regarding the ℓ_2-difference between the initial index β and the true index β_0.

Since we are estimating the direction of β_0 rather than the vector itself, we quantify the distance in terms of the principal angle between two directions. More specifically, for $u, v \in \mathbb{R}$ such that $\|u\|_2 = \|v\|_2 = 1$, let $\angle u, v = \cos^{-1}(|u^\top v|) \in [0, \pi/2]$ be the principal angle between the directions u and v. Note that $\angle u, v = 0$ implies that $|u^\top v| = 1$, i.e., u and v are identical up to sign. At the other extreme, $\angle u, v = \pi/2$ implies that $|u^\top v| = 0$, which means u and v are orthogonal. Equivalently, we can express this in terms of the sine principal angle distance $\sin \angle u, v \in [0, 1]$, where $\sin \angle u, v = 0$ implies that u, v are identical up to sign and $\sin \angle u, v = 1$ implies u and v are orthogonal.

Assumption 4. The initial vector β satisfies

$$\sin \angle \beta, \beta_0 \leq C_0 T^{-\xi/(2\eta+1)} \tag{8}$$

for some $C_0 > 0$ and $\xi \geq 1$.

Note that the inequality (8) implies there exists $o \in \{-1, 1\}$ such that $\|\beta \cdot o - \beta_0\|_2 \leq 2^{1/2} C_0 T^{-\xi/(2\eta+1)}$ (see, e.g., proof of Lemma 2). For future reference, we define $\beta_{sgn} = \beta \cdot o$ which is either $\beta_{sgn} = \beta$ or $\beta_{sgn} = -\beta$ such that the above bound holds. We note that β_{sgn} is an oracle quantity since it depends on the unknown sign. It is used only in the proof and is not required for the actual implementation of the algorithm.

4.1   Regret analysis when a pilot index is available

When a pilot direction satisfying Assumption 4 is provided, our regret analysis follows a similar approach to the adaptive binning with successive elimination method of Perchet and Rigollet [2013], Jiang and Ma [2024], but with non-trivial modifications to accommodate the single-index (GMABC) model setting.

We show that, with an optimal choice of batch size and splitting factor, our regret bound for Algorithm 1 matches (up to logarithmic factors) the minimax rate of non-parametric batched contextual bandits but with $d = 1$ (noting that their γ depends on the covariate dimension d, meaning that their rate for $d > 1$ is significantly slower than ours). To achieve this, we carefully select the batch size and splitting factors to ensure that the regret from one batch does not dominate the regrets from other batches. Specifically, we adopt the allocation rule and splitting factor setup proposed by Jiang and Ma [2024], but with the choice of dimension $d = 1$.

Recall that the list of split factors $\{b_i\}_{i=0}^{M-1}$ determines the number of bins $n_i = \prod_{l=0}^{i-1} b_l$ in the partition $\mathcal{A}_i$ of $[L_\beta, U_\beta]$ and the width $w_i = (U_\beta - L_\beta)/n_i$ of each bin in $\mathcal{A}_i$. Let $\gamma = \frac{\eta(1+\alpha)}{2\eta+1}$ and set $a \asymp (T^{\frac{1-\gamma}{1-\gamma^M}})$. The split factors are then chosen as follows:

$$b_0 = \lfloor a^{\frac{1}{2\eta+1}} \rfloor, \text{ and } b_i = \lfloor b_{i-1}^{\gamma} \rfloor, i = 1, \ldots, M-2. \tag{9}$$

Note that this leads to the following choice of bin widths:

$$w_i \asymp (b_0 b_1 \ldots b_{i-1})^{-1} \asymp b_0^{-(1+\gamma+\cdots+\gamma^{i-1})} \asymp T^{-\frac{1-\gamma^i}{1-\gamma^M}\frac{1}{2\eta+1}}, \ i = 1, \ldots, M-1. \qquad (10)$$

The number of samples allocated to batch i, i.e., $t_i - t_{i-1}$, is chosen so that it increases with the number of bins in the ith layer. Specifically, we let

$$t_i - t_{i-1} = \lfloor c_B \, w_i^{-(2\eta+1)} \log{(Tw_i)} \rfloor, 1 \le i \le M-1. \qquad (11)$$

where $c_B = 4(4L_0 + 1)^{-2}(\bar{c}_X)^{-1}$, with $L_0 = L(2^{3/2}C_0 R_X + 1)^\eta$, is a constant independent of T. With these choices, we now present Theorem 1, which establishes the regret bound for the proposed BIDS algorithm when the batch size M is at most of order $\log(T)$. The proof is provided in Section S2.1 of the Supplementary Material.

Theorem 1. *Suppose Assumptions 1–3 hold, and let a pilot direction β with $\|\beta\|_2 = 1$ be given, satisfying Assumption 4. Assume T is sufficiently large such that $\beta_{sgn} \in \mathbb{B}_2(R_0; \beta_0)$ for $R_0 > 0$ defined in Assumption 3. Suppose $\alpha\eta \le 1$ and $M \le C_1 \log T$ for some $C_1 > 0$. For the BIDS algorithm π described in Algorithm 1, with the choices of split factors and batch size satisfying (9) and (11), the following bound on the expected regret $\mathcal{R}_T(\pi) = \mathbb{E}[R_T(\pi)]$ holds for sufficiently large T:*

$$\mathcal{R}_T(\pi) \le C_2 M \log(T) T^{\frac{1-\gamma}{1-\gamma^M}},$$

where $\gamma = \frac{\eta(1+\alpha)}{2\eta+1}$, where C_2 is a constant depending on model parameters such as $\alpha, \eta, D_0, L, \bar{c}_X, \underline{c}_X$, and R_X, but not on the sample size T.

Corollary 1 shows that when the number of batches is *sufficiently large*, specifically at least of order $\log\log(T)$, the regret bound from the batched bandit matches (up to logarithmic factors) the regret bound in the fully online setting but with $d = 1$, where feedback arrives instantaneously.

Corollary 1. *Assume the same conditions as in Theorem 1. For $M \ge D_1 \log\log{(T)}$ for*

some D_1 that depends on γ, Theorem 1 becomes:

$$\mathcal{R}_T(\pi) \leq C_3 M \log(T) T^{1-\gamma},$$

where $\gamma = \frac{\eta(1+\alpha)}{2\eta+1}$, where C_3 depends on the pilot single index parameter β and other constants such as $\alpha, \eta, D_0, D_1, L, \bar{c}_X, \underline{c}_X$, and R_X, but not on the sample size T.

Proof. It is sufficient to show that for $M \geq D_1 \log\log(T)$, $T^{\frac{\gamma^M}{1-\gamma^M}} = O(1)$ since

$$T^{\frac{1-\gamma}{1-\gamma^M}} = T^{1-\gamma}(T^{\frac{\gamma^M}{1-\gamma^M}})^{1-\gamma}.$$

Note that

$$T^{\frac{\gamma^M}{1-\gamma^M}} \leq C_0$$

$$\leftrightarrow \frac{\gamma^M}{1-\gamma^M} \log(T) \leq \log(C_0)$$

$$\leftrightarrow \gamma^M \cdot (\log(T) + \log(C_0)) \leq \log(C_0), \tag{12}$$

and for $M \geq D_1 \log\log(T)$, $\gamma^M \leq \gamma^{D_1 \log\log(T)} = (\log T)^{-D_1 \log \gamma^{-1}}$, for $D_1 \geq (1/\log \gamma^{-1})$ and $0 < \gamma < 1$, shrinks to zero faster than $\log(T)$ ensuring (12) holds for sufficiently large T. $\qquad\square$

Remark 3 (Number of batches and regret bound rates)**.** Theorem 1 and Corollary 1 show that the BIDS Algorithm, when provided with a sufficiently accurate pilot estimate, achieves near-optimal regret performance across different batch regimes. Specifically, when the number of batches is very small ($M \lesssim \log\log(T)$), the expected regret we obtain in Theorem 1 matches the minimax optimal rate up to logarithmic factors for nonparametric batched bandits in Jiang and Ma [2024] (Theorem 1) with $d = 1$ (without the curse of dimensionality). When the number of batches is moderate ($\log(T) \gtrsim M \gtrsim \log\log(T)$), our proposed algorithm achieves regret that is optimal (up to logarithmic factors) for the setting without batch constraints [Perchet and Rigollet, 2013] but again with $d = 1$. In other words, in this setting, our BIDS algorithm (i.e., Algorithm 1) is able to match the optimal regret for nonparametric bandits (up to log factors) in the fully online setting with only $O(\log\log T)$ policy updates, without the curse of dimensionality.

4.2 Regret analysis when no pilot estimate is available

When no pilot index estimate is available, both the index vector and the link function must be estimated within the batches. We propose using the first batch to estimate the index vector β (Algorithm 2), followed by performing the BIDS algorithm with the estimated index vector β for the remaining batches (Algorithm 1).

Recall that in the initial phase, for $t \in \{1, \ldots, t_{\text{init}}\}$, we draw i.i.d. random samples from each arm. Any suitable single-index model can then be applied in this phase to estimate the index vector. The index vector can generally be estimated at a parametric rate (e.g., Li and Duan [1989], Babichev and Bach [2018], Kuchibhotla and Patra [2020]). Assumption 5 specifies the requirement for the index vector from a Single-Index Regression (SIR) method used in Algorithm 2. Specifically, we require that the SIR algorithm used in Algorithm 2 produces an estimate that satisfies a parametric error bound up to a log term with high probability when applied to an i.i.d dataset of size n_k.

Assumption 5. Let $k \in \{1, 2, \ldots, K\}$ be fixed, and let $\hat{\beta}^{(k)}$ be the estimated vector from an i.i.d sample of size n_k, $(x_i, Y_i^{(k)})_{i=1}^{n_k}$ where $Y_i^{(k)}$ follows the single index model (2). For a sufficiently large n_k, with probability $1 - C_4 n_k^{-\phi}$ for some $\phi \geq 1$ and $C_4 > 0$, the following bound holds:

$$\sin \angle \hat{\beta}^{(k)}, \beta_0 \leq C_5 \frac{\text{polylog}(n_k)}{\sqrt{n_k}}, \tag{13}$$

for some constant $C_5 = C_5(d, \phi)$ which can depend on model parameters but is independent of the sample size n_k.

Remark 4. As an example of a single index estimation algorithm that satisfies Assumption 5, we discuss the Sliced Average Derivative Estimator (SADE) of Babichev and Bach [2018] in Section S4 in Supplementary Material. In particular, Theorem 3 in Supplementary Material establishes that, under mild conditions, the estimates $\hat{\beta}^{(k)}$ obtained using the SADE method satisfy Assumption 5. Please see Section S4 in the Supplementary Material for more details.

The following Lemma 2 shows that under Assumption 5, the estimated direction $\hat{\beta}$ from Algorithm 2 is (up to sign) within a neighborhood of β_0 that shrinks at an approximate

rate of $t_{\text{init}}^{-1/2}$, with an additional log term.

Lemma 2. *Let $\hat{\beta}^{(1)}, \ldots, \hat{\beta}^{(K)}$ be the estimated index vectors from each arm, and let $\hat{\beta}$ be the final estimated direction from Algorithm 2. Suppose Assumption 5 holds for each $k = 1, \ldots, K$. For sufficiently large T, with probability at least $1 - KC_4(t_{\text{init}}/2K)^{-\phi}$, we have:*

$$\sin \angle \hat{\beta}, \beta_0 \leq \tilde{C} \frac{polylog(t_{\text{init}})}{\sqrt{t_{\text{init}}}},$$

for a constant $\tilde{C} = \tilde{C}(d, \phi, K)$. Moreover, there exists $\hat{o} \in \{-1, 1\}$ such that

$$\|\hat{\beta} \cdot \hat{o} - \beta_0\|_2 \leq 2^{1/2} \tilde{C} \frac{polylog(t_{\text{init}})}{\sqrt{t_{\text{init}}}}. \tag{14}$$

The proof for Lemma 2 is provided in Section S2.3 in Supplementary material.

In terms of regret bound analysis, the primary difference in this setting compared to the previous one is that regret will accrue from the observations drawn during the initial phase. In particular, the cumulative regret incurred is given by,

$$\begin{aligned}
\mathcal{R}_T(\pi) &= \mathbb{E}[\sum_{t=1}^{T} g^{(*)}(X_t) - g^{(\pi_t(X_t))}(X_t)] \\
&= \mathbb{E}\left[\sum_{t=1}^{t_{\text{init}}} (g^{(*)}(X_t) - g^{(\pi_t(X_t))}(X_t)) + \sum_{t=t_{\text{init}}+1}^{T} (g^{(*)}(X_t) - g^{(\pi_t(X_t))}(X_t))\right] \\
&\leq t_{\text{init}} + \mathbb{E}\left[\sum_{t=t_{\text{init}}+1}^{T} (g^{(*)}(X_t) - g^{(\pi_t(X_t))}(X_t))\right] \tag{15} \\
&=: t_{\text{init}} + \mathcal{R}_{T-t_{\text{init}}}(\pi; \beta).
\end{aligned}$$

where (15) follows from the fact that $|Y_t| \leq 0.5$.

The size of the first batch t_{init} needs to be chosen to balance two competing factors: achieving sufficient accuracy in estimating the single-index parameter while not incurring too much regret. Assumption 4 requires the working direction β to be within a $T^{-\xi/(2\eta+1)}$ neighborhood of β_0, up to sign, for $\xi \geq 1$. Therefore, to ensure that the estimated direction β is sufficiently accurate to satisfy Assumption 4, we consider the initial phase length as $t_{\text{init}} \asymp \text{polylog}(T) T^{\frac{2}{(2\eta+1)}}$ so that $\frac{\text{polylog}(t_{\text{init}})}{\sqrt{t_{\text{init}}}} \lesssim T^{-\frac{1}{2\eta+1}}$.

Theorem 2. *Suppose Assumptions 1–3 hold. Also, assume that the estimates from Algorithm 2 satisfy Assumption 5. Let $\alpha\eta \leq 1$ and $M = O(\log T)$. Consider the algorithm π, which executes Algorithm 2 during the initial phase with $t_{\text{init}} \asymp \text{polylog}(T) T^{\frac{2}{(2\eta+1)}}$, followed by Algorithm 1 for the remaining batches. Then, the regret for the resulting algorithm π is upper bounded by,*

$$\mathcal{R}_T(\pi) \leq C_6 \, \text{polylog}(T) \max\{T^{\frac{2}{(2\eta+1)}}, T^{\frac{1-\gamma}{1-\gamma^M}}\},$$

where $\gamma = \frac{\eta(1+\alpha)}{2\eta+1}$, where C_6 depends on the single index parameter β and other constants such as $\alpha, \eta, D_0, L, R_X, \bar{c}_X, \underline{c}_X$.

Proof. We know from (15) that,

$$\mathcal{R}_T(\pi) \leq t_{\text{init}} + \mathcal{R}_{T-t_{\text{init}}}(\pi; \beta).$$

Define $\mathcal{E}_\beta$ to be the event that the inequality (13) holds for all $k \in \{1, \ldots, K\}$, which holds with probability at least $1 - KC_4(t_{\text{init}}/2K)^{-\phi}$ under Assumption 5. We have,

$$\mathcal{R}_T(\pi) \leq t_{\text{init}} + \mathbb{E}[R_{T-t_{\text{init}}}(\pi; \beta)1(\mathcal{E}_\beta) + R_{T-t_{\text{init}}}(\pi; \beta)1(\mathcal{E}_\beta^c)]$$
$$\leq t_{\text{init}} + \mathbb{E}[R_{T-t_{\text{init}}}(\pi; \beta)1(\mathcal{E}_\beta)] + (T - t_{\text{init}})\{(2K)^\phi KC_4\}t_{\text{init}}^{-\phi}.$$

On $\mathcal{E}_\beta$, by Lemma 2,

$$\sin \angle\hat{\beta}, \beta_0 \leq \tilde{C}\frac{\text{polylog}(t_{\text{init}})}{\sqrt{t_{\text{init}}}}. \tag{16}$$

Since $t_{\text{init}} \asymp \text{polylog}(T) T^{\frac{2}{(2\eta+1)}}$ so that

$$\frac{\text{polylog}(t_{\text{init}})}{\sqrt{t_{\text{init}}}} \lesssim T^{-\frac{1}{2\eta+1}},$$

the projection vector $\hat{\beta}$ satisfies Assumption 4 on $\mathcal{E}_\beta$ with $\xi = 1$. Then by Theorem 1,

$$\mathbb{E}[R_{T-t_{\text{init}}}(\pi; \beta)1(\mathcal{E}_\beta)] \lesssim M \log(T - t_{\text{init}})(T - t_{\text{init}})^{\frac{1-\gamma}{1-\gamma^M}}.$$

Then,

$$\mathcal{R}_T(\pi) \lesssim \text{polylog}(T)T^{\frac{2}{(2\eta+1)}} + M\log(T)T^{\frac{1-\gamma}{1-\gamma^M}} + T(\text{polylog}(T)T^{\frac{2}{(2\eta+1)}})^{-\phi} \qquad (17)$$
$$\lesssim \text{polylog}(T)\max\{T^{\frac{2}{(2\eta+1)}}, T^{\frac{1-\gamma}{1-\gamma^M}}\},$$

where we use the fact that the first term dominates the third term in (17) since $2 \geq 2\eta + 1 - 2\phi$ since $\eta \in (0,1]$, and $\phi \geq 1$. $\qquad\square$

Now, similar to the characterization of the regret results based on the number of batches in Corollary 1, we get the following special case.

Corollary 2. *For $M \geq D_1 \log\log(T)$, Theorem 2 becomes:*

$$\mathcal{R}_T(\pi) \leq C_7 \, \text{polylog}(T)\max\{T^{\frac{2}{(2\eta+1)}}, T^{1-\gamma}\},$$

where $\gamma = \frac{\eta(1+\alpha)}{2\eta+1}$, where C_7 depends on the single index parameter β and other constants such as $\alpha, \eta, D_0, D_1, L, R_X, \bar{c}_X, \underline{c}_X$.

Note, that in both the bounds in Theorem 2 and Corollary 2, we pay a price for not knowing the pilot index. However, in certain problem instances, we can still achieve the same rates as those in Theorem 1 and Corollary 1. In Theorem 2, it is easy to note that the second term dominates when

$$\frac{2}{2\eta+1} \leq \frac{1-\gamma}{1-\gamma^M},$$

which simplifies to

$$\eta(1-\alpha) \geq 1 - 2\gamma^M. \qquad (18)$$

For instance, when $\eta = 1$ (Lipschitz continuous functions), this condition (18) reduces to $(1+\alpha)^M - \frac{3^M}{2}\alpha \geq 0$. This implies that, for example, when the number of batches after the initial batch is $M = 2$, the rate in Theorem 2 matches with that of Theorem 1 for $0 < \alpha \leq 0.5$. The range of α for which the rate without a pilot estimate matches with the rate with a pilot estimate becomes smaller as the number of batches increases. For

instance, when M is large enough that $\gamma^M \approx 0$, we obtain the condition $\alpha \leq (\eta - 1)/\eta$, which holds only if $\eta = 1, \alpha = 0$. That is, the rate without a pilot estimate is optimal only under the margin condition $\alpha = 0$. At the other extreme, when $\alpha = 1$, the regret grows as $\tilde{O}(T^{\frac{2}{3}})$, whereas when the pilot estimate is known (as in Corollary 1), the regret grows as $\tilde{O}(T^{1/3})$. This gap is likely due to the non-adaptive nature of our index parameter estimation method, and an interesting direction for future work would be to design an algorithm that better leverages the margin condition for settings with a moderate to large number of batches. Nevertheless, it is still encouraging to note that we get a sub-linear regret corresponding to $d = 1$, even when we use some initial data to estimate β_0.

5   Simulation Study

In this section, we present numerical experiments to illustrate the performance of the proposed BIDS algorithm (Algorithm 1) in comparison to the nonparametric analogue: Batched Successive Elimination with Dynamic Binning (BaSEDB) algorithm of Jiang and Ma [2024]. We consider both the cases discussed in Sections 4.1 and 4.2: 1) when the pilot direction is available under varying degrees of accuracy, and 2) when the pilot direction is unknown and estimated using the initial t_{init} amount of data, under varying signal-to-noise level settings.

Simulation settings.   We consider $K = 2$ arm setting, where the mean reward functions $g^{(1)}$ and $g^{(2)}$ follow a single index model structure with the shared parameter $\beta_0 \in \mathbb{R}^d$, i.e.,

$$g^{(k)}(x) = f^{(k)}(x^\top \beta_0), \ k = 1, 2,$$

where $f^{(1)}, f^{(2)} : [l, u] \to \mathbb{R}$ are link functions for arm 1 and 2. We consider $d = 5$ throughout the simulation.

First, the index vector β_0 is simulated by generating a scaled normal random vector. Specifically, we first draw $u \sim N_d(0, I_d)$ and then let $\beta_0 = u/\|u\|_2$. Regarding the covariate distribution, we let each $X_t \in \mathbb{R}^d$ follow a truncated multivariate normal distribution for

$t = 1, \ldots, T$, i.e., $X_t \sim N_T(0, \Sigma_X)$ whose density is given by:

$$f_X(x) = \begin{cases} \frac{1}{Z(\Sigma_X)} \exp\{-\frac{1}{2}x^\top \Sigma_X^{-1} x\} & x \in \mathcal{H} \\ 0 & \text{otherwise,} \end{cases}$$

with $\Sigma_X = 5^2 I_d$. The normalization constant $Z(\Sigma_X)$ is given by $Z(\Sigma_X) = \int_{x \in \mathbb{R}^d} e^{-\frac{1}{2}x^\top \Sigma_X^{-1} x} 1\{x \in \mathcal{H}\} dx$ with the truncation region $\mathcal{H} = \prod_{j=1}^d 1\{|x_j| \leq 3\}$. Additionally, we have considered other covariate distributions, including the Normal distribution without truncation and the uniform distribution. The results were qualitatively similar to those presented below for the truncated normal case and are presented in Section S5 of Supplementary material.

To define 1-dimensional link functions, first let us define,

$$f(x) = a + \frac{2}{B} \sum_{j=1}^{B/2} v_j \, \phi\left(\frac{B}{u-l}(x - q_j)\right), \tag{19}$$

where $q_j = l + (2j - 1)\frac{u-l}{B}$ for $j = 1, \ldots, B/2$, $\phi(x) = (1 - |x|)1\{|x| \leq 1\}$, v_j for $j = 1, \ldots, B/2$ are Rademacher random variables, each taking values ± 1 with equal probability, and $l, u = \mp 3\sqrt{d}$.

We consider two simulation settings for the link functions as illustrated in Figure 3.

Setting 1: $f^{(1)}(x) = f(x)$ with $a = 0.5, B = 8$, and $f^{(2)}(x) = \frac{1}{2} + x$.

Setting 2: $f^{(1)}(x) = f(x)$ with $a = 0.5, B = 8$, and $f^{(2)}(x) = f(x)$ with $a = 0.75, B = 5$.
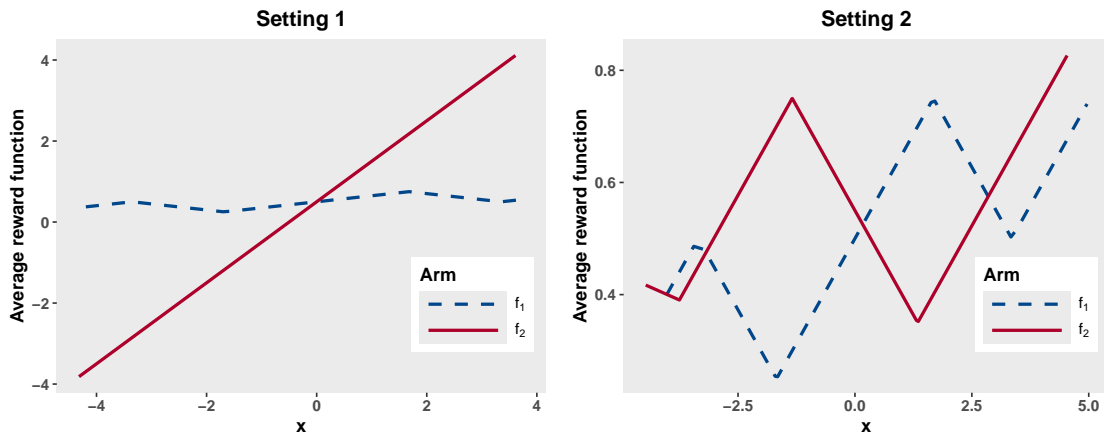

Figure 3: Mean reward functions for the two simulation settings

We let $Y_t^{(k)} = f^{(k)}(X_t) + \epsilon_t$, where $\epsilon_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$ for $t = 1, \ldots, T$, with $\sigma^2 > 0$, representing the noise variance. In the first case, where we test the performance of the BIDS algorithm with varying accuracies of pilot directions, we set $\sigma^2 = 0.01^2$. In the second case, where we estimate the initial direction under different noise levels, we set $\sigma \in \{1, \ldots, 8\}$ for setting 1 and $\sigma \in \{0.02, 0.09, 0.16, \ldots, 1\}$ for setting 2. In both settings, we set the time horizon $T = 10^6$.

Algorithm set-ups. Both BIDS and BaSEDB algorithms require specifying the number of batches M and the grid points $\{t_i\}_{i=0}^M$. We set the total number of batches $M = 5$ in both cases. For the BaSEDB algorithm, we follow the specifications described in Jiang and Ma [2024] for choosing grid points. For the BIDS algorithm (Algorithm 1), in the first case with known pilot directions, we make grid points choices according to (9) and (10), and in the second case with unknown pilot directions, the initial batch size is set to $T^{2/3}$, and the remaining time points are partitioned according to the same rules. In addition, in the latter case, Algorithm 2 requires specifying an SIR algorithm and arm weights. For the SIR algorithm, we use the SADE estimator (Algorithm 3) from Babichev and Bach [2018] and we used equal arm weights $\omega_k = 1/2, k = 1, 2$ for combining directions from each arm. Additionally, both algorithms require specifying the endpoints for hierarchical partitioning: $[L_\beta, U_\beta]$ such that $L_\beta \leq x^\top \beta \leq U_\beta$ for the BIDS algorithm, and $[L, U]$ such that $L \leq x_j \leq U$ for all $j = 1, \ldots, d$ for the BaSEDB algorithm. We constructed these intervals based on the observed minimum and maximum values from i.i.d. samples for each arm in the first batch, and expanded them by 20%. More specifically, we obtained the minimum a and maximum b, where $a = \min_{t \in (t_0, t_1]} x_t^\top \beta$ and $b = \max_{t \in (t_0, t_1]} x_t^\top \beta$ in BIDS algorithm and $a = \min_{t \in (t_0, t_1]} \min_{1 \leq j \leq d} x_{tj}$ and $b = \max_{1 \leq j \leq d} x_{tj}$ in BaSEDB algorithm. The interval was then set as $[\frac{a+b}{2} - \frac{C(b-a)}{2}, \frac{a+b}{2} + \frac{C(b-a)}{2}]$ with $C = 1.2$.

Results. We run each algorithm 20 times and report the average regret for each of them over the 20 runs in Figures 4 and 5, for the two settings, respectively. The batch end points are denoted by the vertical solid black (SIR) and dashed blue lines (nonparametric) in Figures 4 and 5.

Case I (given pilot directions with varying accuracies) In this set-up, we compare

the performance of BIDS and BaSEDB when a pilot direction is available with varying levels of accuracies. Specifically, we set the initial index parameters β for the BIDS algorithm so that $\theta = \angle\beta, \beta_0 \in \{0.01, 0.16, 0.31 \ldots, \pi/2\}$. The corresponding $\sin(\theta)$ ranges from 0 to 1, where, $\sin(\theta) = 0$ implies that β is identical to β_0 up to a sign change, and $\sin(\theta) = 1$ implies that the two vectors are orthogonal. Figure 4 presents the average regrets of the
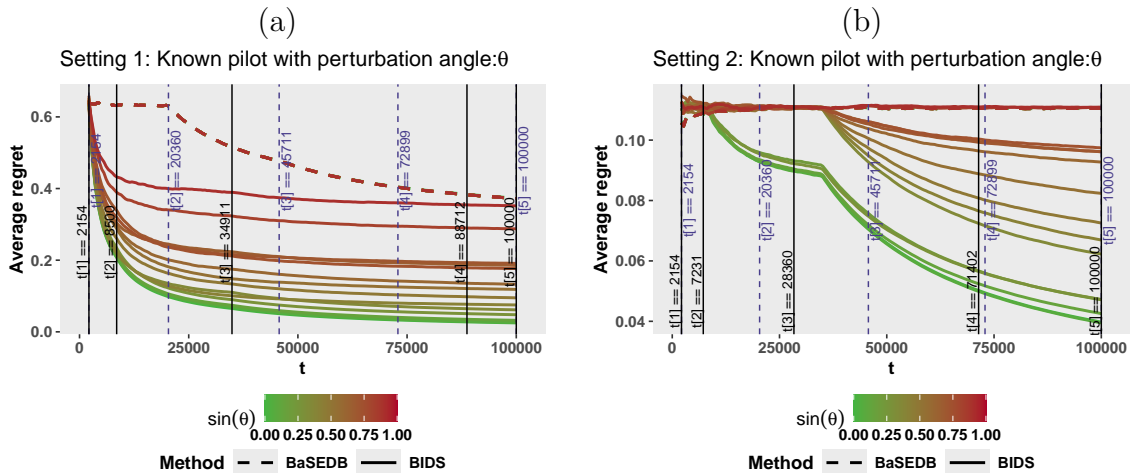
Figure 4: Average regret $((\mathcal{R}_t)_{t=1}^{T})$ with pilot directions β with varying accuracy, measured by $\sin\theta = \sin\angle\beta, \beta_0$ for the two simulation settings. Different colors of the solid lines represent different levels of perturbation, where $\sin\angle\beta, \beta_0 = 0$ corresponds to no perturbation, and $\sin\angle\beta, \beta_0 = 1$ corresponds to orthogonal vectors. As the degree of perturbation increases, performance deteriorates but still beats the nonparametric analogue.

BIDS algorithm with pilot directions of varying accuracies, compared to BaSEDB algorithm. As the perturbation level increases, the performance of the BIDS algorithm with the perturbed pilot estimate declines. However, it consistently outperforms the nonparametric batched bandit algorithm (BaSEDB), even under high perturbations. Interestingly, in Figure 4(b), we observe that in Setting 2—where the two mean reward functions exhibit greater overlap—the BaSEDB algorithm never eliminates an arm. Consequently, its average regret (dashed red line) does not decay over time. Moreover, once the perturbation angle exceeds $\pi/3$ and $\pi/4$, in Settings 1 and 2 respectively, the performance of the BIDS algorithm deteriorates to the level of its nonparametric counterpart.

Case II (no pilot directions) For the case when the pilot estimate is not available, in Figure 5, we assess the performance of these algorithms for varying degrees of model noise, σ. In the simulations, we also included BIDS (oracle), which refers to the BIDS algorithm where the true β_0 is used as the initial direction.
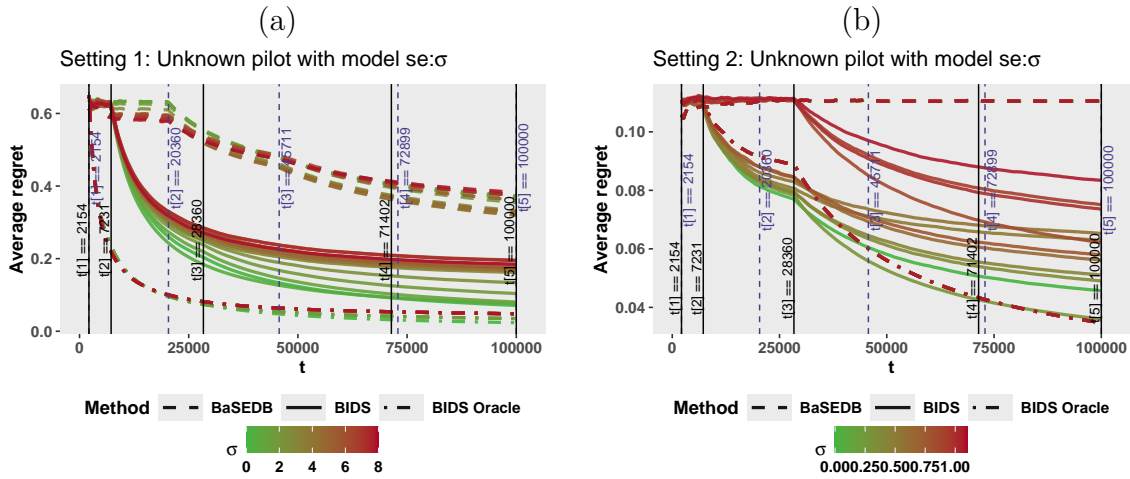
Figure 5: Average regret $((\mathcal{R}_t)_{t=1}^T)$ with varying model noise σ for the two simulation settings. As the noise level increases, while the performance of the BIDS algorithm (solid) remains better than the nonparametric analogue (dashed), but deviates further from the BIDS oracle (dashed-dotted).

Note that in setting 1, the two mean reward functions are well-separated, while in setting 2, they have more of an overlap in various regions. Therefore, even with higher model error in setting 1, it is easier to maintain low regret as can be seen in Figure 5(a). We consider the standard deviation to be ranging from $\sigma \in \{1, 2, \ldots, 8\}$ for setting 1 while $\sigma \in \{0.02, 0.09, 0.16, \ldots, 1\}$ for setting 2. From Figure 5, we see that in both settings, the BIDS algorithm appears to perform better than the BaSEDB algorithm for all the noise variance levels. As expected, the performance of the BIDS algorithm (solid) as compared to the oracle BIDS algorithm (dotted-dashed) deteriorates as the noise grows, as the higher noise levels reduce the accuracy of the initial direction vectors.

Remark 5 (computation considerations). In terms of computation, the GMABC framework and the BIDS algorithm have a key advantage over the BaSEDB algorithm, as the number of bins that needs to be tracked does not grow with the covariate dimension. In contrast, the number of bins in BaSEDB algorithm grows exponentially with the covariate dimension, making implementation challenging even for moderately large dimensions.

6 Application to Real Data

We compare the performance of the batched single-index and batched nonparametric BaSEDB algorithm on three publicly available real datasets:

a) Rice classification [Cinar and Koklu, 2019]: Classifying rice into two common varieties in Turkey, namely, Cammeo and Ormancik, using 7 morphological features extracted from 3810 rice grain's images.

b) Occupancy Detection [Candanedo, 2016]: Experimental data used for binary classification (room occupancy) from Temperature, Humidity, Light and CO_2.

c) EEG Eye State [Roesler, 2013]: This dataset contains instances of EEG measurements where the output is whether eye was open or not. The features correspond to 14 EEG measurements from the headset, labeled AF3, F7, F3, FC5, T7, P, O1, O2, P8, T8, FC6, F4, F8, AF4.

All these datasets involve classification tasks using some features. Accordingly, we take the number of decisions K to be the number of classes and consider a binary reward, which is 1 if we select the correct class and 0 otherwise. The dimension of the features for datasets (a)–(c) is 7, 5, and 14, with two arms each, respectively. The number of rows/instances in (a)-(c) are 3809, 8143, and 14980, respectively, therefore we choose the number of batches to be 5,6, and 7 respectively.

Setup. We leverage supervised learning classification datasets to simulate contextual bandits learning (e.g., see Bietti et al. [2021]). In particular, let $(x_t, c_t) \in \mathbb{R}^d \times \{1, \ldots, K\}$ row in the dataset where x_t is the context and c_t is the true label for the tth instance. We consider this tth row as a contextual bandit instance with x_t as given to the bandit algorithm, and we only reveal a binary reward of the chosen action a_t to be 1 if it matches the true label c_t and 0 otherwise. Therefore, for arms $a_t \in \{1, \ldots, K\}$, we consider the model in (2) and its non-parametric analogue: $Y_t = g^{(a_t)}(X_t) + \epsilon_t$, where $Y_t \in \{0, 1\}$ based on whether the chosen arm is a correct match or not. Note, since we only observe one arm at a given instance t, we only observe the reward corresponding to the chosen arm a_t at that particular instance. Apart from comparing the nonparametric batched bandit (BaSEDB)

performance with the BIDS algorithm proposed in Algorithm 1, we also consider an oracle BIDS algorithm where we estimate the index parameter β_0 using the entire dataset, and then use that for sequential decision-making in the BIDS algorithm. We randomly permute the data 60 times and measure the average regret performance of the three algorithms.

Results. We plot the average regret (rolling fraction of incorrect decisions over 60 trials with randomly permuted rows) as a function of the number of instances (rows) seen thus far for the following algorithms:

1. Nonparametric batched bandit (BaSEDB algorithm) of Jiang and Ma [2024].

2. BIDS algorithm (Algorithm 1) with initial estimator as proposed in Algorithm 2.

3. BIDS algorithm with estimated 'oracle' index, where we estimate oracle direction by using the entire dataset and applying Algorithm 3 on it.
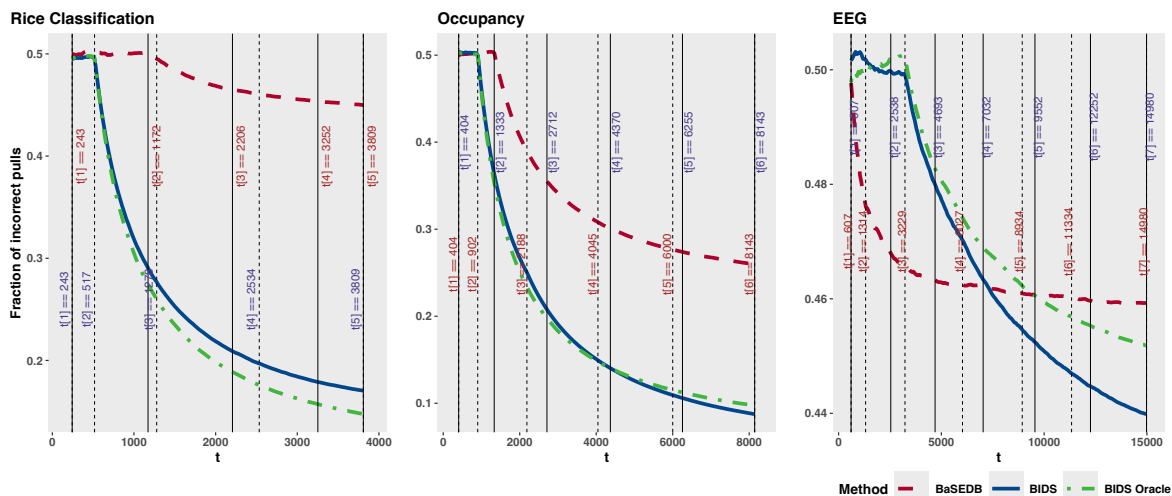

Figure 6: Comparison of expected regret of the proposed semiparametric BIDS algorithm and the nonparametric batched bandit algorithm (BaSEDB) on a) rice classification, b) occupancy detection, and c) EEG datasets, with β_0 estimated in the initial phase with $t_{\text{init}} \approx T^{2/3}$ for their respective data lengths T. Vertical solid and dashed lines denote the batch markings for the BIDS and BaSEDB algorithm, respectively. Observe that the BIDS outperforms BaSEDB in all instances, and for the Occupancy and EEG dataset it even performs similar/better to the BIDS oracle algorithm.

In Figure 6, we notice that in all three datasets, the BIDS algorithm that we propose outperforms the nonparametric batched bandit (BaSEDB) algorithm of Jiang and Ma [2024]. We use $t_{\text{init}} = T^{2/3}$ for each of the datasets. The vertical solid and dashed lines

represent the batch end points for the GMABC and the nonparametric setup, respectively. In the Occupancy dataset, BIDS achieves performance comparable to the BIDS oracle algorithm. In the EEG dataset, although BaSEDB initially experiences a steep decline in regret, it eventually plateaus, whereas the regret for BIDS continues to decrease at a faster rate, surpassing BaSEDB after a certain point. To assess the effect of the initial sample size used for estimating the index parameter β_0, we compare performance across different values of t_{init} in Section S6 of the Supplementary material. The observed trends remain consistent: BIDS outperforms the nonparametric batched analogue across all three datasets. However, as the initial sample size increases, the average regret of BIDS approaches that of the oracle BIDS algorithm.

	Rice Classification ($t_{\text{init}} = 243$)	Occupancy ($t_{\text{init}} = 404$)	EEG ($t_{\text{init}} = 607$)
β_1	Area: 0.0279 (0.0206)	Temp: 0.8326 (0.0817)	AF3: 0.0712 (0.0315)
β_2	Perimeter: -0.2979 (0.0247)	Humidity: -0.0036 (0.0046)	F7: 0.2979 (0.0266)
β_3	MajorAxis: 0.4990 (0.0409)	Light: -0.0769 (0.0083)	F3: 0.2088 (0.0387)
β_4	MinorAxis: -0.8085 (0.0762)	CO_2: -0.1310 (0.0151)	FC5: 0.3310 (0.0170)
β_5	Eccentricity: 0.0446 (0.0185)	HumidRatio: 0.5327 (0.0782)	T7: 0.1372 (0.0638)
β_6	Convex Area: 0.0748 (0.0215)		P7: 0.4034 (0.0512)
β_7	Extent: 0.0093 (0.0234)		O1: 0.2244 (0.0219)
β_8			O2: 0.1807 (0.0236)
β_9			P8: 0.3290 (0.0288)
β_{10}			T8: 0.0832 (0.0304)
β_{11}			FC6: 0.2663 (0.0183)
β_{12}			F4: 0.3146 (0.0314)
β_{13}			F8: 0.3213 (0.0199)
β_{14}			AF4: 0.3164 (0.0266)

Table 2: Index parameter estimates used in the BIDS algorithm for the three datasets.

Interpretability. In Table 2, we present the index parameter estimates for the three datasets when using $t_{\text{init}} = 243$, 404, and 607 ($\approx T^{(2/3)}$) each, respectively. We display the estimate β_i (with standard errors over the 60 replications in the parenthesis) for each $i = 1, \ldots, d$, for $d = 7, 5, 14$, for the three datasets, respectively. We can infer about the relevance of a variable by the absolute value of the corresponding estimates and the largest four values for each dataset are colored in blue. For example, in the Occupancy dataset, we note that the temperature, humidity ratio, light, and CO_2 levels all seem to be significant in predicting whether a room is occupied or not. This is consistent with other findings such as Khan and Rahman [2021]. Similarly, in the Rice Classification example, other supervised learning methods studied in Cinarer et al. [2024] suggest that Extent may not

be a useful variable in classifying rice into Cammeo and Osmancık rice types. Research on the EEG Eye State dataset has identified key features that contribute to distinguishing between eye-open and eye-closed states based on EEG signals. These features are derived from the 14 electrode channels and the ones significant in Table 2 (such as FC5 from top left, P7 from bottom left, P8 from bottom right and, F8 from top right) seem to well represent all the four regions as can be seen from Figure 2 in [Rösler and Suendermann, 2013]. Channels located on the right hemisphere (e.g., O2, P8, and F8) often show larger maximum values for eye-open states, while left-hemisphere channels (e.g., F7, P7, and T7) demonstrate other distinguishing patterns and we see that these appear to be relevant for us as well [Rösler and Suendermann, 2013, Asquith and Ihshaish, 2019].

7   Conclusion

The increasing availability of data streams in domains such as healthcare highlights the growing need for reliable, interpretable, and meaningful sequential decision-making methods that can adapt as data accumulates. In many cases, decisions must be made in batches, as seen in clinical trials. Batched multi-armed bandits with covariates (MABC) offer a natural statistical framework for these scenarios, enabling decision-makers to learn the reward-generating mechanisms of various arms in terms of their dependence on covariates or contexts at the end of each batch. We propose to study a novel Batched Global MABC framework, where we leverage the dependence between arms using a shared global parameter. This is unlike the usual MABC framework where each arm's reward generating mechanism is independent of each other.

Existing work in parametric and nonparametric bandits presents a trade-off: parametric methods provide interpretability under restrictive model assumptions, while nonparametric methods allow modeling of complex relationships but suffer from interpretability challenges and the curse of dimensionality. To address this gap, we introduced a semi-parametric framework using the single-index model, where rewards are modeled in a projected one-dimensional space (parametric shared component) with an unknown link function (nonparametric component). By estimating the single-index direction using established techniques like sliced inverse regression or average derivative estimation, the proposed BIDS algorithm

employs a dynamic binning and successive arm elimination strategy, where the binning of the covariate space is determined by the single-index direction.

We provided rigorous theoretical regret analysis under two scenarios: (1) when the single-index direction is known a priori and (2) when it is estimated from initial data. In the former case, the regret rate matches the minimax optimal rate for nonparametric batched bandits in various batch regimes, but with the effective dimension reduced to $d = 1$, circumventing the curse of dimensionality. In the latter scenario, the same minimax rate can be achieved with $d = 1$, given additional assumptions on smoothness and margin parameters. Simulated and real-world datasets demonstrated the effectiveness of the proposed framework compared to nonparametric bandit algorithms, while also offering interpretable insights into covariate importance through the single-index parameter estimates. In addition to providing better practical performance and interpretability, we also significantly reduce the computation cost as compared to the nonparametric batched bandit based on binning and successive arm elimination.

To the best of our knowledge, this study represents the first exploration of a single-index framework in contextual batched bandits, opening avenues for future research. An immediate open question involves deriving the lower bound for cumulative regret when the single-index direction is unknown. We hypothesize that the regret rate lies between the nonparametric and parametric rates, but this requires formal proof. Another interesting future direction is to estimate the single-index direction adaptively with batches by leveraging the margin condition, especially for settings with moderate to large number of batches. Lastly, providing an adaptive algorithm that learns the smoothness of the underlying mean reward functions, would enhance applicability of the proposed algorithm. The development of this framework lays the groundwork for addressing such questions and further advancing interpretable and flexible sequential decision-making strategies.

8   Acknowledgment

Supplement to "Semi-Parametric Batched Global Multi-Armed Bandits with Covariates"

Sakshi Arya and Hyebin Song

MAMS, Case Western Reserve University and

Department of Statistics, Pennsylvania State University

S1   Proofs for Section 2

S1.1   A summary table of notations

First, to enhance readability, in Table S1, we provide a table of notations that are used in the paper and the proofs presented in this section.

S1.2   Proof for Lemma 1

Proof. For any v, the density of $X^\top v$ is given by

$$
f_{X^\top v}(u) = \begin{cases} \frac{1}{Z(v,\Sigma)} \exp\{-\frac{u^2}{2v^\top \Sigma v}\} & x \in \mathcal{T}_v \\ 0 & \text{otherwise} \end{cases}
$$

where we define $\mathcal{T}_v := \{x^\top v; v \in \mathcal{H}\}$ and $Z(v, \Sigma) := \int_{u \in \mathcal{T}_v} \exp\{-\frac{u^2}{2v^\top \Sigma v}\} du$.

Let a unit vector v be given such that $\|v\|_2 = 1$. First of all, we observe that $\mathcal{T}_v$ is an interval in $\mathbb{R}$. Note that $\mathcal{H}$ is a closed, convex set in $\mathbb{R}^d$. We can find $x_0(v), x_1(v) \in \mathcal{H}$ such that $x_0(v)^\top v = \min_{x \in \mathcal{H}} x^\top v := L_0(v)$ and $x_1(v)^\top v = \max_{x \in \mathcal{H}} x^\top v := L_1(v)$. Moreover, since the dual of the ℓ_∞-norm is the ℓ_1-norm, $L_0(v) = -\|v\|_1$ and $L_1(v) = \|v\|_1$. Now we show for any $u \in [L_0(v), L_1(v)]$, $u \in \mathcal{T}_v$. Since $u \in [L_0(v), L_1(v)]$, we can find $t \in [0, 1]$ such that $u = tL_0(v) + (1-t)L_1(v)$. Then $u = tx_0(v)^\top v + (1-t)x_1(v)^\top v = \{tx_0(v) + (1-t)x_1(v)\}^\top v$. By convexity of $\mathcal{H}$, $tx_0(v) + (1-t)x_1(v) \in \mathcal{H}$, and therefore $u \in \mathcal{T}_v$, which shows that $\mathcal{T}_v = [L_0(v), L_1(v)] \subseteq \mathbb{R}$.

Now let $R_0 = \|\beta_0\|_1/(2\sqrt{d})$. Let $v \in \mathbb{B}_2(R_0; \beta_0)$ be given such that $\|v\|_2 = 1$. We show that for any $u \in \mathcal{T}_v$, the density $f_{X^\top v}(u)$ is bounded below and above by constants $\underline{c}_X$

Category	Notation	Description
Problem setup	T	Total time horizon
	K	Number of arms
	M	Number of batches
	$\mathcal{X}$	Covariate space in $\mathbb{R}^d$
	$\mathcal{G}$	Partition of $\{1, \ldots, T\}$ in M batches
	$\{t_0, t_1, \ldots, t_M\}$	Batch end points
	$R_T(\pi)$	Cumulative regret of π
	$\mathcal{R}_T(\pi)$	Expected cumulative regret of π
	$\angle u, v$	Principal angle between u and v: $\cos^{-1}(|u^\top v|)$
Parameters	β_0	Index parameter
	η	Smoothness parameter in $(0, 1]$
	α	Margin parameter
	$\{\omega_k\}_{k=1}^K$	Weights for the average estimator
Algorithmic and Theory	π	Proposed BIDS algorithm
	β	Working direction
	$\mathcal{I}_\beta := [L_\beta, U_\beta]$	Interval of projected covariates along β
	t_{init}	Initial batch size used when pilot unknown
	$\hat{\beta}^{(k)}$	Single index estimate for kth arm
	$\hat{\beta}$	Initial index estimate of β_0
	$\mathcal{T}$	Tree of depth M
	$\mathcal{A}_i$	Partition of $\mathcal{I}_\beta = [L_\beta, U_\beta]$ at layer i
	$w_i = |\mathcal{I}_\beta|/n_i$	Bin width for ith layer
	b_l	Number of splits in layer l
	n_i	Number of equal width intervals in layer i
	$\mathcal{T}_{\mathcal{A}}$	$\cup_{i=1}^M \mathcal{A}_i$
	$\mathcal{B}_i$	Partition of $\mathcal{X}$ induced by $\mathcal{A}_i$
	$C = C_A(\beta)$	Bin in $\mathcal{X}$ corresponding to $A \in \mathcal{T}_{\mathcal{A}}$
	$|C|_{\mathcal{T}}$	width of A for $C = C_A(\beta)$
	$p(C) = p(C_A(\beta))$	Parent bin of C defined by A
	$\text{child}(C)$	Child bin of C defined by A
	$\mathcal{L}_t, \mathcal{L}^{(i)}$	Set of active bins at time t/at batch i
	$\mathcal{J}_t$	$\cup_{s \leq t} \mathcal{L}_s$
	$\mathcal{I}_C$	Set of active arms in bin C
	$\mathcal{I}'_C$	Set of active arms post arm-elimination in C
	$\underline{\mathcal{I}}_C, \overline{\mathcal{I}}_C, \mathcal{S}_C, \mathcal{G}_C$	Sets defined in (S-2), (S-5), (S-4)
	$U(m, T, C)$	Threshold for arm elimination
	$m_{C,i}$	number of X_t's falling in C during batch i
	$m^*_{C,i}$	$\mathbb{E}[m_{C,i}]$
	SIR	Single-index regression
	$\xi, c_B, R_X, \bar{c}_X, \underline{c}_X, L_0, D_0$	Constants independent of T.

Table S1: Extensive summary of Notations Used in the Paper

and $\bar{c}_X$, which depend on model parameters β_0 and Σ, but independent of v. Recall that $L_0(v) = -\|v\|_1$ and $L_1(v) = \|v\|_1$. Since $|\|v\|_1 - \|\beta_0\|_1| \le \|v - \beta_0\|_1 \le \sqrt{d}R_0$, $|L_0(v) - L_0(\beta_0)| \le \sqrt{d}R_0$. Similarly, $|L_1(v) - L_1(\beta_0)| \le \sqrt{d}R_0$. In particular, $[L_0(\beta_0)/2, L_1(\beta_0)/2] \subseteq [L_0(v), L_1(v)] \subseteq [1.5L_0(\beta_0), 1.5L_1(\beta_0)]$. We let

$$\underline{\mathcal{T}_0} := [L_0(\beta_0)/2, L_1(\beta_0)/2], \overline{\mathcal{T}_0} := [(3/2)L_0(\beta_0), (3/2)L_1(\beta_0)],$$

so that

$$\underline{\mathcal{T}_0} \subseteq \mathcal{T}_v \subseteq \overline{\mathcal{T}_0}.$$

Since $\|v\|_2 = 1$, $\Lambda_{\min}(\Sigma) \le v^\top \Sigma v \le \Lambda_{\max}(\Sigma)$. First, recall $Z(v, \Sigma) = \int_{u \in \mathcal{T}_v} \exp\{-\frac{u^2}{2v^\top \Sigma v}\}du$. We have,

$$Z(v, \Sigma) = \int_{u \in \mathcal{T}_v} \exp\left\{-\frac{u^2}{2v^\top \Sigma v}\right\} du \ge \int_{u \in \underline{\mathcal{T}_0}} \exp\left\{-\frac{u^2}{2\Lambda_{\min}(\Sigma)}\right\} du := \underline{c}_Z$$

Similarly, we have

$$Z(v, \Sigma) \le \int_{u \in \overline{\mathcal{T}_0}} \exp\left\{-\frac{u^2}{2\Lambda_{\max}(\Sigma)}\right\} du := \bar{c}_Z$$

Then for $u \in \mathcal{T}_v$,

$$\frac{1}{\bar{c}_Z} \inf_{u \in \overline{\mathcal{T}_0}} \exp\left\{-\frac{u^2}{2\Lambda_{\min}(\Sigma)}\right\} \le \frac{1}{Z(v, \Sigma)} \exp\left\{-\frac{u^2}{2v^\top \Sigma v}\right\} \le \frac{1}{\underline{c}_Z} \sup_{u \in \overline{\mathcal{T}_0}} \exp\left\{-\frac{u^2}{2\Lambda_{\max}(\Sigma)}\right\},$$

$$(\text{S-1})$$

and we can take $\underline{c}_X = \frac{1}{\bar{c}_Z} \inf_{u \in \overline{\mathcal{T}_0}} \exp\{-\frac{u^2}{2\Lambda_{\min}(\Sigma)}\}$ and $\bar{c}_X = \frac{1}{\underline{c}_Z} \sup_{u \in \overline{\mathcal{T}_0}} \exp\{-\frac{u^2}{2\Lambda_{\max}(\Sigma)}\}$. $\qquad\square$

S2   Proofs for Section 4

S2.1   Proof of Theorem 1

Proof. First we construct two events to capture the elimination process. Let the batch index $i = 1, \ldots, M$ be fixed. For each bin $C \in \mathcal{B}_i$, we define a "good batch elimination event", $\mathcal{S}_C$, associated with C. Note that C may or may not have been born at the

beginning of batch i, and only undergoes the unique batch elimination event if it was born in the beginning of batch i, i.e., when $C \in \mathcal{L}^{(i)}$ (also ref. Remark 2). If $C \notin \mathcal{L}^{(i)}$, simply let $\mathcal{S}_C = \Omega$ where Ω is the whole probability space. When $C \in \mathcal{L}^{(i)}$, let $\mathcal{I}_C$ and $\mathcal{I}'_C$ denote the set of active arms associated with C during batch i and end of batch i after batch elimination process, respectively. Note $|\mathcal{I}'_C| > 1$ will trigger splitting C into its children sets. Define

$$\underline{\mathcal{I}}_C = \left\{ k \in \{1,2\} : \sup_{x \in C}\{f^{(*)}(x^\top \beta_0) - f^{(k)}(x^\top \beta_0)\} \leq c_0|C|_{\mathcal{T}}^\eta \right\}, \tag{S-2}$$

$$\overline{\mathcal{I}}_C = \left\{ k \in \{1,2\} : \sup_{x \in C}\{f^{(*)}(x^\top \beta_0) - f^{(k)}(x^\top \beta_0)\} \leq c_1|C|_{\mathcal{T}}^\eta \right\}, \tag{S-3}$$

for $c_0 = 4L_0 + 1$ with $L_0 = L(2C_0 R_X + 1)^\eta$, $c_1 = 8c_0\gamma_X^{1/2}$ where $\gamma_X = \bar{c}_X/\underline{c}_X$, and

$$f^{(*)}(x^\top \beta_0) = \max_{k \in \{1,2\}} f^{(k)}(x^\top \beta_0).$$

Note that, $\underline{\mathcal{I}}_C \subseteq \overline{\mathcal{I}}_C$. Define a 'good event':

$$\mathcal{S}_C = \{\underline{\mathcal{I}}_C \subseteq \mathcal{I}'_C \subseteq \overline{\mathcal{I}}_C\}. \tag{S-4}$$

This is a good event because it says that all good arms (with small regret) survive the stage i elimination, and all survived arms in $\mathcal{I}'_C$ have not so large regret. In addition, define

$$\mathcal{G}_C = \cap_{C' \in \mathcal{P}(C)} \mathcal{S}_{C'}, \tag{S-5}$$

which is the event where the elimination processes were "good" for all ancestors of C. In the special case when C has no parent since $C \in \mathcal{B}_1$, simply let $\mathcal{G}_C = \Omega$.

We decompose the regret into three terms. Recall that $\mathcal{L}_t$ is the set of active bins at t. Also, we define $\mathcal{J}_t := \cup_{s \leq t}\mathcal{L}_s$ for all the bins that were alive at some time point $s \leq t$. First for a bin $C \in \mathcal{T}$, we define:

$$r_T^l(C) := \sum_{t=1}^T \{g^{(*)}(X_t) - g^{(\pi_t(X_t))}(X_t)\}1(X_t \in C)1(C \in \mathcal{L}_t),$$

which is the amount of regret on C when C is "alive", and also define:

$$r_T^b(C) := \sum_{t=1}^{T}(g^*(X_t) - g^{(\pi_t(X_t))}(X_t))1(X_t \in C)1(C \in \mathcal{J}_t),$$

which is the amount of regret on C since C was "born".

There exists a recursive relationship between $r_T^l(C)$ and $r_T^b(C)$, as introduced in Perchet and Rigollet [2013]. We present this relationship as Lemma S-1 for the convenience of readers and give a proof for the same in Section S2.2.

Lemma S-1. *For $C \in \mathcal{B}_i$, for $i = 1, \ldots, M$, we have*

$$r_T^b(C) = r_T^l(C) + \sum_{C' \in child(C)} r_T^b(C'), \tag{S-6}$$

where we adopt the convention that $\sum_{C \in \emptyset} r_T^b(C) = 0$. In particular, $\sum_{C' \in child(C)} r_T^b(C') = 0$ if $C \in \mathcal{B}_M$.

From Lemma S-1, trivially we obtain,

$$r_T^b(C) = \left\{ r_T^l(C) + \sum_{C' \in child(C)} r_T^b(C') \right\} 1(\mathcal{S}_C) + r_T^b(C)1(\mathcal{S}_C^c)$$

$$= r_T^l(C)1(\mathcal{S}_C) + r_T^b(C)1(\mathcal{S}_C^c) + \sum_{C' \in child(C)} r_T^b(C')1(\mathcal{S}_C) \tag{S-7}$$

Additionally, we can have the following iterative relationship:

$$\sum_{C \in \mathcal{B}_i} \sum_{C' \in child(C)} r_T^b(C')1(\mathcal{S}_C)1(\mathcal{G}_C) \tag{S-8}$$

$$= \sum_{C \in \mathcal{B}_i} \sum_{C' \in child(C)} \left\{ r_T^l(C')1(\mathcal{S}_{C'}) + r_T^b(C')1(\mathcal{S}_{C'}^c) + \sum_{C'' \in child(C')} r_T^b(C'')1(\mathcal{S}_{C'}) \right\} 1(\mathcal{S}_C)1(\mathcal{G}_C)$$

$$= \sum_{C' \in \mathcal{B}_{i+1}} \{ r_T^l(C')1(\mathcal{S}_{C'}) + r_T^b(C')1(\mathcal{S}_{C'}^c) \}1(\mathcal{G}_{C'}) + \sum_{C' \in \mathcal{B}_{i+1}} \sum_{C'' \in child(C')} r_T^b(C'')1(\mathcal{S}_{C'})1(\mathcal{G}_{C'})$$

using the fact that $1(\mathcal{S}_C)1(\mathcal{G}_C) = 1(\mathcal{G}_{C'})$ for $C' \in child(C)$.

Using (S-7) and applying (S-8) iteratively, and using the fact that $\mathcal{G}_C = \Omega$ for $C \in \mathcal{B}_1$,

we have:

$$R_T(\pi) = \sum_{C \in \mathcal{B}_1} r_T^b(C)$$

$$= \sum_{C \in \mathcal{B}_1} r_T^l(C)1(\mathcal{S}_C)1(\mathcal{G}_C) + \sum_{C \in \mathcal{B}_1} r_T^b(C)1(\mathcal{S}_C^c)1(\mathcal{G}_C) + \sum_{C \in \mathcal{B}_1} \sum_{C' \in \text{child}(C)} r_T^b(C')1(\mathcal{S}_C)1(\mathcal{G}_C)$$

$$= \sum_{i=1}^{2} \sum_{C \in \mathcal{B}_i} \{r_T^l(C)1(\mathcal{S}_C) + r_T^b(C)1(\mathcal{S}_C^c)\}1(\mathcal{G}_C) + \sum_{C \in \mathcal{B}_2} \sum_{C' \in \text{child}(C)} r_T^b(C')1(\mathcal{S}_C)1(\mathcal{G}_C)$$

$$\cdots = \sum_{i=1}^{M-1} \sum_{C \in \mathcal{B}_i} \{r_T^l(C)1(\mathcal{S}_C) + r_T^b(C)1(\mathcal{S}_C^c)\}1(\mathcal{G}_C) + \sum_{C \in \mathcal{B}_{M-1}} \sum_{C' \in \text{child}(C)} r_T^b(C')1(\mathcal{S}_C)1(\mathcal{G}_C)$$

$$= \sum_{i=1}^{M-1} \sum_{C \in \mathcal{B}_i} \{r_T^l(C)1(\mathcal{S}_C) + r_T^b(C)1(\mathcal{S}_C^c)\}1(\mathcal{G}_C) + \sum_{C \in \mathcal{B}_M} r_T^b(C)1(\mathcal{G}_C).$$

Define the event that we obtain sufficient samples for all C in $\mathcal{B}_i$ for $1 \leq i \leq M-1$:

$$\mathcal{E} := \{\forall C \in \cup_{i=1}^{M-1}\mathcal{B}_i, m_{C,i} \in [m_{C,i}^\star/2, 3m_{C,i}^\star/2]\} \tag{S-9}$$

We have

$$R_T(\pi) = R_T(\pi)1(\mathcal{E}^c) + R_T(\pi)1(\mathcal{E})$$

Moreover, for a set $C \in \mathcal{T}$, if C has never been born (i.e., if $C \notin \mathcal{J}_T \iff C \notin \mathcal{L}_t$ for all $1 \leq t \leq T$), $r_T^l(C) = r_T^b(C) = 0$. Therefore,

$$R_T(\pi)1(\mathcal{E}) = \sum_{i=1}^{M-1} \sum_{C \in \mathcal{B}_i \cap \mathcal{J}_T} r_T^l(C)1(\mathcal{S}_C \cap \mathcal{G}_C \cap \mathcal{E}) + \sum_{i=1}^{M-1} \sum_{C \in \mathcal{B}_i \cap \mathcal{J}_T} r_T^b(C)1(\mathcal{S}_C^c \cap \mathcal{G}_C \cap \mathcal{E})$$

$$+ \sum_{C \in \mathcal{B}_M \cap \mathcal{J}_T} r_T^b(C)1(\mathcal{G}_C \cap \mathcal{E})$$

$$\leq \sum_{i=1}^{M-1} \sum_{C \in \mathcal{B}_i \cap \mathcal{J}_T} r_T^l(C)1(\mathcal{S}_C \cap \mathcal{G}_C) + \sum_{i=1}^{M-1} \sum_{C \in \mathcal{B}_i \cap \mathcal{J}_T} r_T^b(C)1(\mathcal{S}_C^c \cap \mathcal{G}_C \cap \mathcal{E})$$

$$+ \sum_{C \in \mathcal{B}_M \cap \mathcal{J}_T} r_T^b(C)1(\mathcal{G}_C).$$

Let, for $i = 1, \ldots, M - 1$,

$$U_i := \sum_{C \in \mathcal{B}_i \cap \mathcal{J}_T} r_T^l(C) 1(\mathcal{S}_C \cap \mathcal{G}_C), \quad V_i := \sum_{C \in \mathcal{B}_i \cap \mathcal{J}_T} r_T^b(C) 1(\mathcal{S}_C^c \cap \mathcal{G}_C \cap \mathcal{E}),$$

and $W_M =: \sum_{C \in \mathcal{B}_M \cap \mathcal{J}_T} r_T^b(C) 1(\mathcal{G}_C)$ so that

$$R_T(\pi) 1(\mathcal{E}) \leq \sum_{i=1}^{M-1} (U_i + V_i) + W_M. \tag{S-10}$$

Next, we bound these three terms, namely, U_i, V_i and W_M separately.

Controlling U_i. Let us fix some batch i, $1 \leq i \leq M - 1$, and some bin $C \in \mathcal{B}_i \cap \mathcal{J}_T$. Recall that by definition of $\mathcal{B}_i$, $C = C_A(\beta)$ for some $A \in \mathcal{A}_i$, where $A \subseteq [L_\beta, U_\beta]$ is an interval of length w_i. By definition of $r_T(C)$,

$$\mathbb{E}[r_T^l(C) 1(\mathcal{G}_C \cap \mathcal{S}_C)]$$
$$= \mathbb{E}\left[\sum_{t=1}^{T} \{g^*(X_t) - g^{(\pi_t(X_t))}(X_t)\} 1(X_t \in C) 1(C \in \mathcal{L}_t) 1(\mathcal{G}_C \cap \mathcal{S}_C)\right].$$

We show that the summand is non-zero only for $t \in [t_{i-1} + 1, t_i]$: First, since $C \in \mathcal{B}_i$, $C \notin \mathcal{L}_t$ for $t \leq t_{i-1}$, i.e., $1(C \in \mathcal{L}_t) = 0$ for $t \leq t_{i-1}$. This is because $C \in \mathcal{B}_i$ can only born at the beginning of batch i, that is when $t = t_{i-1} + 1$. Now consider $t > t_i$. At the end of batch i, there are two possibilities: 1. no arms are eliminated (i.e., $|\mathcal{I}_C'| > 1$) : in this case, C is split into its children, and $C \notin \mathcal{L}_t$ for $t > t_i$. 2. one arm is eliminated ($|\mathcal{I}_C'| = 1$): we argue that on $\mathcal{S}_C$, the remaining arm is optimal for all $x \in C$, and therefore $g^*(x) - g^{(\pi_t(x))}(x) = 0$ for $t > t_i$, where we recall that $\pi_t(x)$ is the arm chosen for x by the algorithm. Let $k_1 \in \{1, 2\}$ be the eliminated arm and $k_2 \in \{1, 2\}$ be the remaining arm. On $\mathcal{S}_C$, we have $\underline{\mathcal{I}}_C \subseteq \mathcal{I}_C' = \{k_2\} \subseteq \overline{\mathcal{I}}_C$, therefore $k_1 \notin \underline{\mathcal{I}}_C$. Then, there exists $x_0 \in C$ such that $g^{(k_2)}(x_0) - g^{(k_1)}(x_0) > c_0 |C|_{\mathcal{T}}^\eta$. For any $x \in C$,

$$g^{(k_2)}(x) - g^{(k_1)}(x) \geq g^{(k_2)}(x_0) - g^{(k_1)}(x_0) - \sum_{k \in \{1,2\}} |g^{(k)}(x) - g^{(k)}(x_0)|.$$

By Lemma S-4, for sufficiently large T, $|g^{(k)}(x) - g^{(k)}(x_0)| \leq L_0 w_i^\eta$ for $k \in \{1, 2\}$, and therefore

$$g^{(k_2)}(x) - g^{(k_1)}(x) \geq (c_0 - 2L_0)w_i^\eta = (2L_0 + 1)w_i^\eta > 0,$$

recalling that $c_0 = 4L_0 + 1$. Therefore k_2 is the optimal arm for all $x \in C$. In particular, regret is not incurred for $t > t_i$, i.e., $g^*(X_t) - g^{(\pi_t(X_t))}(X_t) = 0$ for $X_t \in C$, $t > t_i$.

Therefore,

$$\mathbb{E}[r_T^l(C)1(\mathcal{G}_C \cap \mathcal{S}_C)]$$
$$= \mathbb{E}\left[\sum_{t=t_{i-1}+1}^{t_i} \{g^*(X_t) - g^{(\pi_t(X_t))}(X_t)\}1(X_t \in C)1(C \in \mathcal{L}_t)1(\mathcal{G}_C \cap \mathcal{S}_C) \right].$$

On the event $\mathcal{G}_C$, we have that $\mathcal{I}'_{p(C)} \subseteq \overline{\mathcal{I}}_{p(C)}$, that is, for any $k \in \mathcal{I}'_{p(C)}$,

$$\sup_{x \in p(C)} \{g^{(*)}(x) - g^{(k)}(x)\} \leq c_1 |p(C)|_{\mathcal{T}}^\eta.$$

Moreover, regret is only incurred at points where $|g^{(1)}(x) - g^{(2)}(x)| > 0$. Therefore, on $\mathcal{G}_C$, for any $x \in C$ and $k \in \mathcal{I}'_{p(C)}$,

$$g^*(x) - g^{(k)}(x) \leq c_1 |p(C)|_{\mathcal{T}}^\eta 1(0 < |g^{(1)}(x) - g^{(2)}(x)| \leq c_1 |p(C)|_{\mathcal{T}}^\eta).$$

In particular, for any $X_t \in C$, the inequality

$$g^*(X_t) - g^{(\pi_t(X_t))}(X_t) \leq c_1 |p(C)|_{\mathcal{T}}^\eta 1(0 < |g^{(1)}(X_t))) - g^{(2)}(X_t)| \leq c_1 |p(C)|_{\mathcal{T}}^\eta) \qquad \text{(S-11)}$$

holds on $\mathcal{G}_C$ when $t > t_{i-1}$, since for $t > t_{i-1}$, $\pi_t(X_t)$ can be selected from the (subset of) active arms after the $i-1$ batch elimination, and therefore $\pi_t(X_t) \in \mathcal{I}'_{p(C)}$. Therefore, we

obtain,

$$\mathbb{E}\left[\sum_{t=t_{i-1}+1}^{t_i} (g^*(X_t) - g^{(\pi_t(X_t))}(X_t))1(X_t \in C)1(C \in \mathcal{L}_t)1(\mathcal{G}_C \cap \mathcal{S}_C)\right]$$

$$\leq \sum_{t=t_{i-1}+1}^{t_i} c_1|p(C)|_{\mathcal{T}}^{\eta}\mathbb{E}\left[1(0 < |g^{(1)}(X_t) - g^{(2)}(X_t)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta})1(X_t \in C)\right.$$

$$1(C \in \mathcal{L}_t)1(\mathcal{G}_C \cap \mathcal{S}_C)]$$

$$\leq \sum_{t=t_{i-1}+1}^{t_i} c_1|p(C)|_{\mathcal{T}}^{\eta}\mathbb{P}\left(0 < |g^{(1)}(X_t) - g^{(2)}(X_t)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta}, X_t \in C\right)$$

$$= (t_i - t_{i-1})c_1|p(C)|_{\mathcal{T}}^{\eta}\mathbb{P}\left(0 < |g^{(1)}(X) - g^{(2)}(X)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta}, X \in C\right),$$

where the last equality is due to the fact that $X_t \sim \mathbb{P}_X$ iid. Finally,

$$\mathbb{E}[U_i] = \sum_{C \in \mathcal{B}_i \cap \mathcal{J}_T} \mathbb{E}[r_T^l(C)1(\mathcal{G}_C \cap \mathcal{S}_C)]$$

$$\leq \sum_{C \in \mathcal{B}_i \cap \mathcal{J}_T} (t_i - t_{i-1})c_1|p(C)|_{\mathcal{T}}^{\eta}\mathbb{P}\left(0 < |g^{(1)}(X) - g^{(2)}(X)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta}, X \in C\right)$$

$$\leq (t_i - t_{i-1})c_1|p(C)|_{\mathcal{T}}^{\eta}\mathbb{P}\left(0 < |g^{(1)}(X) - g^{(2)}(X)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta}\right),$$

where for the last equality we use the fact that $\mathcal{B}_i$ is the partition of $\mathcal{X}$. Since $|p(C)|_{\mathcal{T}} = w_{i-1}$ by the set-up and $\mathbb{P}\left(0 < |g^{(1)}(X) - g^{(2)}(X)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta}\right) \leq D_0\{c_1|p(C)|_{\mathcal{T}}^{\eta}\}^{\alpha}$ by the margin condition in Assumption 2, for $1 \leq i \leq M - 1$,

$$\mathbb{E}[U_i] \leq (t_i - t_{i-1})D_0\{c_1 w_{i-1}^{\eta}\}^{1+\alpha}. \tag{S-12}$$

Controlling V_i. Similarly, choose some $1 \leq i \leq M - 1$ and bin $C \in \mathcal{B}_i \cap \mathcal{J}_T$. We have $C = C_A(\beta)$ for some $A \in \mathcal{A}_i$. We have from definition of $r_T^b(C)$,

$$
\mathbb{E}[r_T^b(C)1(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E})]
$$

$$
= \mathbb{E}\left[\sum_{t=1}^{T}(g^*(X_t) - g^{(\pi_t(X_t))}(X_t))1(X_t \in C)1(C \in \mathcal{J}_t)1(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E})\right]
$$

$$
= \mathbb{E}\left[\sum_{t=t_{i-1}+1}^{T}(g^*(X_t) - g^{(\pi_t(X_t))}(X_t))1(X_t \in C)1(C \in \mathcal{J}_t)1(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E})\right]
$$

$$
\leq c_1|p(C)|_{\mathcal{T}}^{\eta}\mathbb{E}\left[\sum_{t=t_{i-1}+1}^{T}1(0 < |g^{(1)}(X_t) - g^{(2)}(X_t)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta}, X_t \in C)1(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E})\right],
$$

(S-13)

where for the second equality we use the fact that $C \notin \mathcal{J}_t$ for $t \leq t_{i-1}$, since $C \in \mathcal{B}_i$ can be born only at batch i and we use (S-11) for the last inequality.

We note that $\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E}$ is independent of $\{X_t; t > t_i\}$. This is because $\mathcal{G}_C = \cap_{C \in \mathcal{P}(C)}\mathcal{S}_C$, therefore it only depends on (random) batch elimination events up to $i - 1$ batch, i.e., $\mathcal{G}_C$ only depends on $\{(X_t, Y_t); 1 \leq t \leq t_{i-1}\}$, and $\mathcal{S}_C$ depends on batch elimination event at the end of batch i, and therefore depends on $\{(X_t, Y_t); t_{i-1} + 1 \leq t \leq t_i\}$. Therefore,

$$
\mathbb{E}\left[\sum_{t=t_{i-1}+1}^{T}1(0 < |g^{(1)}(X_t) - g^{(2)}(X_t)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta}, X_t \in C)1(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E})\right]
$$

$$
= \sum_{t=t_{i-1}+1}^{t_i}\mathbb{E}\left[1(0 < |g^{(1)}(X_t) - g^{(2)}(X_t)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta}, X_t \in C)1(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E})\right]
$$

$$
+ \sum_{t=t_i+1}^{T}\mathbb{P}\left[0 < |g^{(1)}(X_t) - g^{(2)}(X_t)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta}, X_t \in C\right]\mathbb{P}(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E})
$$

$$
\leq \sum_{t=t_{i-1}+1}^{t_i}\mathbb{P}\left[0 < |g^{(1)}(X_t) - g^{(2)}(X_t)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta}, X_t \in C\right]
$$

$$
+ \sum_{t=t_i+1}^{T}\mathbb{P}\left[0 < |g^{(1)}(X_t) - g^{(2)}(X_t)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta}, X_t \in C\right]\mathbb{P}(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E}),
$$

where for the last inequality we use $1(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E}) \leq 1$ a.s. Therefore, using this in (S-13)

S10

we obtain,

$$\mathbb{E}[r_T^b(C)1(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E})]$$

$$\leq c_1 |p(C)|_{\mathcal{T}}^{\eta}\{(t_i - t_{i-1}) + (T - t_i)\mathbb{P}(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E})\}$$

$$\times \mathbb{P}\left[0 < |g^{(1)}(X) - g^{(2)}(X)| \leq c_1 |p(C)|_{\mathcal{T}}^{\eta}, X \in C\right].$$

Therefore, using the fact that $\mathcal{B}_i$ is the partition of $\mathcal{X}$, and Assumption 2, we obtain:

$$\mathbb{E}[V_i] = \sum_{i \in \mathcal{B}_i \cap \mathcal{J}_T} \mathbb{E}[r_T^b(C)1(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E})]$$

$$\leq c_1 |p(C)|_{\mathcal{T}}^{\eta}\{(t_i - t_{i-1}) + (T - t_i)\mathbb{P}(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E})\}$$

$$\mathbb{P}\left[0 < |g^{(1)}(X) - g^{(-1)}(X)| \leq c_1 |p(C)|_{\mathcal{T}}^{\eta}\right]$$

$$\leq D_0\{c_1 w_{i-1}^{\eta}\}^{1+\alpha}\{(t_i - t_{i-1}) + (T - t_i)\mathbb{P}(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E})\}.$$

From Lemma S-5, we have that $P(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E}) \leq \dfrac{3m_{C,i}^*}{2T|C|_{\mathcal{T}}}$. Recalling the definition $m_{C,i}^* = \mathbb{E}[\sum_{t=t_{i-1}+1}^{t_i} 1\{X_t \in C\}] = (t_i - t_{i-1})P_X(C)$, we have

$$(T - t_i)\mathbb{P}(\mathcal{G}_C \cap \mathcal{S}_C^c \cap \mathcal{E}) \leq \frac{(T - t_{i-1})\{3\bar{c}_X(t_i - t_{i-1})|C|_{\mathcal{T}}\}}{T|C|_{\mathcal{T}}} \leq 3\bar{c}_X(t_i - t_{i-1}),$$

since $P_X(C) = P_X(C_A(\beta)) = P(X^{\top}\beta \in A) = \int_{u \in A} f_{x^{\top}\beta}(u)du \leq \bar{c}_X|A| = \bar{c}_X|C|_{\mathcal{T}}$ from Assumption 3. Then,

$$\mathbb{E}[V_i] \leq D_0\{c_1 w_{i-1}^{\eta}\}^{1+\alpha}(3\bar{c}_X + 1)(t_i - t_{i-1}). \tag{S-14}$$

Controlling W_M. Finally, for $C = C_A(\beta) \in \mathcal{B}_M \cap \mathcal{J}_T$ with $A \in \mathcal{A}_M$, since $C \in \mathcal{J}_t$ only for $t > t_{M-1}$,

$$
\begin{aligned}
\mathbb{E}[r_T^b(C)1(\mathcal{G}_C)] &= \mathbb{E}[\sum_{t=1}^{T}\{g^*(X_t) - g^{(\pi_t(X_t))}(X_t)\}1(X_t \in C)1(C \in \mathcal{J}_t)1(\mathcal{G}_C)] \\
&= \mathbb{E}[\sum_{t=t_{M-1}+1}^{T}\{g^*(X_t) - g^{(\pi_t(X_t))}(X_t)\}1(X_t \in C)1(\mathcal{G}_C)] \\
&\leq \mathbb{E}\Big[\sum_{t=t_{M-1}+1}^{T} c_1|p(C)|_{\mathcal{T}}^{\eta}1(0 < |g^{(1)}(X_t) - g^{(2)}(X_t)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta},\, X_t \in C)1(\mathcal{G}_C)\Big] \\
&\leq \sum_{t=t_{M-1}+1}^{T} c_1|p(C)|_{\mathcal{T}}^{\eta}\mathbb{P}(0 < |g^{(1)}(X_t) - g^{(2)}(X_t)| \leq c_1|p(A)|^{\eta},\, X_t \in C),
\end{aligned}
$$

where the first inequality is due to (S-11). In particular,

$$
\begin{aligned}
\mathbb{E}[W_M] &= \sum_{C \in \mathcal{B}_M \cap \mathcal{J}_T} \mathbb{E}[r_T^b(C)1(\mathcal{G}_C)] \\
&\leq (T - t_{M-1})c_1|p(C)|_{\mathcal{T}}^{\eta}\mathbb{P}(0 < |g^{(1)}(X) - g^{(2)}(X)| \leq c_1|p(C)|_{\mathcal{T}}^{\eta}) \\
&\leq (T - t_{M-1})D_0\{c_1 w_{M-1}^{\eta}\}^{1+\alpha}. \tag{S-15}
\end{aligned}
$$

Regret upper bound. Putting the results from (S-12), (S-14) and (S-15) together in (S-10), we get,

$$
\begin{aligned}
\mathbb{E}[R_T(\pi)1(\mathcal{E})] &\leq \sum_{1 \leq i \leq M-1} \{\mathbb{E}[U_i] + \mathbb{E}[V_i]\} + \mathbb{E}[W_M] \\
&\leq \sum_{1 \leq i \leq M-1} D_0(3\bar{c}_X + 2)\{c_1 w_{i-1}^{\eta}\}^{1+\alpha}(t_i - t_{i-1}) \\
&\qquad\qquad + D_0\{c_1 w_{M-1}^{\eta}\}^{1+\alpha}(T - t_{M-1}).
\end{aligned}
$$

By the choice of the batch sizes in (11), for $1 \leq i \leq M-1$, we have

$$
w_{i-1}^{\eta(1+\alpha)}(t_i - t_{i-1}) \asymp w_{i-1}^{\eta(1+\alpha)}w_i^{-(2\eta+1)}\log(Tw_i) \lesssim T^{\frac{1-\gamma}{1-\gamma^M}}\log(T),
$$

since $w_{i-1}^{\eta(1+\alpha)} w_i^{-(2\eta+1)} \asymp T^{-\frac{1-\gamma^{i-1}}{1-\gamma^M}\frac{\eta(1+\alpha)}{2\eta+1}+\frac{1-\gamma^i}{1-\gamma^M}} = T^{\frac{1-\gamma}{1-\gamma^M}}$ recalling the definition of $\gamma = \frac{\eta(1+\alpha)}{2\eta+1}$.

For the last term,

$$(T - t_{M-1}) w_{M-1}^{\eta(1+\alpha)} \lesssim T^{1-\frac{1-\gamma^{M-1}}{1-\gamma^M}\frac{\eta(1+\alpha)}{2\eta+1}} = T^{\frac{1-\gamma}{1-\gamma^M}}.$$

Therefore,

$$\mathbb{E}[R_T(\pi)1(\mathcal{E})] \lesssim MT^{\frac{1-\gamma}{1-\gamma^M}} \log(T).$$

On the other hand, since we have $Y_i \in [0, 1]$,

$$\mathbb{E}[R_T(\pi)1(\mathcal{E}^c)] \leq T\mathbb{P}(\mathcal{E}^c) \leq 1,$$

by Lemma S-3. Therefore, we prove the result of Theorem 1. $\qquad\square$

S2.2   Proof for Lemma S-1

Proof. There exists three cases for $C \in \mathcal{B}_i$ for $i = 1, \ldots, M - 1$.

1. C is not born at the beginning of batch i,

2. C is born at the beginning of batch i, and is not split into its children sets after the batch elimination at the end of batch i, and

3. C is born at the beginning of batch i, and is split into its children sets after the batch elimination at the end of batch i.

In case 1, C is never born, i.e., $C \notin \mathcal{L}_t$ for all $1 \leq t \leq T$, as a set $C \in \mathcal{B}_i$ can be born only at batch i by the set up of the algorithm. Moreover, since C is not born, its child $C' \in \text{child}(C)$ will not be born. Therefore $r_T^b(C) = r_T^l(C) = r_T^b(C') = 0$, and equation (S-6) is trivially true. In case 2, $C \notin \mathcal{J}_t$ for $t \leq t_{i-1}$ (before batch i) and $C \in \mathcal{L}_t$ for $t \geq t_{i-1} + 1$

(batch i and onward). Therefore,

$$
\begin{aligned}
r_T^b(C) &= \sum_{t=t_{i-1}+1}^{T} \{g^{(*)}(X_t) - g^{(\pi_t(X_t))}(X_t)\} 1(X_t \in C) 1(C \in \mathcal{J}_t) \\
&= \sum_{t=t_{i-1}+1}^{T} \{g^{(*)}(X_t) - g^{(\pi_t(X_t))}(X_t)\} 1(X_t \in C) 1(C \in \mathcal{L}_t) = r_T^l(C).
\end{aligned}
$$

Since child$(C) \notin \mathcal{J}_t$ for all t (C is not split), $r_T^b(C') = 0$ for any $C' \in$ child(C), and therefore equation (S-6) holds. In the last case,

$$
\begin{aligned}
r_T^b(C) &= \sum_{t=t_{i-1}+1}^{t_i} \{g^{(*)}(X_t) - g^{(\pi_t(X_t))}(X_t)\} 1(X_t \in C) 1(C \in \mathcal{L}_t) \\
&\quad + \sum_{t=t_i+1}^{T} \{g^{(*)}(X_t) - g^{(\pi_t(X_t))}(X_t)\} 1(X_t \in C) 1(C \in \mathcal{J}_t) \\
&= \sum_{t=t_{i-1}+1}^{t_i} \{g^{(*)}(X_t) - g^{(\pi_t(X_t))}(X_t)\} 1(X_t \in C) 1(C \in \mathcal{L}_t) \\
&\quad + \sum_{t=t_i+1}^{T} \sum_{C' \in \text{child}(C)} \{g^{(*)}(X_t) - g^{(\pi_t(X_t))}(X_t)\} 1(X_t \in C') 1(C' \in \mathcal{J}_t) \\
&= r_T^l(C) + \sum_{C' \in \text{child}(C)} r_T^b(C'),
\end{aligned}
$$

where the second equality is due to the fact that $C = \cup_{C' \in \text{child}(C)} C'$ and children sets are disjoint, and $1(C \in \mathcal{J}_t) = 1(C' \in \mathcal{J}_t) = 1$ for $t_i + 1 \le t \le T$. Therefore,

$$
r_T^b(C) = r_T^l(C) + \sum_{C' \in \text{child}(C)} r_T^b(C').
$$

The equation (S-6) is also true for $i = M$, where only the first two cases happen, and we treat $\sum_{C' \in \text{child}(C)} r_T^b(C') = \sum_{C' \in \emptyset} r_T^b(C') = 0$. $\qquad \square$

S2.3   Proof of Lemma 2

Proof. Let $\hat{\beta}^{(1)}, \dots, \hat{\beta}^{(K)}$ be the estimated index vectors. Let n_k be the number of samples used for $\hat{\beta}^{(k)}$ for $k = 1, \dots, K$. By the setup of the Algorithm 2, we have $t_{\text{init}}/(2K) \le n_k \le (2t_{\text{init}})/K$. Then, for sufficiently large t_{init}, from Assumption 5, with probability at least

$1 - KC_4(t_{\text{init}}/2K)^{-\phi}$ the following inequality holds for all $k = 1, \ldots, K$:

$$\sin \angle \hat{\beta}^{(k)}, \beta_0 \leq C_5 \frac{\text{polylog}(2t_{\text{init}}/K)}{\sqrt{t_{\text{init}}/2K}} = C_6 \frac{\text{polylog}(t_{\text{init}})}{\sqrt{t_{\text{init}}}}, \tag{S-16}$$

for another constant $C_6 = C_6(d, \phi, K)$.

Note for any u, v such that $\|u\|_2 = \|v\|_2 = 1$,

$$\|uu^\top - vv^\top\|_F^2 = 2 - 2(u^\top v)^2 = 2(\sin \angle u, v)^2, \tag{S-17}$$

since $\cos(\angle u, v) = |u^\top v|$ by the definition of the principal angle between u and v.

Then, for $\hat{\mathcal{P}} = \sum_{k=1}^K \omega_k \hat{\beta}^{(k)} (\hat{\beta}^{(k)})^\top$ with $\sum_k \omega_k = 1$,

$$\begin{aligned}
\|\hat{\mathcal{P}} - \mathcal{P}_0\|_F &= \|\sum_{k=1}^K \omega_k \{\hat{\beta}^{(k)}(\hat{\beta}^{(k)})^\top - \beta_0 \beta_0^\top\}\|_F \\
&\leq \sum_{k=1}^K \omega_k \|\hat{\beta}^{(k)}(\hat{\beta}^{(k)})^\top - \beta_0 \beta_0^\top\|_F \\
&\leq \sqrt{2} C_6 \frac{\text{polylog}(t_{\text{init}})}{\sqrt{t_{\text{init}}}}.
\end{aligned} \tag{S-18}$$

Then by a variant of Davis-Kahan inequality (Theorem 2 in Yu et al. [2015]) with $r = s = 1$ and the bound (S-18), we have,

$$\sin \angle \hat{\beta}, \beta_0 = 2\|\hat{\mathcal{P}} - \mathcal{P}_0\|_F \leq 2^{3/2} C_6 \frac{\text{polylog}(t_{\text{init}})}{\sqrt{t_{\text{init}}}}.$$

Taking $\tilde{C} = 2^{3/2} C_6$, we obtain the first inequality.

For the second inequality, note that for any u, v such that $\|u\|_2 = \|v\|_2 = 1$, if $u^\top v \geq 0$, we have

$$\|u - v\|_2^2 = 2(1 - u^\top v) \leq 2(1 - (u^\top v)^2) = 2(\sin \angle u, v)^2. \tag{S-19}$$

On the other hand, if $u^\top v \leq 0$, we have,

$$\|u + v\|_2^2 \leq \|uu^\top - vv^\top\|_F^2 = 2(\sin \angle u, v)^2, \tag{S-20}$$

which can be obtained by replacing v with $-v$ in (S-19). In particular, there exists $\hat{o} =$

$\text{sgn}(\hat{\beta}^{\top}\beta_0) \in \{-1, 1\}$ such that

$$\|\hat{\beta} \cdot \hat{o} - \beta_0\|_2 \leq \sqrt{2} \sin \angle \hat{\beta}, \beta_0 \leq 2^{1/2} \tilde{C} \frac{\text{polylog}(t_{\text{init}})}{\sqrt{t_{\text{init}}}}.$$

$\square$

S3   Lemmas

Lemma S-2. ***Multiplicative Chernoff Bound:*** *Suppose $X_1, ..., X_n$ are independent random variables taking values in $\{0, 1\}$. Let X denote their sum and let $\mu = \mathbb{E}[X]$ denote the sum's expected value. Then for any $\delta > 0$,*

$$\mathbb{P}(|X - \mu| \geq \delta\mu) \leq 2e^{-\delta^2 \mu/3}.$$

More details on multiplicative Chernoff bound and its extensions can be found in Kuszmaul and Qi [2021]. Next, we use the multiplicative Chernoff bound to provide a concentration result on the number of covariates falling in a bin contained in the tree $\mathcal{T}$.

Lemma S-3. *Suppose Assumption 3 holds. Suppose $M \leq C_1 \log T$ for some $C_1 > 0$. Suppose Assumption 4 holds, and T is sufficiently large so that $\beta_{sgn} \in \mathbb{B}_2(R_0; \beta_0)$ for $R_0 > 0$ defined in Assumption 3. For a sufficiently large T, for all $1 \leq i \leq M-1$ and $C \in \mathcal{B}_i$, we have $m_{C,i} \in [m^{\star}_{C,i}/2, 3m^{\star}_{C,i}/2]$ with probability at least $1/T$, i.e.,*

$$\mathbb{P}(\forall C \in \cup_{i=1}^{M-1} \mathcal{B}_i, m_{C,i} \in [m^{\star}_{C,i}/2, 3m^{\star}_{C,i}/2]) \geq 1 - \frac{1}{T}$$

where we define $m_{C,i} = \sum_{t=t_{i-1}+1}^{t_i} 1\{X_t \in C\}$ as the number of times X_t visits C during batch i, and $m^*_{C,i} = \mathbb{E}[m_{C,i}]$.

Proof. Let $i \in \{1, \ldots, M-1\}$ be given, and choose a set $C \in \mathcal{B}_i$. We have $C = C_A(\beta)$ with $A \in \mathcal{A}_i$. In addition, let $\Delta_{ti} = t_i - t_{i-1}$ be the size of batch i. Let $\mathcal{E}_C$ be the event that $m_{C,i} \in [m^{\star}_{C,i}/2, 3m^{\star}_{C,i}/2]$. Using the multiplicative Chernoff bound from Lemma S-2,

using $\delta = \frac{1}{2}$, we get:

$$\mathbb{P}(|\sum_{t=t_{i-1}+1}^{t_i} 1\{X_t \in C_A(\beta)\} - m^\star_{C,i}| \geq \frac{m^*_{C,i}}{2}) \leq 2\exp(-\frac{m^*_{C,i}}{12}).$$

as each $1\{X_t \in C_A(\beta)\} \in [0,1]$ a.s. Note since (X_t) are iid, $m^\star_{C,i} = \sum_{t=t_{i-1}+1}^{t_i} \mathbb{P}(X_t \in C_A(\beta)) = \Delta_{ti}\mathbb{P}_X(C_A(\beta))$. Also, note that $\mathbb{P}_X(C_A(\beta)) = \mathbb{P}(X^\top \beta \in A) = \mathbb{P}(X^\top(-\beta) \in -A)$. Defining $A_{sgn} = A$ if $\beta_{sgn} = \beta$ and $-A$ otherwise, we have $\mathbb{P}_X(C_A(\beta)) = P(X^\top \beta_{sgn} \in A_{sgn}) = \int_{u \in A_{sgn}} f_{x^\top \beta_{sgn}}(u)du$. In particular,

$$\underline{c}_X|A| \leq \mathbb{P}_X(C_A(\beta)) \leq \bar{c}_X|A| \tag{S-21}$$

by Assumption 3. Therefore, $m^*_{C,i} \geq \underline{c}_X\Delta_{ti}|A|$, and

$$P(\mathcal{E}^c_C) \leq 2\exp(-m^\star_{C,i}/12) \leq 2\exp(-\Delta_{ti}\underline{c}_X|A|/12).$$

For $1 \leq i \leq M-1$, $\Delta_{ti} = \lfloor c_B w_i^{-(2\eta+1)} \log(2Tw_i)\rfloor \asymp |A|^{-2\eta-1}\log(T|A|)$ since $|A| = w_i$ and c_B does not depend on T. Also, recall that $|A|^{-1} = w_i^{-1} = (b_0 b_1 \cdots b_{i-1})/(U_\beta - L_\beta)$ for $(b_i)_{i=1}^{M-1}$ defined in (9). In particular, for sufficiently large T, $b_i \geq 1$ for all i, and

$$\frac{\underline{c}_X}{12}\Delta_{ti}|A| \asymp |A|^{-2\eta}\log(2T|A|) \gtrsim |A|^{-2\eta} \gtrsim b_0^{2\eta} \asymp T^{(\frac{1-\gamma}{1-\gamma^M})(\frac{2\eta}{2\eta+1})}. \tag{S-22}$$

Therefore, for a sufficiently large T, $\frac{\underline{c}_X}{12}\Delta_{ti}|A| \geq 3\log(T)$, and $P(\mathcal{E}^c_i) \leq 2/T^3$.

Now we obtain a union bound over all sets in $\cup_{i=1}^{M-1}\mathcal{B}_i$. Recall the number of sets in $\mathcal{B}_i$ is $n_i = \prod_{l=0}^{i-1} b_l$, and thus the total number of sets in $\cup_{i=1}^{M-1}\mathcal{B}_i$ is $\sum_{i=1}^{M-1} n_i = \sum_{i=1}^{M-1} \prod_{l=0}^{i-1} b_l \leq M\prod_{l=0}^{M-2} b_l$. Therefore, we have

$$\mathbb{P}(\exists C \in \cup_{i=1}^{M-1}\mathcal{B}_i \text{ s.t. } m_{C,i} \notin [m^\star_{C,i}/2, 3m^\star_{C,i}/2]) \leq \sum_{C \in \cup_{i=1}^{M-1}\mathcal{B}_i} P(\mathcal{E}^c_C) \leq \frac{2M}{T^3}\prod_{l=0}^{M-2} b_l.$$

Since $\prod_{l=0}^{M-2} b_l = b_0^{1+\gamma+\cdots+\gamma^{M-3}} = b_0^{\frac{1-\gamma^{M-2}}{1-\gamma}} \asymp T^{(\frac{1-\gamma^{M-2}}{1-\gamma^M})(\frac{1}{2\eta+1})} \lesssim T$ and $M \leq C_1\log T$,

$$P(\exists C \in \cup_{i=1}^{M-1}\mathcal{B}_i \text{ such that } m_{C,i} \notin [m^\star_{C,i}/2, 3m^\star_{C,i}/2]) \lesssim \frac{2C_1\log T}{T^2} \leq \frac{1}{T},$$

S17

when T is sufficiently large. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Lemma S-4. *For $i = 1, \ldots, M - 1$, choose $C \in \mathcal{B}_i$. Suppose Assumptions 1 and 3 hold. Also assume Assumption 4, and T is sufficiently large so that $\beta_{sgn} \in \mathbb{B}_2(R_0; \beta_0)$ for $R_0 > 0$ defined in Assumption 3. For each $k \in \{1, 2\}$, define $\bar{g}_C^{(k)} = \frac{1}{\mathbb{P}_X(C)} \int_{x \in C} g^{(k)}(x) d\mathbb{P}_X(x)$. For any $x, y \in C$, $k \in \{1, 2\}$, we have*

1. *$|g^{(k)}(x) - g^{(k)}(y)| \leq L\{2R_X C_0 T^{-\xi/(2\eta+1)} + w_i\}^\eta$ and*

2. *$|\bar{g}_C^{(k)} - g^{(k)}(x)| \leq L\{2C_0 R_X T^{-\xi/(2\eta+1)} + w_i\}^\eta$.*

In particular, for a sufficiently large T, $|g^{(k)}(x) - g^{(k)}(y)| \leq L_0 w_i^\eta$ and $|\bar{g}_C^{(k)} - g^{(k)}(x)| \leq L_0 w_i^\eta$ for $L_0 := L(2^{3/2} C_0 R_X + 1)^\eta$.

Proof. We have $C = C_A(\beta)$ for an $A \in \mathcal{A}_i$. We have

$$\left| \bar{g}_C^{(k)} - g^{(k)}(x) \right| = \left| \frac{1}{\mathbb{P}_X(C)} \int_{y \in C} g^{(k)}(y) - g^{(k)}(x) d\mathbb{P}_X(y) \right|$$

by definition. Since for any $x, y \in C$, we have $x^\top \beta \in A$ and $y^\top \beta \in A$ by the set-up of C. In particular, $|x^\top \beta - y^\top \beta| = |x^\top \beta_{sgn} - y^\top \beta_{sgn}| \leq |A|$. For any $x, y \in C$ we have,

$$\begin{aligned} |g^{(k)}(x) - g^{(k)}(y)| &= |f^{(k)}(x^\top \beta_0) - f^{(k)}(y^\top \beta_0)| \\ &\leq L|x^\top \beta_0 - y^\top \beta_0|^\eta \\ &\leq L\{|(x - y)^\top \beta_{sgn}| + |(x - y)^\top (\beta_{sgn} - \beta_0)|\}^\eta \\ &\leq L\{|A| + \|x - y\|_2 \|\beta_{sgn} - \beta_0\|_2\}^\eta \\ &\leq L\{|A| + 2^{3/2} R_X C_0 T^{-\xi/(2\eta+1)}\}^\eta, \end{aligned}$$

where we use the smoothness condition of $f^{(k)}$ in Assumption 1, Assumption 3 to bound $\|y - x\|_2 \leq 2R_X$, and Assumption 4 to bound $\|\beta_{sgn} - \beta_0\|_2$. Therefore,

$$\begin{aligned} \left| \bar{g}_C^{(k)} - g^{(k)}(x) \right| &\leq \frac{1}{\mathbb{P}_X(C)} \int_{y \in C} L\{2^{3/2} C_0 R_X T^{-\xi/(2\eta+1)} + w_i\}^\eta d\mathbb{P}_X(y) \\ &\leq L\{2^{3/2} C_0 R_X T^{-\xi/(2\eta+1)} + w_i\}^\eta. \end{aligned}$$

From (10), we note that $w_i \asymp T^{-\frac{1-\gamma^i}{1-\gamma^M} \frac{1}{2\eta+1}}$. Therefore for $\xi \geq 1$, there exists $T_0 < \infty$ such

that $T^{-\xi/(2\eta+1)} \leq w_i$ for $T \geq T_0$. For such T,

$$\left|\bar{g}_C^{(k)} - g^{(k)}(x)\right| \leq \sup_{x,y \in C} \left|g^{(k)}(y) - g^{(k)}(x)\right| \leq L(2^{3/2}C_0 R_X + 1)^\eta w_i^\eta = L_0 w_i^\eta. \qquad \text{(S-23)}$$

$\square$

Lemma S-5. *Let $C \in \cup_{l=1}^{M-1} \mathcal{B}_l$ be given. We have $i \in \{1, \ldots, M-1\}$ such that $C = C_A(\beta) \in \mathcal{B}_i$ and $A \in \mathcal{A}_i$. Suppose Assumptions 1 and 3 hold. Suppose Assumption 4, and T is sufficiently large so that $\beta_{sgn} \in \mathbb{B}_2(R_0; \beta_0)$ for $R_0 > 0$ defined in Assumption 3 and $m_{C,i}^* \geq 4$. Then, we have,*

$$\mathbb{P}(\mathcal{E} \cap \mathcal{G}_C \cap \mathcal{S}_C^c) \leq \frac{3m_{C,i}^*}{2T|C|_{\mathcal{T}}},$$

where,

$$\mathcal{E} = \{\forall C \in \cup_{i=1}^{M-1} \mathcal{B}_i, m_{C,i} \in [m_{C,i}^\star/2, 3m_{C,i}^\star/2]\},$$
$$\mathcal{S}_C = \{\underline{\mathcal{I}}_C \subseteq \mathcal{I}_C' \subseteq \overline{\mathcal{I}}_C\},$$
$$\mathcal{G}_C = \cap_{C' \in \mathcal{P}(C)} \mathcal{S}_{C'},$$

and we recall the definition of $\underline{\mathcal{I}}_C$ and $\overline{\mathcal{I}}_C$ as

$$\underline{\mathcal{I}}_C = \left\{k \in \{1,2\} : \sup_{x \in C}\{f^{(*)}(x^\top \beta_0) - f^{(k)}(x^\top \beta_0)\} \leq c_0 |C|_{\mathcal{T}}^\eta\right\},$$
$$\overline{\mathcal{I}}_C = \left\{k \in \{1,2\} : \sup_{x \in C}\{f^{(*)}(x^\top \beta_0) - f^{(k)}(x^\top \beta_0)\} \leq c_1 |C|_{\mathcal{T}}^\eta\right\}$$

for $c_0 = 4L_0 + 1$ with $L_0 := L(2^{3/2}C_0 R_X + 1)^\eta$ and $c_1 = 8c_0 \gamma_X^{1/2}$.

Proof. Since $\mathcal{S}_C = \{\underline{\mathcal{I}}_C \subseteq \mathcal{I}_C' \subseteq \overline{\mathcal{I}}_C\}$, we have $\mathcal{S}_C^c = \{\underline{\mathcal{I}}_C \not\subseteq \mathcal{I}_C'\} \cup [\{\underline{\mathcal{I}}_C \subseteq \mathcal{I}_C'\} \cap \{\mathcal{I}_C' \not\subseteq \overline{\mathcal{I}}_C\}]$. Therefore,

$$\mathbb{P}(\mathcal{E} \cap \mathcal{G}_C \cap \mathcal{S}_C^c) = \mathbb{P}(\mathcal{E} \cap \mathcal{G}_C \cap \{\underline{\mathcal{I}}_C \not\subseteq \mathcal{I}_C'\}) + \mathbb{P}(\mathcal{E} \cap \mathcal{G}_C \cap \{\underline{\mathcal{I}}_C \subseteq \mathcal{I}_C'\} \cap \{\mathcal{I}_C' \not\subseteq \overline{\mathcal{I}}_C\})$$

Also, suppose for now that the following inequalities

$$2c_0|C|_{\mathcal{T}}^{\eta} \leq U(m_{C,i}, T, C) \leq \frac{2}{3}(c_1 - 2L_0)|C|_{\mathcal{T}}^{\eta} \tag{S-24}$$

hold on $\mathcal{E}$, which we later will show. Here, we recall that $|C|_{\mathcal{T}} = |A|$ for $C = C_A(\beta)$.

For the first term, since $\underline{\mathcal{I}}_C \nsubseteq \mathcal{I}'_C$, there exists an arm $k_1 \in \underline{\mathcal{I}}_C$ such that $k_1 \notin \mathcal{I}'_C$, i.e., k_1 was eliminated at the end of batch i within the bin C. By the arm elimination mechanism, $\exists k_2 \in \mathcal{I}_{p(C)}$ such that,

$$\bar{Y}_{C,i}^{(k_2)} - \bar{Y}_{C,i}^{(k_1)} > U(m_{C,i}, T, C). \tag{S-25}$$

We argue that this implies that there exists $k \in \{1, 2\}$ such that $|\bar{Y}_{C,i}^{(k)} - \bar{g}_C^{(k)}| > \frac{1}{4}U(m_{C,i}, T, C)$. We have,

$$\bar{g}_C^{(k_2)} - \bar{g}_C^{(k_1)} = \frac{1}{\mathbb{P}_X(C)}\int_{x \in C}\{g^{(k_2)}(x) - g^{(k_1)}(x)\}d\mathbb{P}_X(x) \leq \frac{1}{\mathbb{P}_X(C)}\int_{x \in C}\{g^{(*)}(x) - g^{(k_1)}(x)\}d\mathbb{P}_X(x)$$

and since $k_1 \in \underline{\mathcal{I}}_C$, $\sup_{x \in C}\{g^{(*)}(x) - g^{(k_1)}(x)\} \leq c_0|A|^{\eta}$, and thus

$$\bar{g}_C^{(k_2)} - \bar{g}_C^{(k_1)} \leq c_0|A|^{\eta}.$$

Then, if both $k \in \{k_1, k_2\}$ satisfy $|\bar{Y}_{C,i}^{(k)} - \bar{g}_C^{(k)}| \leq \frac{1}{4}U(m_{C,i}, T, C)$, then

$$
\begin{aligned}
\bar{Y}_{C,i}^{(k_2)} - \bar{Y}_{C,i}^{(k_1)} &= \bar{Y}_{C,i}^{(k_2)} - \bar{g}_C^{(k_2)} + \bar{g}_C^{(k_2)} - \bar{g}_C^{(k_1)} + \bar{g}_C^{(k_1)} - \bar{Y}_{C,i}^{(k_1)} \\
&\leq |\bar{Y}_{C,i}^{(k_2)} - \bar{g}_C^{(k_2)}| + \{\bar{g}_C^{(k_2)} - \bar{g}_C^{(k_1)}\} + |\bar{Y}_{C,i}^{(k_1)} - \bar{g}_C^{(k_1)}| \\
&\leq \frac{1}{2}U(m_{C,i}, T, C) + c_0|A|^{\eta} \\
&\leq U(m_{C,i}, T, C),
\end{aligned}
$$

which is a contradiction, and therefore on $\mathcal{E}$, there exists $k \in \{1, 2\}$ such that $|\bar{Y}_{C,i}^{(k)} - \bar{g}_C^{(k)}| > \frac{1}{4}U(m_{C,i}, T, C)$. In particular, we can bound the first term as follows:

$$\mathbb{P}(\mathcal{E} \cap \mathcal{G}_C \cap \{\underline{\mathcal{I}}_C \subseteq \mathcal{I}'_C\}^c) \leq \mathbb{P}\left(\mathcal{E} \cap \left\{\exists k \in \{1, 2\} \text{ s.t.} |\bar{Y}_{C,i}^{(k)} - \bar{g}_C^{(k)}| > \frac{1}{4}U(m_{C,i}, T, C)\right\}\right).$$

For the second term where $\{\underline{\mathcal{I}}_C \subseteq \mathcal{I}'_C\} \cap \{\mathcal{I}'_C \not\subseteq \overline{\mathcal{I}}_C\}$, there exists $k_1 \in \mathcal{I}'_C$ such that $k_1 \notin \overline{\mathcal{I}}_C$. By the definition of $\overline{\mathcal{I}}_C$, there exists $x_0 \in C$ such that

$$g^{(k_2)}(x_0) - g^{(k_1)}(x_0) > c_1 |A|^\eta \tag{S-26}$$

for $k_2 \neq k_1$. Then, for any $x \in C$,

$$g^{(k_2)}(x) - g^{(k_1)}(x) \geq g^{(k_2)}(x_0) - g^{(k_1)}(x_0) - \sum_{k \in \{1,2\}} |g^{(k)}(x) - g^{(k)}(x_0)|$$

$$\geq c_1 |A|^\eta - 2L_0 |A|^\eta = (c_1 - 2L_0)|A|^\eta > 0 \tag{S-27}$$

where the last inequality is due to the fact that for sufficiently large T, $|g^{(k)}(x) - g^{(k)}(x_0)| \leq L_0 |A|^\eta$ by (S-23), and

$$c_1 - 2L_0 \geq 8c_0 \gamma_X^{1/2} - c_0 = c_0(8\gamma_X^{1/2} - 1) \geq 7c_0 \gamma_X^{1/2} > 0 \tag{S-28}$$

since $c_1 = 8c_0 \gamma_X^{1/2}$, $c_0 = 4L_0 + 1 \geq 2L_0$, and $\gamma_X \geq 1$.

Note the bound (S-27) implies that k_2 is universally better than k_1 on C. In particular, $k_2 \in \underline{\mathcal{I}}_C \subseteq \mathcal{I}'_C$ as well. Since both $k_1, k_2 \in \mathcal{I}'_C$,

$$|\bar{Y}_{C,i}^{(k_1)} - \bar{Y}_{C,i}^{(k_2)}| \leq U(m_{C,i}, T, C).$$

We argue that on $\mathcal{E}$, when T is sufficiently large, this implies that there exists $k \in \{1, 2\}$ such that $|\bar{Y}_{C,i}^{(k)} - \bar{g}_C^{(k)}| > \frac{1}{4} U(m_{C,i}, T, C)$. We have

$$\bar{g}_C^{(k_2)} \geq g^{(k_2)}(x_0) - |\bar{g}_C^{(k)} - g^{(k)}(x_0)|$$

$$\geq g^{(k_2)}(x_0) - L_0 |A|^\eta$$

$$> \{g^{(k_1)}(x_0) + c_1 |A|^\eta\} - L_0 |A|^\eta,$$

where the second inequality is due to Lemma S-4, and the third inequality is due to the

choice of x_0 in (S-26). Applying Lemma S-4 again,

$$\begin{aligned}
\bar{g}_C^{(k_2)} &> g^{(k_1)}(x_0) + c_1|A|^\eta - L_0|A|^\eta \\
&> \{\bar{g}_C^{(k_1)} - |\bar{g}_C^{(k_1)} - g^{(k_1)}(x_0)|\} + c_1|A|^\eta - L_0|A|^\eta \\
&> \bar{g}_C^{(k_1)} + (c_1 - 2L_0)|A|^\eta \\
&> \bar{g}_C^{(k_1)} + \frac{3}{2}U(m_{C,i}, T, C),
\end{aligned}$$

where for the last inequality we use (S-24). On the other hand,

$$\begin{aligned}
|\bar{g}_C^{(k_2)} - \bar{g}_C^{(k_1)}| &\le |\bar{g}_C^{(k_2)} - \bar{Y}_{C,i}^{(k_2)}| + |\bar{Y}_{C,i}^{(k_2)} - \bar{Y}_{C,i}^{(k_1)}| + |\bar{g}_C^{(k_2)} - \bar{Y}_{C,i}^{(k_1)}| \\
&\le |\bar{g}_C^{(k_2)} - \bar{Y}_{C,i}^{(k_2)}| + U(m_{C,i}, T, C) + |\bar{g}_C^{(k_2)} - \bar{Y}_{C,i}^{(k_1)}|.
\end{aligned}$$

Therefore if both $k \in \{k_1, k_2\}$ satisfy $|\bar{Y}_{C,i}^{(k)} - \bar{g}_C^{(k)}| \le \frac{1}{4}U(m_{C,i}, T, C)$, then $|\bar{g}_C^{(k_2)} - \bar{g}_C^{(k_1)}| \le \frac{3}{2}U(m_{C,i}, T, C)$, which is a contradiction. Therefore,

$$\mathbb{P}(\mathcal{E} \cap \mathcal{G}_C \cap \{\underline{\mathcal{I}}_C \subseteq \mathcal{I}_C'\} \cap \{\mathcal{I}_C' \not\subseteq \overline{\mathcal{I}}_C\}) \le \mathbb{P}(\mathcal{E} \cap \{\exists k \in \{1, 2\} \text{ s.t.} |\bar{Y}_{C,i}^{(k)} - \bar{g}_C^{(k)}| > \frac{1}{4}U(m_{C,i}, T, C)\}).$$

Combining two inequalities and by Lemma S-7, we have

$$\begin{aligned}
\mathbb{P}(\mathcal{E} \cap \mathcal{G}_C \cap \mathcal{S}_C^c) &\le 2\mathbb{P}(\mathcal{E} \cap \{\exists k \in \{1, 2\} \text{ s.t. } |\bar{Y}_{C,i}^{(k)} - \bar{g}_C^{(k)}| > \frac{1}{4}U(m_{C,i}, T, C)\}) \\
&\le \frac{3m_{C,i}^*}{2T|A|}.
\end{aligned}$$

It remains to show (S-24) on $\mathcal{E}$. Recall $U(m, T, C) = 4\sqrt{2\log(2T|A|)/m}$. First we show that

$$c_0|A|^\eta \le \frac{1}{2}U(\frac{3}{2}m_{C,i}^*, T, C) \quad \text{and} \quad \frac{3}{2}U(\frac{1}{2}m_{C,i}^*, T, C) \le (c_1 - 2L_0)|A|^\eta. \tag{S-29}$$

Recall for $1 \le i \le M - 1$, $m_{C,i}^* = (t_i - t_{i-1})\mathbb{P}_X(C)$, and we have $\underline{c}_X|A| \le \mathbb{P}_X(C) \le \overline{c}_X|A|$ (ref. Equation (S-21)) under the stated assumptions. Moreover, we have $t_i - t_{i-1} = \lfloor c_B|A|^{-(2\eta+1)}\log(2T|A|)\rfloor$ and $c_B = 4/(c_0^2\overline{c}_X) = 4(4L_0 + 1)^{-2}(\overline{c}_X)^{-1}$ in (11). Therefore, we

have

$$\frac{1}{2}U(\frac{3}{2}m_{C,i}^*, T, C) \geq 2\sqrt{\frac{2\log(2T|A|)}{(3/2)c_B|A|^{-2\eta-1}\log(2T|A|)\mathbb{P}_X(C)}}$$

$$\geq 2\sqrt{\frac{2\log(2T|A|)c_0^2\bar{c}_X}{(3/2)\cdot 4\cdot|A|^{-2\eta-1}\log(2T|A|)(\bar{c}_X|A|)}}$$

$$\geq \frac{2}{\sqrt{3}}|A|^{\eta}\sqrt{\frac{c_0^2\bar{c}_X|A|}{\bar{c}_X|A|}} = c_0|A|^{\eta}.$$

On the other hand,

$$\frac{3}{2}U(\frac{1}{2}m_{C,i}^*, T, C) = 6\sqrt{\frac{2\log(2T|A|)}{(1/2)\lfloor c_B|A|^{-2\eta-1}\log(2T|A|)\rfloor\mathbb{P}_X(C)}}.$$

To upper-bound RHS,

$$\lfloor c_B|A|^{-2\eta-1}\log(2T|A|)\rfloor \geq c_B|A|^{-2\eta-1}\log(2T|A|) - 0.5 \geq (1-\delta)c_B|A|^{-2\eta-1}\log(2T|A|)$$

for sufficiently large T, for any given $\delta > 0$, since $|A|^{-2\eta-1}\log(2T|A|)$ grows with T. In particular, taking $\delta = 3/4$ and using $\mathbb{P}_X(C) \geq \underline{c}_X|A|$,

$$\frac{3}{2}U(\frac{1}{2}m_{C,i}^*, T, C) \leq 6\sqrt{\frac{2\log(2T|A|)c_0^2\bar{c}_X}{(1/2)(3/4)4|A|^{-2\eta-1}\log(2T|A|)\underline{c}_X|A|}}$$

$$\leq (12/\sqrt{3})|A|^{\eta}\sqrt{\frac{c_0^2\bar{c}_X|A|}{\underline{c}_X|A|}}$$

$$\leq 7c_0\gamma_X^{1/2}|A|^{\eta}$$

$$\leq (c_1 - 2L_0)|A|^{\eta},$$

where for the last inequality we use (S-28).

Finally, on $\mathcal{E}$, we have that $\frac{1}{2}m_{C,i}^* \leq m_{C,i} \leq \frac{3}{2}m_{C,i}^*$, therefore

$$U(1.5m_{C,i}^*, T, C) \leq U(m_{C,i}, T, C) \leq U(0.5m_{C,i}^*, T, C) \tag{S-30}$$

By combining (S-29) and (S-30), we obtain (S-24).

$\square$

Lemma S-6. *Let $i \in \{1, \ldots, M\}$ be given, and fix $C \in \mathcal{B}_i$. Let $\tau_{C,i}(s)$ be the sth time at which the sequence X_t is in C during $[t_i, t_{i+1})$. Fix $k \in \{1, 2\}$. Assume $|Y_t^{(k)}| \leq 1$ almost surely for any t, k. Consider $\{Y_{\tau_{C,i}(s)}^{(k)}; s = 1, \ldots, N\}$ for some $N < \infty$. Then $\{Y_{\tau_{C,i}(s)}^{(k)}; s = 1, \ldots, N\}$ are independent random variables with expectation $\bar{g}_C^{(k)}$, where*

$$\bar{g}_C^{(k)} := \frac{1}{\mathbb{P}(X \in C)} \int_{x \in C} g^{(k)}(x) d\mathbb{P}_X(x) = \frac{1}{\mathbb{P}(X \in C)} \int_{x \in C} f^{(k)}(x^\top \beta_0) d\mathbb{P}_X(x).$$

Proof. Recall that $\tau_{C,i}(s) = \inf\{n \geq \tau_{C,i}(s-1) + 1; X_n \in C\}$ represents the time of the sth visit to the set C from t_{i-1}, for $s = 1, 2, \ldots$ and $\tau_{C,i}(0) = t_{i-1}$. Without loss of generality, assume $i = 1$; otherwise we can redefine the sequence $X_{t_{i-1}+1}, X_{t_{i-1}+2}, \ldots$ as $X_1, X_2, \ldots$. Also, let $\tau_C(s) = \tau_{C,i}(s)$ for notational simplicity.

We note that for any s, $\tau_C(s)$ is a stopping time with respect to filtration $\mathcal{F}_t^X = \sigma(X_1, \ldots, X_t)$, as for any $t \in \mathbb{N}$, $\{\tau_C(s) > t\} = \{\sum_{n=1}^t 1\{X_n \in C\} < s\}$ and therefore $\{\tau_C(s) > t\}$ is $\mathcal{F}_t^X$-measurable.

First, we compute $\mathbb{E}[Y_{\tau_C(s)}^{(k)}]$. First note that $1 = \sum_{t=s}^\infty 1\{\tau_C(s) = t\}$ almost surely and

$$\{\tau_C(s) = t\}$$

$$= \bigcup_{\substack{(i_1, \ldots, i_{s-1}) \subseteq \{1, \ldots, t-1\} \\ (j_1, \ldots, j_{t-s}) \subseteq \{1, \ldots, t-1\} \setminus (i_1, \ldots, i_{s-1})}} \{X_{i_1} \in C, \ldots, X_{i_{s-1}} \in C, X_{j_1} \in C^c, \ldots, X_{j_{n-s}} \in C^c\} \bigcap \{X_t \in C\}$$

as $\{\tau_C(s) = t\}$ is the event where X_n visits C for $s - 1$ times during $n = 1, \ldots, t-1$ and $X_t \in C$. For future reference, we define for $a < b$, and $s \in \{0, \ldots, b - a\}$,

$$\mathcal{E}_C(a, b, s) = \bigcup_{\substack{(i_1, \ldots, i_s) \subseteq \{a+1, \ldots, b\} \\ (j_1, \ldots, j_{b-a-s}) \subseteq \{a+1, \ldots, b\} \setminus (i_1, \ldots, i_s)}} \{X_{i_1} \in C, \ldots, X_{i_{s-1}} \in C, X_{j_1} \in C^c, \ldots, X_{j_{b-a-s}} \in C^c\}$$

to be the event that during $n = a + 1, \ldots, b$, $X_n \in C$ for s times. With this notation,

$$\{\tau_C(s) = t\} = \mathcal{E}_C(0, t-1, s-1) \cap \{X_t \in C\}. \tag{S-31}$$

Since $(X_t)_{t \geq 1}$ are independent and identically distributed, we have,

$$\mathbb{P}(\mathcal{E}_C(a,b,s)) = \binom{b-a}{s} \mathbb{P}(X_1 \in C^c)^{(b-a)-s} \mathbb{P}(X_1 \in C)^s$$

Therefore, we have,

$$\mathbb{E}[Y^{(k)}_{\tau_C(s)}] = \mathbb{E}[\sum_{t=s}^{\infty} Y^{(k)}_t 1\{\tau_C(s) = t\}]$$

$$= \sum_{t=s}^{\infty} \mathbb{E}[Y^{(k)}_t 1\{\tau_C(s) = t\}]$$

$$= \sum_{t=s}^{\infty} \mathbb{E}[Y^{(k)}_t 1_{\mathcal{E}_C(0,t-1,s-1)} 1\{X_t \in C\}]$$

$$= \sum_{t=s}^{\infty} \binom{t-1}{s-1} \mathbb{P}(X_1 \in C^c)^{t-s} \mathbb{P}(X_1 \in C)^{s-1} \mathbb{E}[Y^{(k)}_t 1\{X_t \in C\}]$$

where for the second line we use the Fubini's theorem and the fact that $|Y^{(k)}_t|$ is bounded almost surely, and for the third line we use the independence between $(X_1, \ldots, X_{t-1})$ and (X_t, Y_t). Since $\sum_{t=s}^{\infty} \binom{t-1}{s-1} \mathbb{P}(X_1 \in C^c)^{t-s} \mathbb{P}(X_1 \in C)^{s-1} = \mathbb{P}(X_1 \in C)^{-1}$, we have

$$\mathbb{E}[Y^{(k)}_{\tau_C(s)}] = \frac{\mathbb{E}[Y^{(k)}_1 1\{X_1 \in C\}]}{\mathbb{P}(X_1 \in C)} = \frac{1}{\mathbb{P}_X(C)} \int_{x \in C} g^{(k)}(x) d\mathbb{P}_X(x) = \bar{g}^{(k)}_C$$

where we note that $\mathbb{E}[Y^{(k)}_1 1\{X_1 \in C\}] = \mathbb{E}_{X_1}[\mathbb{E}_{\epsilon|X_1}[Y^{(k)}_1 | X_1] 1\{X_1 \in C\}] = \mathbb{E}_{X_1}[g^{(k)}(X_1) 1\{X_1 \in C\}]$.

Now we show the independence of $\{Y^{(k)}_{\tau_C(s)}; s = 1, \ldots, N\}$. Fix $m \leq N$. Let $(i_1, \ldots, i_m) \subseteq \{1, \ldots, N\}$ be given such that $i_1 < i_2 < \cdots < i_m$, as well as $B_1, \ldots, B_m \in \mathscr{B}_{\mathbb{R}}$. It is sufficient to show $\mathbb{P}(Y^{(k)}_{\tau_C(i_1)} \in B_1, \ldots, Y^{(k)}_{\tau_C(i_m)} \in B_m) = \prod_{j=1}^{m} \mathbb{P}(Y^{(k)}_{\tau_C(i_j)} \in B_j)$.

$$\mathbb{P}(Y^{(k)}_{\tau_C(i_1)} \in B_1, \ldots, Y^{(k)}_{\tau_C(i_m)} \in B_m)$$

$$= \sum_{n_1, n_2, \ldots, n_m} \mathbb{P}(Y^{(k)}_{n_1} \in B_1, \ldots, Y^{(k)}_{n_m} \in B_m, \tau_C(i_1) = n_1, \ldots, \tau_C(i_m) = n_m)$$

Recall $\{\tau_C(i_1) = n_1, \ldots, \tau_C(i_m) = n_m\}$ is the event that the time point for the i_1th visit $= n_1$, time point for the i_2th visit $= n_2, \ldots$, and the time point for the i_mth visit $= n_m$. Note that there are some restrictions in the possible values of $(n_1, \ldots, n_m)$. For example,

the earliest time X_t can visit C for i_1 times is i_1, when $X_t \in C$ for $1 \leq t \leq i_1$, so $n_1 \geq i_1$. When $\tau_C(i_1) = n_1$, the earliest time that X_t can visit C for i_2 times is $n_1 + (i_2 - i_1)$, so n_2 has to be at least $n_1 + (i_2 - i_1)$. With this consideration, we have,

$$\mathbb{P}(Y^{(k)}_{\tau_C(i_1)} \in B_1, \ldots, Y^{(k)}_{\tau_C(i_m)} \in B_m)$$

$$= \sum_{n_1, n_2, \ldots, n_m} \mathbb{P}(Y^{(k)}_{n_1} \in B_1, \ldots, Y^{(k)}_{n_m} \in B_m, \tau_C(i_1) = n_1, \ldots, \tau_C(i_m) = n_m)$$

$$= \sum_{n_1=i_1}^{\infty} \sum_{n_2=n_1+(i_2-i_1)}^{\infty} \cdots \sum_{n_m=n_{m-1}+(i_m-i_{m-1})}^{\infty} \mathbb{E}(1\{\mathcal{E}_C(0, n_1 - 1, i_1 - 1) \cap \{X_{n_1} \in C, Y^{(k)}_{n_1} \in B_1\} \quad \cdots$$

$$\cap \, \mathcal{E}_C(n_{m-1}, n_m - 1, i_m - i_{m-1} - 1) \cap \{X_{n_m} \in C, Y^{(k)}_{n_m} \in B_m\}\})$$

$$= \sum_{n_1=i_1}^{\infty} \sum_{n_2=n_1+(i_2-i_1)}^{\infty} \cdots \sum_{n_m=n_{m-1}+(i_m-i_{m-1})}^{\infty} \prod_{j=1}^{m} \mathbb{P}(\mathcal{E}_C(n_{j-1}, n_j - 1, i_j - i_{j-1} - 1))$$

$$\times \, \mathbb{P}(X_{n_j} \in C, Y^{(k)}_{n_j} \in B_j)$$

where we define $n_0 = 0, i_0 = 0$, and use independence for the last equation. Since

$$\mathbb{P}(\mathcal{E}_C(n_{j-1}, n_j - 1, i_j - i_{j-1} - 1)) = \binom{n_j - n_{j-1} - 1}{i_j - i_{j-1} - 1}(1 - p)^{(n_j - n_{j-1}) - (i_j - i_{j-1})} p^{i_j - i_{j-1} - 1},$$

for $p = \mathbb{P}(X \in C)$, we have,

$$\sum_{n_1=1}^{\infty} \sum_{n_2=n_1+(i_2-i_1)}^{\infty} \cdots \sum_{n_m=n_{m-1}+(i_m-i_{m-1})}^{\infty} \prod_{j=1}^{m} \mathbb{P}(\mathcal{E}_C(n_{j-1}, n_j - 1, i_j - i_{j-1} - 1))$$

$$\times \, \mathbb{P}(X_{n_j} \in C, Y^{(k)}_{n_j} \in B_j)$$

$$= \prod_{j=1}^{m} \left\{ \sum_{n_j=n_{j-1}+(i_j-i_{j-1})}^{\infty} \mathbb{P}(\mathcal{E}_C(n_{j-1}, n_j - 1, i_j - i_{j-1} - 1)) \mathbb{P}(X_1 \in C, Y^{(k)}_1 \in B_j) \right\}$$

$$= \prod_{j=1}^{m} \frac{\mathbb{P}(X_1 \in C, Y^{(k)}_1 \in B_j)}{\mathbb{P}(X_1 \in C)} \tag{S-32}$$

where for the last equality we use the fact that for any $j \in \{1, \ldots, m\}$,

$$\sum_{n_j=n_{j-1}+(i_j-i_{j-1})}^{\infty} \mathbb{P}(\mathcal{E}_C(n_{j-1}, n_j - 1, i_j - i_{j-1} - 1))$$

$$= \sum_{n_j=n_{j-1}+(i_j-i_{j-1})}^{\infty} \binom{n_j - n_{j-1} - 1}{i_j - i_{j-1} - 1} (1-p)^{(n_j-n_{j-1})-(i_j-i_{j-1})} p^{(i_j-i_{j-1})-1}$$

$$= \sum_{k=i_j-i_{j-1}}^{\infty} \binom{k-1}{(i_j - i_{j-1}) - 1} (1-p)^{k-(i_j-i_{j-1})} p^{(i_j-i_{j-1})-1} = \frac{1}{p}. \tag{S-33}$$

Here for the last equality, we use the following identity $\sum_{k=r}^{\infty} \binom{k-1}{r-1} p^k (1-p)^{n-r} = 1$ with $r = i_j - i_{j-1}$.

On the other hand, for any $j \in \{1, \ldots, m\}$,

$$\mathbb{P}(Y_{\tau_C(i_j)}^{(k)} \in B_j) = \sum_{n=i_j}^{\infty} \mathbb{E}[1\{Y_n^{(k)} \in B_j, \tau_C(i_j) = n\}]$$

$$= \sum_{n=i_j}^{\infty} \mathbb{E}[1\{Y_n^{(k)} \in B_j, X_n \in C\}1\{\mathcal{E}_C(0, n-1, i_j - 1)\}]$$

$$= \mathbb{P}(Y_1^{(k)} \in B_j, X_1 \in C) \sum_{n=i_j}^{\infty} \mathbb{P}(\mathcal{E}_C(0, n-1, i_j-1))$$

$$= \frac{\mathbb{P}(Y_1^{(k)} \in B_j, X_1 \in C)}{\mathbb{P}(X_1 \in C)} \tag{S-34}$$

where we use (S-33) with $j = 1$ for the last equality.

Therefore $\mathbb{P}(Y_{\tau_C(i_1)}^{(k)} \in B_1, \ldots, Y_{\tau_C(i_m)}^{(k)} \in B_m) = \prod_{j=1}^{m} \mathbb{P}(Y_{\tau_C(i_j)}^{(k)} \in B_j)$ by (S-32) and (S-34) and the proof is complete. □

Lemma S-7. *Fix $i \in \{1, \ldots, M-1\}$ and $C \in \mathcal{B}_i$. Suppose T is sufficiently large that $m_{C,i}^* \geq 4$. Assume $|Y_t^{(k)}| \leq 1$ almost surely for any t, k. Define $U(m, T, C) = 4\sqrt{\frac{2\log(2T|C|_{\mathcal{T}})}{m}}$. We have*

$$\mathbb{P}\left(\mathcal{E} \cap \left\{\exists k \in \{1, 2\}; |\bar{Y}_{C,i}^{(k)} - \bar{g}_{C_A}^{(k)}| \geq \frac{1}{4}U(m_{C,i}, T, C)\right\}\right) \leq \frac{3m_{C,i}^*}{2T|C|_{\mathcal{T}}}.$$

where $\mathcal{E} = \{\forall C \in \cup_{i=1}^{M-1}\mathcal{B}_i, m_{C,i} \in [m_{C,i}^*/2, 3m_{C,i}^*/2]\}$ and for $\bar{Y}_{C,i}^{(k)}$ defined in (5).

Proof. We have

$$
\mathbb{P}\left(\mathcal{E} \cap \left\{\exists k \in \{1,2\} \text{ s.t. } |\bar{Y}_{C,i}^{(k)} - \bar{g}_{C_A}^{(k)}| \geq \frac{1}{4}U(m_{C,i}, T, C)\right\}\right)
$$

$$
\leq \mathbb{P}\left(2 \leq m_{C,i} \leq \frac{3}{2}m_{C,i}^*, \exists k \in \{1,2\} \text{ s.t. } |\bar{Y}_{C,i}^{(k)} - \bar{g}_{C_A}^{(k)}| \geq \frac{1}{4}U(m_{C_A,i}, T, C)\right)
$$

$$
\leq \sum_{k=1}^{2} \mathbb{P}\left(2 \leq m_{C,i} \leq \frac{3}{2}m_{C,i}^*, |\bar{Y}_{C,i}^{(k)} - \bar{g}_{C_A}^{(k)}| \geq \frac{1}{4}U(m_{C,i}, T, C)\right)
$$

$$
\leq \sum_{k=1}^{2} \sum_{n=2}^{\lfloor 1.5m_{C,i}^* \rfloor} \mathbb{P}\left(m_{c,i} = n, |\bar{Y}_{C,i}^{(k)} - \bar{g}_{C_A}^{(k)}| \geq \frac{1}{4}U(m_{C,i}, T, C)\right).
$$

For any $n > 0$, $\{Y_{\tau_{C,i}(s)}^{(k)}; 1 \leq s \leq n\}$ consists of bounded independent random variables with mean $\bar{g}_C^{(k)}$ by Lemma S-6. Define

$$
\tilde{Y}_n^{(k)} = \frac{1}{|\{1 \leq s \leq n; s \bmod 2 \equiv k\}|} \sum_{\substack{1 \leq s \leq n \\ s \bmod 2 \equiv k}} Y_{\tau_{C,i}(s)}^{(k)},
$$

which represents the average of the $Y_{\tau_{C,i}(s)}^{(k)}$ values over the indices s satisfying $s \bmod 2 \equiv k$, corresponding to either the odd ($k = 1$) or even ($k = 2$) or terms of the sequence $\{Y_{\tau_{C,i}(s)}^{(k)}; 1 \leq s \leq n\}$ of length n. Also note that when n is even, $\tilde{Y}_n^{(k)}$ is the average of $n/2$ terms, and when n is odd, $\tilde{Y}_n^{(k)}$ is the average of $(n+1)/2$ terms for $k = 1$ and $(n-1)/2$ terms for $k = 2$. In all cases, $\tilde{Y}_n^{(k)}$ is the average of at least $(n-1)/2$ terms.

On $\{m_{C,i} = n\}$, we have $\bar{Y}_{C,i}^{(k)} = \tilde{Y}_n^{(k)}$. For $n \geq 2$ (note this guarantees that $\tilde{Y}_n^{(k)}$ is the average of at least 1 term), by Hoeffding's inequality,

$$
\mathbb{P}\left(|\tilde{Y}_n^{(k)} - \bar{g}_C^{(k)}| \geq \sqrt{\frac{2\log(2T|C|_\mathcal{T})}{n}}\right) \leq \exp\left(-2 \cdot \frac{\log(2T|C|_\mathcal{T})}{(1/2)n} \cdot \frac{n}{4}\right) = \frac{1}{2T|C|_\mathcal{T}}
$$

where we use the fact that $n/2 - 1/2 \geq n/4$ for any $n \geq 2$. Then, by using the union bound,

$$
\mathbb{P}\left(\mathcal{E} \cap \left\{\exists k \in \{1,2\} \text{ s.t. } |\bar{Y}_{C,i}^{(k)} - \bar{g}_C^{(k)}| \geq \frac{1}{4}U(m_{C,i}, T, C)\right\}\right) \leq 2 \cdot \lfloor 1.5m_{C,i}^* \rfloor \frac{1}{2T|C|_\mathcal{T}} \leq \frac{3m_{C,i}^*}{2T|C|_\mathcal{T}}.
$$

$\square$

S4 Single index vector estimation from the initial phase

In this section, we present an example of constructing the initial vector $\hat{\beta}$ which satisfies Assumption 5. We propose using the Sliced Average Derivative Estimator (SADE) introduced by Babichev and Bach [2018], which combines the Average Derivative Estimator and Sliced Inverse Regression . This approach offers provable improvements over non-sliced versions and provides non-asymptotic bounds for estimating a matrix whose column space lies within the effective dimension reduction (e.d.r) space. Using this bound and the Davis-Kahan inequality, we will derive a non-asymptotic bound for the initial vector that satisfies Assumption 5.

SADE algorithm   We briefly describe the SADE algorithm and the non-asymptotic bound for the matrix whose column space belongs to the e.d.r of the model by Babichev and Bach [2018]. Consider for now a dataset with iid observations $(X_i, Y_i)_{i=1}^n$. Babichev and Bach [2018] makes the following assumptions on the model and the distribution of X:

1. (A1) For all $x \in \mathbb{R}^d$, we have $f(x) = g\left(w^\top x\right)$ for a certain matrix $w \in \mathbb{R}^{d \times k}$ and a function $g : \mathbb{R}^k \to \mathbb{R}$. Moreover, $Y = f(X) + \varepsilon$ with ε independent of X with zero mean and finite variance.

2. (A2) The distribution of X has a strictly positive density $p(x)$ which is differentiable with respect to the Lebesgue measure, and such that $p(x) \to 0$ when $\|x\| \to \infty$.

Note that when $k = 1$ in (A1), the model corresponds to the single-index model.

Let $\mathcal{S}_1(x)$ be the negative derivative of the log density of $\mathbb{P}_X$, i.e., $\mathcal{S}_1(x) = -\nabla \log p(x) = \frac{-1}{p(x)} \nabla p(x)$ where $p(x)$ is the density function of $\mathbb{P}_X$ with respect to Lebesgue measure, which is assumed to be known. For example, if X is normally distributed with mean vector μ and covariance matrix Σ, then $\mathcal{S}_1(x) = \Sigma^{-1}(x - \mu)$.

From Lemma 2 in Babichev and Bach [2018], under (A1)–(A2), $\mathbb{E}(\mathcal{S}_1(X)|Y = y)$ belongs to the e.d.r space $\mathrm{span}(w_1, \ldots, w_k)$ for almost every (a.e.) y. Then $\mathcal{V}_{1,\mathrm{cov}} = \mathbb{E}[\mathbb{E}(\mathcal{S}_1(X)|Y)\mathbb{E}(\mathcal{S}_1(X)|Y)^\top] = \mathrm{Cov}[\mathbb{E}(\mathcal{S}_1(x)|Y)]$ will be at most a rank-k matrix whose eigenvectors corresponding to non-zero eigenvalues belong to $\mathrm{span}(w_1, \ldots, w_k)$. The process to estimate $\mathcal{V}_{1,\mathrm{cov}}$ given a data $(x_i, y_i)_{i=1}^n$ is summarized in Algorithm 3.

Algorithm 3 SADE Algorithm to estimate β_0 for i.i.d. dataset

---

1: **Input:** Data $(x_i, y_i)_{i=1}^n$, score function $\mathcal{S}_1$, number of slices H
2: **Output:** β = the scaled eigenvector corresponding to the largest eigenvalue of $\hat{\mathcal{V}}_{1,\mathrm{cov}}$
3: Slice $[0,1]$ into H slices $I_1, \ldots, I_H$
4: Let $\hat{p}_h$ be the empirical proportion of y_i that fall in the slice I_h:

$$\hat{p}_h = \frac{\sum_{i=1}^n \mathbb{1}\{y_i \in I_h\}}{n}$$

5: Estimate $(\mathcal{S}_1)_h = \mathbb{E}[\mathcal{S}_1(x) \mid y \in I_h]$ by:

$$(\hat{\mathcal{S}}_1)_h = \frac{1}{\sum_{i=1}^n \mathbb{1}\{y_i \in I_h\}} \sum_{i=1}^n \mathbb{1}\{y_i \in I_h\} \mathcal{S}_1(x_i)$$

6: Estimate $\mathrm{Cov}(\mathcal{S}_1(x) \mid y \in I_h)$ by:

$$(\hat{\mathcal{S}}_1)_{\mathrm{cov},h} = \frac{1}{n\hat{p}_h - 1} \sum_{i=1}^n \mathbb{1}\{y_i \in I_h\}(\mathcal{S}_1(x_i) - (\hat{\mathcal{S}}_1)_h)(\mathcal{S}_1(x_i) - (\hat{\mathcal{S}}_1)_h)^\top$$

7: Compute:

$$\hat{\mathcal{V}}_{1,\mathrm{cov}} = \frac{1}{n} \sum_{i=1}^n \mathcal{S}_1(x_i)\mathcal{S}_1(x_i)^\top - \sum_{h=1}^H \hat{p}_h \cdot (\hat{\mathcal{S}}_1)_{\mathrm{cov},h}$$

8: Let u be the eigenvector corresponding to the largest eigenvalue of $\hat{\mathcal{V}}_{1,\mathrm{cov}}$.
9: If $u_1 < 0$, let $u \leftarrow -u$.
10: **Return:** $\beta = u/\|u\|_2$

Babichev and Bach [2018] derive a non-asymptotic bound on $\|\mathcal{V}_{1,\mathrm{cov}} - \hat{\mathcal{V}}_{1,\mathrm{cov}}\|_*$, where $\|\cdot\|_*$ denotes the nuclear norm, under the additional assumptions (L1)–(L4) listed below.

(L1) The function $m : \mathbb{R} \to \mathbb{R}^d$ such that $\mathbb{E}(\mathcal{S}_1(X) \mid Y = y) = m(y)$ is L-Lipschitz continuous.

(L2) The random variable $Y \in \mathbb{R}$ is sub-Gaussian, i.e., such that $\mathbb{E}e^{t(Y - Ey)} \leqslant e^{\tau_y^2 t^2/2}$, for some $\tau_y > 0$.

(L3) The random variables $\mathcal{S}_{1j}(X) \in \mathbb{R}$ are sub-Gaussian, i.e., such that $\mathbb{E}e^{t\mathcal{S}_{1j}(X)} \leqslant e^{\tau_\ell^2 t^2/2}$ for each component $j \in \{1, \ldots, d\}$, for some $\tau_\ell > 0$.

(L4) The random variables $\eta_j = \mathcal{S}_{1j}(X) - m_j(Y) \in \mathbb{R}$ are sub-Gaussian, i.e., such that $\mathbb{E}e^{t\eta_j} \leqslant e^{\tau_\eta^2 t^2/2}$ for each component $j \in \{1, \ldots, d\}$, for some $\tau_\eta > 0$.

Under (A1)–(A2) and (L1)–(L4), Babichev and Bach [2018] proves the following bound in Theorem 1: for any $\delta < \frac{1}{n}$, with probability not less than $1 - \delta$:

$$\left\| \hat{\mathcal{V}}_{1,\text{cov}} - \mathcal{V}_{1,\text{cov}} \right\|_* \leqslant \frac{d\sqrt{d}\left(195\tau_\eta^2 + 2\tau_\ell^2\right)}{\sqrt{n}}\sqrt{\log\frac{24d^2}{\delta}}$$
$$+ \frac{8L^2\tau_y^2 + 16\tau_\eta\tau_y L\sqrt{d} + \left(157\tau_\eta^2 + 2\tau_\ell^2\right)d\sqrt{d}}{n}\log^2\frac{32d^2n}{\delta}. \qquad \text{(S-35)}$$

Non-asymptotic bound for the estimated initial vector Now, combining the non-asymptotic bound for $\mathcal{V}_{1,\text{cov}}$ and Davis-Kahan Theorem, we present the non-asymptotic bound for $\hat{\beta}^{(k)}$ where $\hat{\beta}^{(k)}$ is the estimated index vector using an i.i.d sample $(X_t, Y_t^{(k)})$ of size n_k from the single index model (2).

Theorem 3. *Assume the single index model (2) and Assumption 3, along with (L1)–(L4). Let $\phi \geq 1$ be given. For sufficiently large n_k, the following bound holds with probability at least $1 - n_k^{-\phi}$:*

$$\sin \angle \hat{\beta}^{(k)}, \beta_0 \leq c(d, \tau_\eta, \tau_\ell, \lambda_1, \phi)\sqrt{\frac{\log(n_k)}{n_k}}.$$

Here $c(d, \tau_\eta, \tau_\ell, \lambda_1, \phi)$ is a constant which depends on model parameters $d, \tau_\eta, \tau_\ell, \lambda_1, K$ but not on the sample size n.

Proof. Let $\hat{\mathcal{V}}_{1,\text{cov}}^{(k)}$ be the estimated covariance matrix from Algorithm 3 using the dataset $\mathcal{D}_{\text{init}}^{(k)}$ for $k = 1, \ldots, K$. For $A \in \mathbb{R}^{d \times d}$ with singular values $\sigma_1, \ldots, \sigma_d$, we have $\|A\|_* = \sum_{i=1}^{d} \sigma_i \leq (\sum_{i=1}^{d} \sigma_i^2)^{1/2}(\sum_{i=1}^{d} 1)^{1/2} = d^{1/2}\|A\|_F$. Then from (S-35), for any $\delta < 1/n_k$, we have with probability at least $1 - \delta$:

$$\left\| \hat{\mathcal{V}}_{1,\text{cov}}^{(k)} - \mathcal{V}_{1,\text{cov}} \right\|_F \leqslant \frac{d^2\left(195\tau_\eta^2 + 2\tau_\ell^2\right)}{\sqrt{n_k}}\sqrt{\log\frac{24d^2}{\delta}}$$
$$+ \frac{8L^2\tau_y^2\sqrt{d} + 16\tau_\eta\tau_y Ld + \left(157\tau_\eta^2 + 2\tau_\ell^2\right)d^2}{n_k}\log^2\frac{32d^2n_k}{\delta}.$$

Now, by applying a variant of Davis-Kahan inequality (ref. Theorem 2 in Yu et al. [2015])

to this bound,

$$\sin \angle \hat{\beta}^{(k)}, \beta_0 \leq \frac{2\|\hat{\mathcal{V}}_{1,\text{cov}} - \mathcal{V}_{1,\text{cov}}\|_F}{\lambda_1 - \lambda_2},$$

where $\beta_0, \hat{\beta}^{(k)}$ correspond to the first eigenvector of $\mathcal{V}_{1,\text{cov}}$ and $\hat{\mathcal{V}}_{1,\text{cov}}^{(k)}$ and $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_d$ are eigenvalues of $\mathcal{V}_{1,\text{cov}}$.

Note since $k = 1$, $\mathcal{V}_{1,\text{cov}}$ should have only one non-zero eigenvalue, i.e., $\lambda_2 = 0$. Under condition where SADE is consistent, $\lambda_1 > 0$. In particular, choose $\delta = n_k^{-\phi}$ for some $\phi \geq 1$. Then with probability at least $1 - n_k^{-\phi}$,

$$
\begin{aligned}
&\sin \angle \hat{\beta}^{(k)}, \beta_0 \\
&\leq \frac{2}{\lambda_1} \left\{ \frac{d^2 \left(195\tau_\eta^2 + 2\tau_\ell^2\right)}{\sqrt{n_k}} \sqrt{\log(24d^2 n_k^\phi)} \right. \\
&\quad \left. + \frac{8L^2\tau_y^2\sqrt{d} + 16\tau_\eta\tau_y Ld + \left(157\tau_\eta^2 + 2\tau_\ell^2\right)d^2}{n_k} \log^2(32d^2 n_k^{\phi+1}) \right\} \\
&\leq \frac{2d^2 \left(195\tau_\eta^2 + 2\tau_\ell^2\right)}{\lambda_1} \sqrt{\frac{\log(24d^2 n_k^\phi)}{n_k}} \\
&\quad + \frac{2(8L^2\tau_y^2\sqrt{d} + 16\tau_\eta\tau_y Ld + \left(157\tau_\eta^2 + 2\tau_\ell^2\right)d^2)}{\lambda_1} \frac{\log^2(32d^2 n_k^{\phi+1})}{n_k} \\
&\leq \frac{2^{3/2}d^2 \left(195\tau_\eta^2 + 2\tau_\ell^2\right)\phi^{1/2}}{\lambda_1} \sqrt{\frac{\log(n_k)}{n_k}}
\end{aligned}
$$

for sufficiently large n_k, as the first term is the leading order term.

$\square$

S5 Additional Simulation Results

In addition to the simulation study in Section 5, we explore alternative covariate distributions beyond the truncated multivariate normal distribution. Specifically, for $X_t \in \mathbb{R}^d$ for $t = 1, \ldots, T$, we consider: 1) $X_t \sim N(0, \Sigma_X)$, where $\Sigma_X = 5I$, where I is the identity matrix, 2) $X_{ti} \sim \text{Unif}(-L, L)$ for $i = 1, \ldots, d$ and with $L = 3$. We consider Setting 2 from Section 5 with $T = 10^6$. The true index vector β_0 and rewards are generated exactly as in Section 5. As before, we consider both the cases:

- When the pilot direction β_0 is available under varying degree of angular permutations θ, i.e., we perturb β_0 by an angle θ ranging from $\{0.01, \ldots, \pi/2\}$ use the resulting perturbed direction in Algorithm 1.

- When the pilot direction is unknown and we use the initial $t_0 = T^{2/3}$ data to estimate using SADE algorithm [Babichev and Bach, 2018] described in Algorithm 3 for each arm and then using Algorithm 2 to construct the average index estimator. We consider varying level of model noise σ and compare the performance of the proposed Algorithm 1 with the nonparametric analogue, i.e., the BaSEDB algorithm of Jiang and Ma [2024].

The average regret over 20 replications of each algorithm is shown in Figures S7 and S8 for normally and uniformly distributed covariates, respectively. Note, the black solid and blue dashed vertical lines in all the four plots denote the $M = 5$ batches for BIDS and nonparametric analogue (BaSEDB), respectively, chosen according to the theory as described in Section 3.1. Since the width of the BaSEDB algorithm depends on the covariate dimension d, we notice that the bins are much wider in the nonparametric setting as compared to the semiparametric GMABC setting. For the case where the pilot direction is available, both for Normally distributed covariates [Figure S7(b)] and Uniformly distributed covariates [Figure S8(b)], we observe that as the perturbation, $\sin(\theta)$, increases from 0 to 0.8 (corresponds to $\theta \leq \pi/4$), the performance of the proposed algorithm deteriorates (solid green to solid red lines) and stops learning if the perturbation is larger, similar to the nonparametric analogue. However for $\theta \leq \pi/4$, it still outperforms the nonparametric analogue (dashed lines), where no arm elimination appears to occur. The decline in performance seems to be more pronounced for Normally distributed covariates compared to Uniform ones.
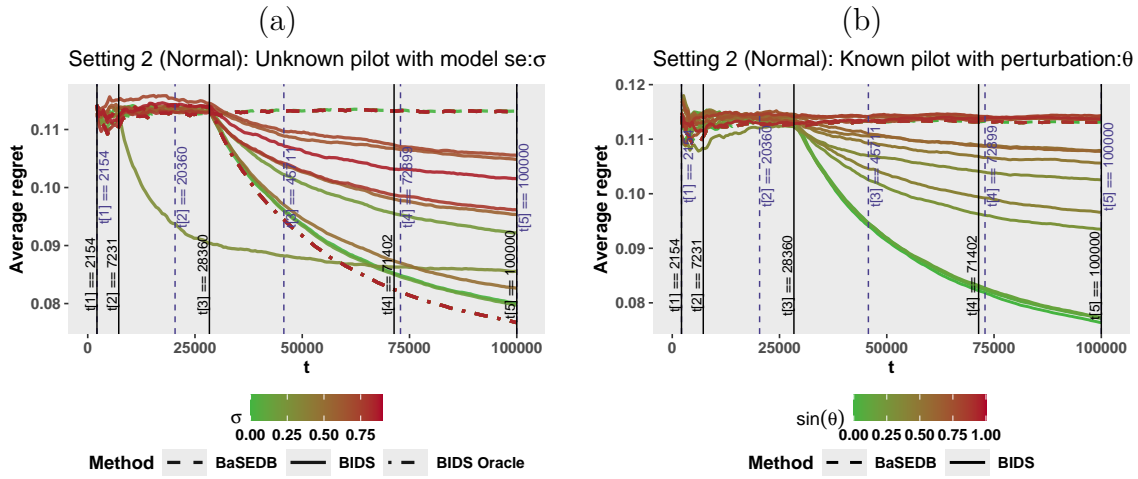
Figure S7: Average regret $((\mathcal{R}_t)_{t=1}^T)$ with normally distributed covariates. As the noise gets larger, the performance of the SIR batched bandit (solid) still beats the nonparametric analogue (dashed) but gets further way from the oracle (dashed-dotted).

When the pilot direction is unknown and Algorithm 3 is employed with the initial index estimator as described in Algorithm 2, we note that for both Normal [Figure S7(a)] and Uniform covariates [Figure S8(a)], the average regret for the proposed Algorithm 1 decreases faster than for the nonparametric analogue (dashed lines). Nonetheless, its performance degrades as the model error grows from 0.1 to 0.8 (solid green to red lines), with the decline being more pronounced for Normally distributed covariates compared to Uniform ones.
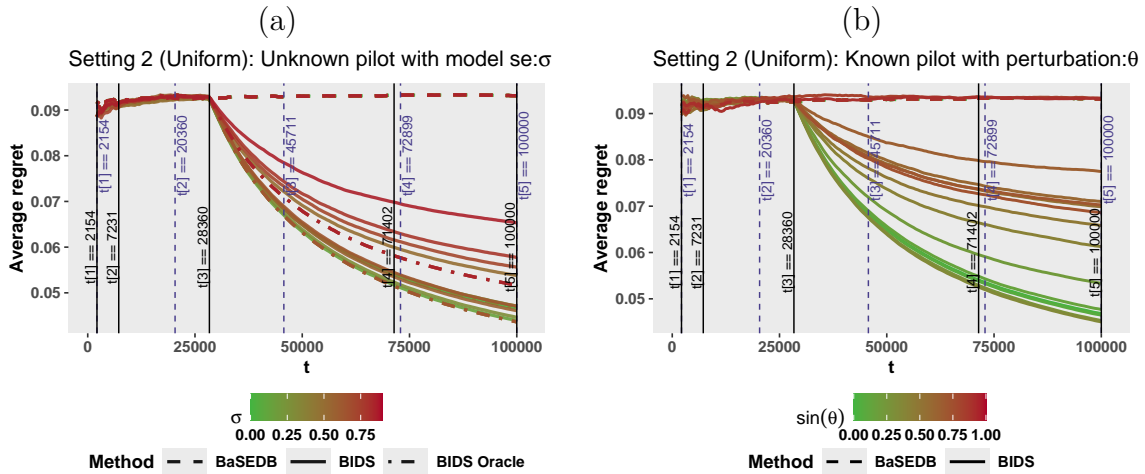

Figure S8: Average regret $((\mathcal{R}_t)_{t=1}^T)$ with uniformly distributed covariates with perturbed true direction β_0 by an angle θ. As the perturbation gets larger, the performance of the SIR batched bandit still beats the nonparametric analogue but gets further way from the oracle direction.

Finally, the performance of the proposed algorithm with the oracle direction (dashed-dotted lines) shows slight variation as model noise increases, but it remains consistently better than the other algorithms, as expected. This variation in the oracle's performance could be attributed to variability across different simulation runs of the decision-making process.

S6 Additional real data results

We compare the performance of the BIDS algorithm and the BaSEDB algorithm of Jiang and Ma [2024] when different initial batch sizes are used to estimate the direction β_0. We let $t_0 = 1$. In Figure S9, note that the columns denote increasing initial batch size $t_1 = t_{\text{init}}$, as denoted by the labels on the first vertical lines in the plots. Vertical solid lines denote the batch end points for the GMABC framework as proposed in (11), and the dashed lines denote the batch end points for the nonparametric batched bandits framework as suggested by Jiang and Ma [2024]. Since the bin-widths depend on d in nonparametric batched bandits, we see that the batch sizes are much larger than the corresponding GMABC setup where the bin-width does not depend on the number of covariates.

Similar to Section 5, we notice that BIDS outperforms BaSEDB algorithm, even though we do not know the true data generating mechanism in any of these datasets. While in the EEG dataset, for a small initial batch size ($t_{\text{init}} = 75$), the BIDS algorithm incurred large regret in the beginning, the rate of decrease is much faster. We notice that as the initial sample size increases, the average regret for the BIDS algorithm gets closer to the oracle BIDS algorithm. In fact, the regret rate for the BIDS algorithm decreases even faster than that of the oracle BIDS algorithm. This may be because, as we incorporate more data to learn the direction, we estimate the direction for each arm separately before combining them using Algorithm 2. In contrast, the oracle direction utilizes the entire dataset to determine a single direction, which could correspond to a possibly mis-specified model.
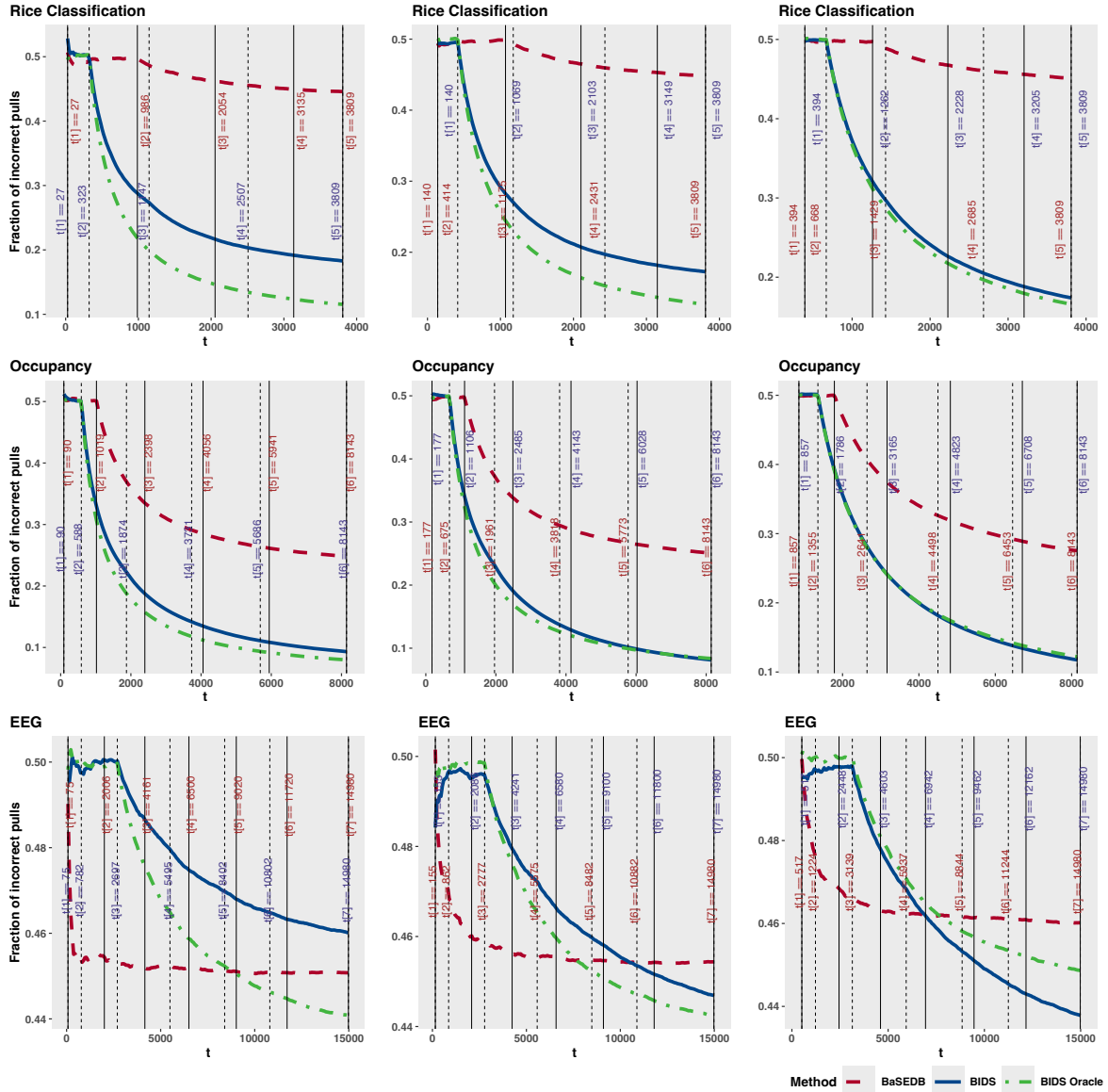
Figure S9: Comparison of expected regret of the proposed semiparametric BIDS algorithm and the nonparametric batched bandit algorithm (BaSEDB) on a) rice classification, b) occupancy detection, and c) EEG datasets with β_0 estimated in the initial phase with $t_1 = t_{\text{init}}$ increasing as we go from left to right for the respective datasets. Vertical lines denote the batch markings for both the algorithms. Observe that the BIDS outperforms BaSEDB in all instances, and for the Occupancy dataset it even performs similar to the BIDS oracle algorithm.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 127–135, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

Sakshi Arya and Bharath K Sriperumbudur. Kernel ϵ-greedy for contextual bandits. *arXiv preprint arXiv:2306.17329*, 2023.

Phoebe M Asquith and Hisham Ihshaish. Classification of eye-state using eeg recordings: speed-up gains using signal epochs and mutual information measure. In *Proceedings of the 23rd International Database Applications & Engineering Symposium*, pages 1–6, 2019.

Onur Atan, Cem Tekin, and Mihaela Van der Schaar. Global multi-armed bandits with hölder continuity. In *Artificial Intelligence and Statistics*, pages 28–36. PMLR, 2015.

Onur Atan, Cem Tekin, and Mihaela van der Schaar. Global bandits. *IEEE Transactions on Neural Networks and Learning Systems*, 29(12):5798–5811, 2018.

Dmitry Babichev and Francis Bach. Slice inverse regression with score functions. *Electronic Journal of Statistics*, 12(1):1507 – 1543, 2018.

Hamsa Bastani and Mohsen Bayati. Online decision making with high-dimensional covariates. *Operations Research*, 68(1):276–294, 2020.

Alberto Bietti, Alekh Agarwal, and John Langford. A contextual bandit bake-off. *Journal of Machine Learning Research*, 22(133):1–49, 2021.

Luis Candanedo. Occupancy Detection . UCI Machine Learning Repository, 2016.

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 208–214, Fort Lauderdale, FL, USA, 11–13 Apr 2011. PMLR.

Ilkay Cinar and Murat Koklu. Rice (Cammeo and Osmancik). UCI Machine Learning Repository, 2019.

Gokalp Cinarer, Nizamettin Erbaş, and Abdurrahman Öcal. Rice classification and quality detection success with artificial intelligence technologies. *Brazilian Archives of Biology and Technology*, 2024.

Ran Dai, Hyebin Song, Rina Foygel Barber, and Garvesh Raskutti. Convergence guarantee for the sparse monotone single index model. *Electronic Journal of Statistics*, 16(2):4449–4496, 2022.

Hossein Esfandiari, Amin Karbasi, Abbas Mehrabian, and Vahab Mirrokni. Regret bounds for batched bandits. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35 (8):7340–7348, May 2021.

Yasong Feng, Zengfeng Huang, and Tianyu Wang. Lipschitz bandits with batched feedback. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.

Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric Bandits: The generalized linear case. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010.

Alexander Goldenshluger and Assaf Zeevi. A linear response bandit problem. *Stochastic Systems*, 3(1):230–261, 2013.

Kristjan Greenewald, Ambuj Tewari, Susan Murphy, and Predag Klasnja. Action centered contextual bandits. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Quanquan Gu, Amin Karbasi, Khashayar Khosravi, Vahab Mirrokni, and Dongruo Zhou. Batched neural bandits. *ACM / IMS J. Data Sci.*, 1(1), January 2024.

Yanjun Han, Zhengqing Zhou, Zhengyuan Zhou, Jose Blanchet, Peter W Glynn, and Yinyu Ye. Sequential batch learning in finite-action linear contextual bandits. *arXiv preprint arXiv:2004.06321*, 2020.

Wolfgang Hardle, Peter Hall, and Hidehiko Ichimura. Optimal smoothing in single-index models. *The Annals of Statistics*, 21(1):157–178, 1993.

Hidehiko Ichimura. Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of econometrics*, 58(1-2):71–120, 1993.

Heinrich Jiang. Non-asymptotic uniform rates of consistency for k-nn regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3999–4006, 2019.

Rong Jiang and Cong Ma. Batched nonparametric contextual bandits. *arXiv preprint arXiv:2402.17732*, 2024.

Tianyuan Jin, Jing Tang, Pan Xu, Keke Huang, Xiaokui Xiao, and Quanquan Gu. Almost optimal anytime algorithm for batched multi-armed bandits. In *International Conference on Machine Learning*, pages 5065–5073. PMLR, 2021.

Cem Kalkanli and Ayfer Ozgur. Batched thompson sampling. In *Advances in Neural Information Processing Systems*, volume 34, pages 29984–29994. Curran Associates, Inc., 2021.

Galib Hassan Khan and Md. Akibur Rahman. Room occupancy detection from temperature, light, humidity, and carbon dioxide measurements using deep learning. In *2021 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, pages 1–4, 2021.

Gi-Soo Kim and Myunghee Cho Paik. Contextual multi-armed bandit algorithm for semi-parametric reward model. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3389–3397. PMLR, 09–15 Jun 2019.

Akshay Krishnamurthy, Zhiwei Steven Wu, and Vasilis Syrgkanis. Semiparametric contextual bandits. In *International Conference on Machine Learning*, pages 2776–2785. PMLR, 2018.

Arun K Kuchibhotla and Rohit K Patra. Efficient estimation in single index models through smoothing splines. 2020.

William Kuszmaul and Qi Qi. The multiplicative version of azuma's inequality, with an application to contention analysis. *arXiv preprint arXiv:2102.05077*, 2021.

Tze Leung Lai. Adaptive treatment allocation and the multi-armed bandit problem. *The Annals of Statistics*, pages 1091–1114, 1987.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

Ke Li, Yun Yang, and Naveen N Narisetty. Regret lower bound and optimal algorithm for high-dimensional contextual linear bandit. *Electronic Journal of Statistics*, 15(2): 5652–5695, 2021.

Ker-Chau Li and Naihua Duan. Regression analysis under link violation. *The Annals of Statistics*, pages 1009–1052, 1989.

Wenhao Li, Ningyuan Chen, and L Jeff Hong. Dimension reduction in contextual online learning via nonparametric variable selection. *Journal of Machine Learning Research*, 24 (136):1–84, 2023.

Wenjie Li, Adarsh Barik, and Jean Honorio. A simple unified framework for high dimensional bandit problems. In *International Conference on Machine Learning*, pages 12619–12655. PMLR, 2022.

Vianney Perchet and Philippe Rigollet. The multi-armed bandit problem with covariates. *The Annals of Statistics*, 2013.

Vianney Perchet, Philippe Rigollet, Sylvain Chassang, and Erik Snowberg. Batched bandit problems. *The Annals of Statistics*, 44(2):660 – 681, 2016.

Wei Qian and Yuhong Yang. Kernel estimation and model combination in a bandit problem with covariates. *Journal of Machine Learning Research*, 17(149), 2016.

Wei Qian, Ching-Kang Ing, and Ji Liu. Adaptive algorithm for multi-armed bandit problem with high-dimensional covariates. *Journal of the American Statistical Association*, 119 (546):970–982, 2024.

Zhimei Ren, Zhengyuan Zhou, and Jayant R. Kalagnanam. Batched learning in generalized linear contextual bandits with general decision sets. *IEEE Control Systems Letters*, 6: 37–42, 2022.

Philippe Rigollet and Assaf Zeevi. Nonparametric bandits with covariates. *Conference on Learning Theory (COLT)*, page 54, 2010.

Oliver Roesler. EEG Eye State. UCI Machine Learning Repository, 2013.

Oliver Rösler and David Suendermann. A first step towards eye state prediction using eeg. *Proc. of the AIHLS*, 1:1–4, 2013.

Cong Shen, Ruida Zhou, Cem Tekin, and Mihaela van der Schaar. Generalized global bandit and its application in cellular coverage optimization. *IEEE Journal of Selected Topics in Signal Processing*, 12(1):218–232, 2018.

Nirandika Wanigasekara and Christina Yu. Nonparametric contextual bandits in metric spaces with unknown metric. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Yuhong Yang and Dan Zhu. Randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates. *The Annals of Statistics*, 30(1):100–121, 2002.

Yi Yu, Tengyao Wang, and Richard J Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102:315–323, 2015.